

Estimation of time-varying covariance matrices for large datasets*

Yiannis Dendramis^{1†} Liudas Giraitis²
George Kapetanios^{3‡}

¹*Athens University of Economics and Business* ²*Queen Mary University of London*

³*King's College London*

November 9, 2020

Abstract

Time variation is a fundamental problem in statistical and econometric analysis of macroeconomic and financial data. Recently there has been considerable focus on developing econometric modelling that enables stochastic structural change in model parameters and on model estimation by Bayesian or non-parametric kernel methods. In the context of the estimation of covariance matrices of large dimensional panels, such data requires taking into account time variation, possible dependence and heavy-tailed distributions. In this paper we introduce a non-parametric version of regularisation techniques for sparse large covariance matrices, developed by Bickel and Levina (2008) and others. We focus on the robustness of such a procedure to time variation, dependence and heavy-tailedness of distributions. The paper includes a set of results on Bernstein type inequalities for dependent unbounded variables which are expected to be applicable in econometric analysis beyond estimation of large covariance matrices. We discuss the utility of the robust thresholding method, comparing it with other estimators in simulations and an empirical application on the design of minimum variance portfolios.

Keywords: covariance matrix estimation, large dataset, regularization, thresholding, shrinkage, exponential inequalities, minimum variance portfolio.

JEL classification: C13; C22; C51.

1 Introduction

This paper considers estimation of large covariance matrices under structural change, possible dependence and heavy-tailed distributions. The problem of structural change in the

*We thank the Editor, the Co-Editor and the Referees for constructive comments.

[†]Yiannis Dendramis acknowledges support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 708501.

[‡]Corresponding address: School of Management & Business, Kings College, 30 Aldwych, London, U.K. Tel: +44 20 7848 3770. E-mail address: george.kapetanios@kcl.ac.uk (G.K.)

econometric literature has been mainly addressed within either univariate or relatively small multivariate models and settings. Existing research is primarily focused on factor models where datasets are summarised by a finite set of unobserved time series, usually referred to as factors. Examples of this literature include Stock and Watson (2002a, 2002b). There has been little work done on structural change in large datasets. This paper addresses the issue without recourse to factor modelling.

Estimation of large covariance matrices is particularly demanding since the number of estimated objects rises as a square of the dimension of the dataset leading to a large amount of aggregated estimation error. Several regularization techniques for improved estimation of large covariance matrices have been proposed which are summarised briefly in the body of the paper. These include Ledoit and Wolf (2004), Bickel and Levina (2008), Cai and Liu (2011) and Abadir, Distaso, and Zikes (2014). See also the excellent review article by Fan, Liao, and Liu (2016) and references therein. That review article focuses on structure-based estimators of covariance and precision matrices - that is estimators that assume sparsity or a factor model. For structure-free estimators, see Pourahmadi (2013).

In this paper we develop regularized thresholding estimation in the presence of structural change of a general unspecified form. The characterisation of the change is an important aspect of the problem. In contrast to the majority of the work we allow for smooth deterministic or stochastic change of the covariance matrix rather than structural breaks. Some new characterizations of smooth change introduced in this paper expand upon the notion of smoothness used in nonparametric inference.

There is limited literature on regularized estimation under deterministic structural change, see Chen, Xu, and Wu (2013), Zhou, Lafferty, and Wasserman (2010) and Kolar and Xing (2011); strong mixing, see Fan, Liao, and Mincheva (2013), and heavy-tailed data (independent, modelled by elliptical distribution), see, e.g., Wegkamp and Zhao (2016), Han and Liu (2017), and Fan, Wang, and Zhong (2016). All this work assumes that the volatility process is a deterministic function or a constant which is considered extremely restrictive for economic and financial data, both from a theoretical and an empirical point of view. In a different stochastic setup than the one considered in the paper, Bickel, Wang, and Zhou (2013) have provided a theoretical analysis for Bickel-Levina thresholding in the context of realized covariance, estimated from noisy data.

We provide a unified framework for estimation of the paths of large covariance matrices, that change over time, in a potentially stochastic way, with temporal dependence, characterised by mixing, and with potentially heavy tails. The key characteristic of such regularized estimation is robustness to all those data features, which are highly likely to be present in

financial and economic data.

Deterministic change of covariance corresponds to the typical characterization of heteroscedasticity. Stochastic (persistent) change of volatility, also considered in our paper, is a standard vehicle for the modeling of structural change in the economy. It differs from ARCH type volatility models that have dominated the financial econometrics literature since their introduction in Engle (1982). It aligns with the empirically established fact that stochastic change in volatility may be more persistent and smooth than that allowed by stationary conditional volatility models, see Kapetanios (2010).

In order to achieve such a characterization of persistent volatility, we follow the modelling framework of Giraitis, Kapetanios and Yates (2014, 2018) and specify that the volatility process is, potentially stochastic, bounded and has small increments. That framework, in turn, extended the work of, e.g., Dahlhaus (1997) on locally stationary processes, where change was assumed to be smooth and deterministic, in nature.

This paper is organised as follows. In Section 2, we present the main results. We show that the thresholding estimation procedure developed by Bickel and Levina (2008) and others, in general, is robust to dependence, random scaling (volatility) and heteroscedasticity and the type of distribution of the data. In Section 3, we discuss cross-validation methods for the selection of tuning parameters and use simulations to compare the performance of regularized estimation methods. In that Section, we further provide an application of regularized estimation methods for designing minimum variance portfolios. Section 4 contains new results on Bernstein type inequalities for dependent data, that are of independent interest. They are particularly important for the rigorous analysis of penalised regression methods, like Lasso, in the presence of time series dependence. The online supplement contains proofs and technical results.

2 Thresholding Estimation

Given a sample $(\mathbf{y}_1, \dots, \mathbf{y}_T)$ of a p -variate process $\mathbf{y}_t = (y_{1,t}, \dots, y_{p,t})'$, estimation of the population covariance $p \times p$ matrix Σ by the sample covariance is a well defined procedure when p is fixed. For large covariance matrices, when p increases with T the poor performance of the sample covariance matrix estimate $\hat{\Sigma}$ can be improved by various regularization procedures which include the thresholding methods developed by Bickel and Levina (2008), Cai and Liu (2011), Fan, Liao, and Mincheva (2013) and others.

The aim of this paper is to investigate the impact of dependence, heteroscedasticity and distribution of the data on consistency rates for regularised estimation of the covariance matrix using the Bickel and Levina (2008) thresholding procedure. Under heteroscedasticity, the

covariance matrix $\Sigma_t = \text{var}(\mathbf{y}_t) = E[(\mathbf{y}_t - E\mathbf{y}_t)(\mathbf{y}_t' - E\mathbf{y}_t)']$ varies over time, and can be estimated by the local sample covariance matrix $\widehat{\Sigma}_t$, see (10). For data with a random scaling (volatility) component, the limit of $\widehat{\Sigma}_t$ is less obvious. As shown in the following analysis, such a limit comprises a time-varying deterministic matrix and a random scale. The novelty of our theoretical findings is showing that the Bickel-Levina thresholding procedure is robust to dependence, light or heavy tailedness of the data and it can be adjusted for heteroscedasticity.

To reflect the robustness properties of regularised estimation of Σ_t , we consider three settings for \mathbf{y}_t . Before proceeding further, we provide assumptions, notation and definitions.

Assumption M. (i) The centered stochastic process $\mathbf{y}_t - E\mathbf{y}_t$ is an α -mixing (but not necessarily stationary) process with mixing coefficients α_k^1 such that for some $0 < \phi < 1$ and $c > 0$,

$$\alpha_k \leq c\phi^k, \quad k \geq 1. \quad (1)$$

(ii) The elements of $E\mathbf{y}_t = (\mu_{1,t}, \dots, \mu_{p,t})'$ and $\text{var}(\mathbf{y}_t) = \Sigma_t = (\sigma_{jk,t})_{j,k=1,\dots,p}$ are such that

$$|\mu_{j,t}| \leq C, \quad |\mu_{j,t} - \mu_{j,s}| \leq C \frac{|t-s|}{\max(t,s)}, \quad |\sigma_{jk,t}| \leq C, \quad |\sigma_{jk,t} - \sigma_{jk,s}| \leq C \frac{|t-s|}{\max(t,s)} \quad (2)$$

for $1 \leq t, s \leq T$ where $C < \infty$ does not depend on j, k and t, s and T .

The components of $\mathbf{y}_t = (y_{1,t}, \dots, y_{p,t})'$ can have a wide variety of tail behavior. We write $(\mathbf{y}_t) \in \mathcal{E}(s)$, $s > 0$ to denote a thin-tailed distribution for $y_{j,t}$:

$$\max_{j,t} E \exp(a|y_{j,t}|^s) < \infty \quad \text{for some } a > 0. \quad (3)$$

The notation $(\mathbf{y}_t) \in \mathcal{H}(\theta)$, $\theta > 2$ denotes a heavy-tailed distribution property:

$$\max_{j,t} E|y_{j,t}|^\theta < \infty. \quad (4)$$

The definitions (3) and (4) imply that there exists $c_0 > 0$, $c_1 > 0$ such that for all $\zeta > 0$, $t \geq 1$,

$$P(|y_{i,t}| \geq \zeta) \leq \begin{cases} c_0 \exp(-c_1 \zeta^s) & \text{if } (y_{j,t}) \in \mathcal{E}(s), s > 0, \\ c_0 \zeta^{-\theta} & \text{if } (y_{j,t}) \in \mathcal{H}(\theta), \theta > 2. \end{cases}$$

Thresholding under stationarity. Suppose that \mathbf{y}_t is a stationary α -mixing process. Then, $E\mathbf{y}_t = \boldsymbol{\mu}$ and $\Sigma = \text{var}(\mathbf{y}_t) = [\sigma_{ij}]$ do not depend on t . Denote by $\widehat{\Sigma} = [\widehat{\sigma}_{ij}]$ the sample covariance estimate of Σ ,

$$\widehat{\Sigma} = T^{-1} \sum_{j=1}^T \mathbf{y}_j \mathbf{y}_j' - \bar{\mathbf{y}} \bar{\mathbf{y}}', \quad \bar{\mathbf{y}} = T^{-1} \sum_{j=1}^T \mathbf{y}_j. \quad (5)$$

¹Let $\mathcal{F}_{-\infty}^j$, \mathcal{F}_j^∞ denote σ -fields generated by $\{\mathbf{y}_t, t \leq j\}$ and $\{\mathbf{y}_t, t \geq j\}$ respectively. Define the α -mixing coefficient as $\alpha_k = \sup_j \sup_{A \in \mathcal{F}_{-\infty}^j, B \in \mathcal{F}_{j+k}^\infty} |P(A)P(B) - P(A \cap B)|$.

Hard and adaptive thresholding, introduced by Bickel and Levina (2008) and Cai and Liu (2011), are two standard procedures to regularize $\widehat{\Sigma}$, when p increases with T . Hard thresholding is based on the idea of setting the elements of $\widehat{\Sigma}$, whose absolute values are smaller than some threshold λ , to zero. It yields the estimate

$$T_\lambda(\widehat{\Sigma}) = (\widehat{\sigma}_{ij} I(|\widehat{\sigma}_{ij}| > \lambda)).$$

Procedures based on other thresholding operators can be defined, but they have similar properties to hard thresholding, asymptotically, although they may differ in finite samples.

In their seminal work Bickel and Levina (2008) showed that if \mathbf{y}_t is an i.i.d. Gaussian process, then the regularized estimator $T_\lambda(\widehat{\Sigma})$ of Σ under a sparsity assumption is consistent,

$$\|T_\lambda(\widehat{\Sigma}) - \Sigma\| = O_P(n_p \lambda) \quad \text{with} \quad \lambda = \kappa \sqrt{\frac{\log p}{T}}, \quad (6)$$

$$\|T_\lambda(\widehat{\Sigma})^{-1} - \Sigma^{-1}\| = O_P(n_p \lambda) \quad \text{if } n_p \lambda = o(1) \text{ and } \|\Sigma\| \geq c > 0 \quad (7)$$

where $\|\cdot\|$ denotes the spectral norm and κ is a tuning parameter. Thresholding estimation as a rule assumes that Σ is approximately sparse, i.e. the sparsity parameter, n_p , which is the maximum number of non-zero elements in a row of Σ ,

$$\max_{i=1, \dots, p} \sum_{j=1}^p I(\sigma_{ij} \neq 0) = n_p, \quad (8)$$

does not grow too fast with p .

As in Bickel and Levina (2008) and Fan, Liao, and Mincheva (2013), we consider, as a leading case, the one where Σ is sparse, although the theory can accommodate large $n_p \leq p$. We assume that $T, p \rightarrow \infty$.

Theorem 1 *Let $\mathbf{y}_1, \dots, \mathbf{y}_T$ be a sample from a stationary process (\mathbf{y}_t) which satisfies Assumption M. Let $c > 0$, $\varepsilon > 0$ and*

$$T \geq cp^\varepsilon, \quad T, p \rightarrow \infty. \quad (9)$$

Then, for sufficiently large $\kappa > 0$, the regularised estimate $T_\lambda(\widehat{\Sigma})$ of Σ satisfies (6) and (7) in the following cases.

- (i) *If $(\mathbf{y}_t) \in \mathcal{E}(s)$, $s > 0$.*
- (ii) *If $(\mathbf{y}_t) \in \mathcal{H}(\theta)$, $\theta > 4$ and $\varepsilon > 8/(\theta - 4)$ in (9).*

This theorem shows that the Bickel-Levina thresholding procedure is valid for a stationary α -mixing process (\mathbf{y}_t) which may have a non-Gaussian, heavy-tailed distribution. For \mathbf{y}_t with

a heavy-tailed distribution, it requires at least four finite moments and p should be relatively small compared to T .

Thresholding under heteroscedasticity. In economic applications, heteroscedasticity may arise due to smooth changes of the mean and unconditional variance of observations. Next we consider the case when \mathbf{y}_t is a heteroscedastic process, i.e. $\text{var}(\mathbf{y}_t) = \boldsymbol{\Sigma}_t = [\sigma_{ij,t}]$ and $E[\mathbf{y}_t] = \boldsymbol{\mu}_t$ vary with t . To account for variation in $\boldsymbol{\Sigma}_t$, we use a kernel type sample covariance estimator $\widehat{\boldsymbol{\Sigma}}_t = [\widehat{\sigma}_{ij,t}]$ of $\boldsymbol{\Sigma}_t$ at time $t = 1, \dots, T$ given by

$$\widehat{\boldsymbol{\Sigma}}_t \equiv \widehat{\boldsymbol{\Sigma}}_{T,t} = K_t^{-1} \sum_{k=1}^T b_{H,|t-k|} \mathbf{y}_k \mathbf{y}_k' - \bar{\mathbf{y}}_t \bar{\mathbf{y}}_t', \quad \bar{\mathbf{y}}_t = K_t^{-1} \sum_{k=1}^T b_{H,|t-k|} \mathbf{y}_k \quad (10)$$

where $K_t = \sum_{k=1}^T b_{H,|t-k|}$ and $b_{H,|t-k|}$ are kernel weights defined in (12). The regularized sample covariance estimate of $\boldsymbol{\Sigma}_t$ is defined as

$$T_\lambda(\widehat{\boldsymbol{\Sigma}}_t) = (\widehat{\sigma}_{ij,t} I(|\widehat{\sigma}_{ij,t}| > \lambda)). \quad (11)$$

Below we modify the thresholding procedure to account for heteroscedasticity.

We assume that $\boldsymbol{\Sigma}_t$ is approximately sparse, i.e. the maximum number of non-zero elements in each row of $\boldsymbol{\Sigma}_t$ for all t , n_p , is either finite, or does not grow too fast with p . The rate of change of $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\mu}_t$ will be controlled by (2) of Assumption M.

We will consider kernel estimates (10) with weights

$$b_{H,|t-k|} = K(|t-k|/H), \quad (12)$$

where $H \rightarrow \infty$, $H = o(T)$. $K(x)$, $x \in (0, a)$ is a non-negative continuous function with finite or infinite support, such that, for some $C > 0$ and $\nu > 3$,

$$K(x) \leq C(1+x^\nu)^{-1}, \quad |(d/dx)K(x)| \leq C(1+x^\nu)^{-1}, \quad x \in (0, a). \quad (13)$$

In particular, the functions $K(x) = I(0 \leq x \leq 1)$, $K(x) = (1+x^\nu)^{-1}$ with $\nu > 2$ and $K(x) = \exp(-cx^\alpha)$ with $\alpha > 0$ satisfy (13).

The following theorem establishes the consistency properties of the regularised estimate $T_\lambda(\widehat{\boldsymbol{\Sigma}}_t)$ of $\boldsymbol{\Sigma}_t$. Under heteroscedasticity the threshold parameter λ in (15) is larger than under stationarity in Theorem 1. It depends on the bandwidth H (“window size”) and accounts for the bias-variance tradeoff of change in mean $\boldsymbol{\mu}_t = E[\mathbf{y}_t]$ and covariance $\boldsymbol{\Sigma}_t = \text{var}(\mathbf{y}_t)$. We denote by n_p the sparsity parameter of $\boldsymbol{\Sigma}_t$ defined as in (8).

Theorem 2 *Suppose that $\mathbf{y}_1, \dots, \mathbf{y}_T$ is a sample from a p -dimensional heteroscedastic process (\mathbf{y}_t) which satisfies Assumption M. Let $c_0, c, \delta, \varepsilon > 0$.*

Then, the regularised estimate $T_\lambda(\widehat{\Sigma}_t)$ of Σ_t , $\delta T \leq t \leq T$, with bandwidth H such that

$$c_0 p^\varepsilon \leq H = o(T/\sqrt{\log T}) \quad \text{as } T, p \rightarrow \infty \quad (14)$$

for sufficiently large κ has the properties

$$\|T_\lambda(\widehat{\Sigma}_t) - \Sigma_t\| = O_P(n_p \lambda) \quad \text{with } \lambda = \kappa \sqrt{\log p} \max\left(\frac{1}{\sqrt{H}}, \frac{H}{T}\right), \quad (15)$$

$$\|T_\lambda(\widehat{\Sigma}_t)^{-1} - \Sigma_t^{-1}\| = O_P(n_p \lambda) \quad \text{if } n_p \lambda = o(1) \text{ and } \|\Sigma_t\| \geq c > 0 \quad (16)$$

in the following two cases.

(i) If $(\mathbf{y}_t) \in \mathcal{E}(s)$, $s > 0$.

(ii) If $(\mathbf{y}_t) \in \mathcal{H}(\theta)$, $\theta > 4$ and $\varepsilon > 8/(\theta - 4)$ in (14).

The bandwidth $H_{opt} = T^{2/3}$ yields the lowest threshold $\lambda_{opt} = \kappa \sqrt{\frac{\log p}{H_{opt}}} = \kappa \frac{\sqrt{\log p}}{T^{1/3}}$ in (15).

Condition (14) on H is similar to that on T in equation (9), Theorem 1. The choice of λ reflects the bias–variance tradeoff of nonparametric inference. Under heteroscedasticity, dependence and thin-tailed or heavy-tailed distribution of (\mathbf{y}_t) , the optimal bandwidth, $H_{opt} = T^{2/3}$ yields the threshold $\lambda_{opt} = \kappa \sqrt{\frac{\log p}{H_{opt}}}$ which is the Bickel-Levina threshold (6) with T replaced by H_{opt} . The tuning parameter κ can be selected by cross-validation.

We complete this subsection with typical examples of μ_t .

Example 3 The following two sequences of real numbers μ_1, \dots, μ_T satisfy assumption (2).

a) $\mu_t = t^{-1} \sum_{j=1}^t a_j$, $t \geq 1$ where a_j is a bounded sequence of real numbers.

b) $\mu_t = \mu_{Tt} = g(t/T)$, $1 \leq t \leq T$ where $g(x)$, $x \in [0, 1]$ is a bounded function with a bounded derivative. Then,

$$|\mu_t - \mu_s| \leq C \frac{|t - s|}{T} \leq C \frac{|t - s|}{\max(t, s)} \quad \text{for } 1 \leq t, s \leq T.$$

Thresholding under random scaling. Allowing for heteroscedasticity that involves stochastic scaling is crucial in modelling macroeconomic and financial data.

Such data can be put in the form

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t, \quad (17)$$

where \mathbf{x}_t is a heteroscedastic α -mixing process with time-varying mean $\boldsymbol{\mu}_t^{(x)} = E\mathbf{x}_t$ and covariance matrix $\Sigma_t^{(x)} = \text{var}(\mathbf{x}_t)$, and $\mathbf{H}_t = (h_{ij,t})$ is a random persistent $p \times p$ matrix-valued scaling process. We show below, that for such \mathbf{y}_t the limit of $T_\lambda(\widehat{\Sigma}_t)$ is a time varying matrix

$$\Sigma_t = \mathbf{H}_t \Sigma_t^{(x)} \mathbf{H}_t'$$

which itself is a random process. Although no restriction on the dependence between \mathbf{H}_t and \mathbf{x}_t is imposed, the elements $(h_{ij,t})$ of \mathbf{H}_t are assumed to be persistent (smooth) and thin-tailed processes. The latter is formalized as follows.

Assumption H. The elements $(h_{ij,t})$ of \mathbf{H}_t have the following property. For all $j, k = 1, \dots, p$, $1 \leq t, s \leq T$, $T \geq 2$,

$$|h_{jk,t} - h_{jk,s}| \leq \left(\frac{|t-s|}{\max(t,s)} \right)^{1/2} h_{jk,ts} \quad (18)$$

where $(h_{jk,t}) \in \mathcal{E}(\alpha)$, $(h_{jk,ts}) \in \mathcal{E}(\alpha)$ for some $\alpha > 0$.

Under Assumption H, all processes $h_{jk,t}$ and $h_{jk,ts}$ satisfy the thin tail property (3) with the same parameters a .

In Theorem 4 we assume that the sparsity parameter n_H of \mathbf{H}_t is bounded for all t and p , while the sparsity parameter n_p of $\Sigma_t^{(x)}$ may increase with p .

Theorem 4 *Suppose that $\mathbf{y}_1, \dots, \mathbf{y}_T$ is a sample from a p -dimensional process, \mathbf{y}_t as in (17). Assume that \mathbf{x}_t satisfies Assumption M and \mathbf{H}_t satisfies Assumption H with parameter $\alpha > 0$. Let $\nu = (\alpha + 4)/(2\alpha)$ and $c_0, c, \delta, \varepsilon > 0$.*

Then, the regularised estimate $T_\lambda(\widehat{\Sigma}_t)$ of $\Sigma_t = \mathbf{H}_t \Sigma_t^{(x)} \mathbf{H}_t'$, $\delta T \leq t \leq T$, with the bandwidth H such that

$$c_0 p^\varepsilon \leq H = o(T/(\log T)^\nu) \quad \text{as } T, p \rightarrow \infty \quad (19)$$

for sufficiently large κ has the property that

$$\|T_\lambda(\widehat{\Sigma}_t) - \Sigma_t\| = O_P(n_p \lambda) \quad \text{with } \lambda = \kappa (\log p)^\nu \max\left(\frac{1}{\sqrt{H}}, \left(\frac{H}{T}\right)^{1/2}\right), \quad (20)$$

$$\|T_\lambda(\widehat{\Sigma}_t)^{-1} - \Sigma_t^{-1}\| = O_P(n_p \lambda) \quad \text{if } n_p \lambda = o(1) \text{ and } \|\Sigma_t\| \geq c > 0 \quad (21)$$

in the following two cases.

(i) If $(\mathbf{x}_t) \in \mathcal{E}(s)$, $s > 0$.

(ii) If $(\mathbf{x}_t) \in \mathcal{H}(\theta)$, $\theta > 4$ and $\varepsilon > 8/(\theta - 4)$ in (19).

The bandwidth $H_{opt} = T^{1/2}$ yields the lowest threshold $\lambda_{opt} = \kappa \sqrt{\frac{(\log p)^{2\nu}}{H_{opt}}} = \kappa \frac{(\log p)^\nu}{T^{1/4}}$ in (20).

In the presence of random scaling, the threshold, λ , in Theorem 4 is larger and the consistency rate slower than in Theorem 2 under heteroscedasticity. The optimal threshold λ_{opt} depends on the unknown parameter $\nu = (\alpha + 4)/(2\alpha)$ where α is the tail parameter of \mathbf{H}_t . If $\alpha = 2$, then $\nu = 3/2$. If \mathbf{H}_t is bounded, then $\alpha = \infty$ and $\nu = 1/2$. In this case $\lambda_{opt} = \kappa \sqrt{(\log p)/H_{opt}}$. Parameter κ can be selected by cross-validation.

Next, we provide two examples of processes satisfying (18).

Example 5 Let $\xi_j \sim IIDN(0, 1)$. The following sequences ζ_t , $t \geq 1$ satisfy (18) with $\alpha = 2$.

a) $\zeta_t = t^{-1/2} \sum_{j=1}^t \xi_j$, $1 \leq t \leq T$,

b) $\zeta_t = T^{-1/2} \sum_{j=1}^t \xi_j$.

To verify a), let $t > s$. Write

$$\begin{aligned} |\zeta_t - \zeta_s| &= \left| t^{-1/2} \sum_{j=s+1}^t \xi_j + (t^{-1/2} - s^{-1/2}) \sum_{j=1}^s \xi_j \right| \\ &\leq \frac{(t-s)^{1/2}}{t^{1/2}} \left| \frac{1}{(t-s)^{1/2}} \sum_{j=s+1}^t \xi_j \right| + \frac{|t^{1/2} - s^{1/2}|}{t^{1/2}} \left| \frac{1}{s^{1/2}} \sum_{j=1}^s \xi_j \right| \\ &\leq \frac{(t-s)^{1/2}}{t^{1/2}} \zeta_{ts}, \quad \zeta_{ts} = \frac{1}{(t-s)^{1/2}} \left| \sum_{j=s+1}^t \xi_j \right| + \frac{1}{s^{1/2}} \left| \sum_{j=1}^s \xi_j \right| \end{aligned}$$

noting that $t^{1/2} - s^{1/2} \leq (t-s)^{1/2}$. Since $\zeta_t \sim N(0, 1)$, then $(t-s)^{-1/2} \sum_{j=s+1}^t \xi_j \sim N(0, 1)$ which implies $(\zeta_t) \sim \mathcal{E}(2)$, $(\zeta_{ts}) \sim \mathcal{E}(2)$ and verifies (18) with $\alpha = 2$.

b) Observe that $|\zeta_t| \leq t^{-1/2} \left| \sum_{j=1}^t \xi_j \right|$ and $|\zeta_t - \zeta_s| \leq (t-s)^{1/2} T^{-1/2} \left| \sum_{j=s+1}^t \xi_j \right|$ which implies (18) with $\alpha = 2$.

Throughout our analysis we have assumed a significant degree of sparsity for the covariance matrix. The issue of sparsity can be addressed by extending the method developed by Fan, Liao, and Mincheva (2013), based on a factor structure, to the time varying case. To do so note that by Stock and Watson (2002a), factors can be consistently estimated, using principal component analysis, in the presence of structural change. Then, one can use time varying regression analysis, based on Giraitis, Kapetanios, and Yates (2018) to obtain time varying factor loadings and associated residuals. The (sparse) covariance matrix of these residuals can then be analysed using our proposed method. Due to the considerable technical arguments needed to implement this approach rigorously we leave this to future research.

3 Implementation of regularized estimation

In this section we compare the finite sample performance of various approaches for the estimation of large dimensional covariance matrices. We also examine their usefulness for one step ahead out-of-sample forecasting of such matrices.

Besides thresholding, another popular method of regularizing the sample covariance estimator is based on shrinkage. Ledoit and Wolf (2003, 2004) have promoted this approach in a series of papers. We consider the full sample and time-varying version of shrinkage estimators, defined as

$$\widehat{\Sigma}_{LW} = \rho \mu_T \mathbf{I} + (1 - \rho) \widehat{\Sigma}, \quad \widehat{\Sigma}_{LW,t} = \rho \mu_{t,T} \mathbf{I} + (1 - \rho) \widehat{\Sigma}_t,$$

where $\mu_T = p^{-1} \text{tr}(\widehat{\Sigma})$, $\mu_{t,T} = p^{-1} \text{tr}(\widehat{\Sigma}_t)$. Here the matrices $\mu_t \mathbf{I}$ and $\mu_{t,T} \mathbf{I}$ approximate the shrinkage target $\text{diag}(\widehat{\Sigma})$ and $\text{diag}(\widehat{\Sigma}_t)$ and the shrinkage intensity parameter $\rho \in [0, 1]$ can be

obtained through cross-validation. Ledoit and Wolf (2004) suggested to evaluate $\widehat{\Sigma}_{LW}$ with the theoretically optimal

$$\rho_{opt} = \frac{b_T^2}{d_T^2}, \quad d_T^2 = p^{-1} \text{tr}(\widehat{\Sigma}^2) - \mu_T^2, \quad b_T^2 = \min(\bar{b}_T^2, d_T^2),$$

$$\bar{b}_T^2 = \frac{1}{pT^2} \sum_{t=1}^T \left(\sum_{i=1}^p y_{it}^2 \right)^2 - \frac{1}{pT} \text{tr}(\widehat{\Sigma}^2).$$

It is of relevance and interest to see how linear shrinkage and its time-varying version compares to thresholding. Therefore, we implement the linear shrinkage estimator for both simulation and empirical exercise. For i.i.d. data, the linear shrinkage estimator has by now been superseded by the nonlinear shrinkage estimator of Ledoit and Wolf (2015). It is important to note, however, that the latter, unlike its linear counterpart, depends crucially on random matrix theoretical arguments. Therefore, it requires assuming that data are i.i.d. and p/T has a finite limit, both of which we consider as too restrictive for our purposes. While the analysis of the linear shrinkage estimator has also been carried out under similar assumptions, it is likely that a number of desirable properties of this estimator carry over to less strict settings than the i.i.d. one. In our empirical application, we also consider this type of shrinkage estimator.

In the literature, other popular methods for large covariance estimation have been proposed, which do not account explicitly for dependent data. For instance, Cai and Liu (2011) in a recent paper provide adaptive threshold estimators for the large covariance matrix Σ . In their approach the threshold λ_{ij} depends on the i, j -th entry of the matrix $\Sigma = \text{var}(\mathbf{y}_t)$ as follows:

$$\lambda_{ij} = \delta \sqrt{\frac{\widehat{\theta}_{ij} \log p}{T}}, \quad \delta > 0, \tag{22}$$

$$\widehat{\theta}_{ij} = T^{-1} \sum_{t=1}^T ((y_{it} - \bar{y}_i)(y_{jt} - \bar{y}_j) - \widehat{\sigma}_{ij})^2, \quad \bar{y}_i = T^{-1} \sum_{j=1}^T y_{it}.$$

Now, the lower bound for an off diagonal element of Σ is not global, but depends on the variability of the individual point estimate of the i, j -th element of the sample covariance matrix $\widehat{\Sigma}$.

The adaptive thresholding approach of Cai and Liu is further extended in another stimulating paper by Fan, Liao, and Mincheva (2013). The authors introduce the Principal Orthogonal complement Thresholding (POET) estimator, to account for non sparsity, using an approximate factor model. This is done by combining the factor based covariance matrix estimator and the thresholding approach developed in Cai and Liu (2011). To this end, consider the

approximate factor model for \mathbf{y}_t ,

$$\mathbf{y}_t = \mathbf{b}\mathbf{f}_t + \mathbf{u}_t, \quad (23)$$

where \mathbf{b} is the $p \times K$ matrix of factor loadings, \mathbf{f}_t is the $K \times 1$ vector of unobserved factors, with $K \ll p$, and \mathbf{u}_t is the idiosyncratic error which is uncorrelated with the factor \mathbf{f}_t . The factors \mathbf{f}_t summarize information of the large vector \mathbf{y}_t . Model (23) implies that

$$\Sigma = \mathbf{b}\Sigma_f\mathbf{b}' + \Sigma_u$$

where Σ_f , Σ_u are defined as $\Sigma_f = \text{var}(\mathbf{f}_t)$, and $\Sigma_u = \text{var}(\mathbf{u}_t)$. When (23) is an approximate factor model, Σ_u is a non diagonal but sparse matrix. A natural estimator is defined as

$$\widehat{\Sigma}_{poet} = \widehat{\mathbf{b}}\widehat{\Sigma}_f\widehat{\mathbf{b}}' + T_{\lambda_{ij}}(\widehat{\Sigma}_u)$$

with $\widehat{\mathbf{b}}$, $\widehat{\mathbf{f}}_t$ estimated by PCA, $\widehat{\Sigma}_f = T^{-1}\widehat{\mathbf{f}}'\widehat{\mathbf{f}}$, $\widehat{\Sigma}_u = T^{-1}\widehat{\mathbf{u}}'\widehat{\mathbf{u}}$, $\widehat{\mathbf{u}}_t = \mathbf{y}_t - \widehat{\mathbf{f}}_t\widehat{\mathbf{b}}$, and $T_{\lambda_{ij}}(\widehat{\Sigma}_u)$ is the regularized estimate of $\widehat{\Sigma}_u$, which is adaptively thresholded (see (22)). In practice, the number of factors K is chosen by information criteria, K_{opt} , as the one used in (34).

3.1 Cross Validation methods

Estimation of large covariance matrices requires the use of a number of tuning parameters. All the estimators presented above can be generically denoted as $m_\gamma(\widehat{\Sigma})$ and $m_\gamma(\widehat{\Sigma}_t)$, where m_γ is some function and γ is a vector of tuning parameters, e.g., ρ for the Ledoit and Wolf estimator, κ for the Bickel and Levina, (ρ, H) for time varying Ledoit and Wolf, (κ, H) or only κ for the time varying Bickel and Levina, and δ for the Cai and Liu and POET estimators. A popular approach to obtaining values for tuning parameters is via cross-validation. Cross-validation schemes are especially well-suited to the time-varying framework. We propose for this framework two objective functions that the optimal choice of parameter(s) γ minimize over a sensible parameter space:

$$Q_{T,\gamma}^{(1)} := \left\| \frac{1}{T - T_a} \sum_{t=T_a+1}^T (m_\gamma(\widehat{\Sigma}_{t|t-1}) - \mathbf{y}_t^0 \mathbf{y}_t^{0'}) (m_\gamma(\widehat{\Sigma}_{t|t-1}) - \mathbf{y}_t^0 \mathbf{y}_t^{0'})' \right\|_F, \quad (24)$$

$$Q_{T,\gamma}^{(2)} := \left\| \frac{1}{T - T_a} \sum_{t=T_a+1}^T (m_\gamma(\widehat{\Sigma}_{t|t-1})^{-1/2} \mathbf{y}_t^0 \mathbf{y}_t^{0'} m_\gamma(\widehat{\Sigma}_{t|t-1})^{-1/2} - I) \right. \\ \left. \times (m_\gamma(\widehat{\Sigma}_{t|t-1})^{-1/2} \mathbf{y}_t^0 \mathbf{y}_t^{0'} m_\gamma(\widehat{\Sigma}_{t|t-1})^{-1/2} - I)' \right\|_F \quad (25)$$

where $\mathbf{y}_t^0 = \mathbf{y}_t - \bar{\mathbf{y}}_t$ with $\bar{\mathbf{y}}_t$ as in (10). They are based on the estimate of $\text{var}(\mathbf{y}_t | \mathbf{y}_{t-1})$, $\widehat{\Sigma}_{t|t-1}$, defined as

$$\widehat{\Sigma}_{t|t-1} = \left(\sum_{j=1}^{t-1} b_{H,|j-t|} \right)^{-1} \sum_{j=1}^{t-1} b_{H,|j-t|} \mathbf{y}_j \mathbf{y}_j' - \bar{\mathbf{y}}_{t|t-1} \bar{\mathbf{y}}_{t|t-1}', \quad \bar{\mathbf{y}}_{t|t-1} = \left(\sum_{j=1}^{t-1} b_{H,|j-t|} \right)^{-1} \sum_{k=1}^{t-1} b_{H,|t-k|} \mathbf{y}_k. \quad (26)$$

The optimal choice of γ is then given by $\hat{\gamma} = \arg \min_{\gamma} Q_{T,\gamma}^{(j)}$ for $j = 1$ or 2 . In (25), the regularized covariance $m_{\gamma}(\hat{\Sigma}_t)$ matrix is constrained to be positive definite. If $E\mathbf{y}_t = 0$ (or $E\mathbf{x}_t = 0$ in (17)), then \mathbf{y}_t^0 can be replaced by \mathbf{y}_t . For full sample methods, $m_{\gamma}(\Sigma)$, that do not account for time variation (i.e. no H parameter), to select γ , we use the same objective functions with $\hat{\Sigma}_{|t-1}$ computed setting $b_{H,|j-t|} = 1$ in (26).

The choice of the objective function depends on the empirical application at hand, while in different situations, the objectives (24) and (25) can give slightly different estimation results. In practice, one can use the (25) for choices of γ that yield a positive definite covariance matrix $m_{\gamma}(\hat{\Sigma}_t)$, while dropping all other alternatives, that do not guarantee this property. To ensure this property, the eigenvalues of $m_{\gamma}(\hat{\Sigma}_t)$ should always be positive. Giraitis, Kapetanios, and Price (2013) have shown that cross-validation methods can return tuning parameters that optimize the forecasting mean squared error in time-varying settings.

All tuning parameters are calibrated over a sensible parameter space. For the time varying estimator $\hat{\Sigma}_t$ we search for tuning parameter $H = T^h$, with h on the interval $[0.45, 1]$. For the thresholding methods, the parameter space depends on the data at hand. For instance, κ depends on the individual entries of the $\hat{\Sigma}_t$ and the specific thresholding approach that is considered. To this end, we focus on values of $\kappa \in [\kappa_{\min}, \kappa_{\max}]$, where $0 < \kappa_{\min}$, and κ_{\max} is the minimum value that implies $m_{\gamma}(\hat{\Sigma}_t) \equiv \text{diag}(\hat{\Sigma}_t)$ for $H \in [T^{0.45}, T]$. In the adaptive thresholding approach, as suggested by Cai and Liu, we search for δ on $[0, 12]$. In this case, one theoretically reasonable value is $\delta_{opt} = 2$. This result relies on the fact the $\hat{\theta}_{ij}$ (see (22)) provides an estimate of the variability of each individual entry of $\hat{\Sigma}_t$. The same parameter space is used for the POET estimator. For the Ledoit and Wolf methods, we search for ρ in the interval $[0, 1]$. We look for optimal values using a grid of 20 points for h , 100 points for κ , 50 points for δ , and 20 points for ρ in all the above parameter spaces. In the simulation experiments we use $Q_{T,\gamma}^{(1)}$ although results do not alter significantly when we use the $Q_{T,\gamma}^{(2)}$.

3.2 Monte Carlo experiments

Following the theoretical exposition, we carry out a Monte Carlo study to explore the properties of estimation and forecasting of large deterministic and stochastic time-varying covariance matrices. We compare the performance of the estimates $\hat{\Sigma}$, $\hat{\Sigma}_t$, Ledoit-Wolf shrinkage estimates $\hat{\Sigma}_{LW}$, $\hat{\Sigma}_{LW,t}$ and Bickel-Levina thresholding estimates $T_{\lambda}(\hat{\Sigma})$, $T_{\lambda}(\hat{\Sigma}_t)$, $T_{\lambda_{opt}}(\hat{\Sigma}_t)$, Cai-Liu thresholding estimates $T_{\lambda_{ij}}(\hat{\Sigma})$ and POET estimators $\hat{\Sigma}_{poet}$.

Monte Carlo design. The generation of Σ_t presents particular challenges. We wish to have a reasonably realistic generation mechanism for Σ_t that corresponds to economic or financial data. Previous work on large dimensional covariance matrices of stationary processes offers

little guidance since its designs for Σ with no time variation are rather simplistic for our purposes.

We choose to have a design for Σ_t based on the one factor setting used in Bailey, Kapetanios, and Pesaran (2016). It allows for varying degrees of sparsity $n_p \leq p$. We consider the data generating model

$$\mathbf{y}_t = (\Sigma_t)^{1/2} \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T \quad (27)$$

with two settings for Σ_t : deterministic and random. We consider two different settings for $\boldsymbol{\varepsilon}_t$: in the first one $\boldsymbol{\varepsilon}_t$ is an i.i.d. p -dimensional standard normal variable while in the second $\boldsymbol{\varepsilon}_t$ is an i.i.d. p -dimensional random variable where each element follows a Student- $t(12)$ distribution with 12 degrees of freedom. As we consider relatively large values of p compared to T , we feel that our choice of the number of degrees of freedom is reasonable. Denote

$$\mathbf{V}_t = \text{diag}(\mathbf{e}_t) + \mathbf{b}_t \mathbf{b}_t' = \{v_{ij,t}\}$$

where $\mathbf{b}_t = (b_{1t}, b_{2t}, \dots, b_{n_p t}, 0, \dots, 0)'$, $\mathbf{e}_t = (e_{1t}, e_{2t}, \dots, e_{pt})'$ are $p \times 1$ vectors. Then, we set $\Sigma_t = \{v_{ij,t}/(v_{ii,1}v_{jj,1})^{1/2}\}$. We assume $e_{it} = h_{it}d_i$ where b_{it} , e_{it} and d_i will be specified below. This simulation design can be seen as a one factor model with time-varying factor loadings. Stock and Watson (2002a) argue that macroeconomic and/or financial time series do possess small instabilities, which can be amplified when these span a significant long time period, and the proposed simulation scheme is consistent with this idea.

We set $T = 400$ and consider three cases: $p = 10$ with sparsity parameter 3, 5, 10, $p = 50$ with $n_p = 5, 20, 50$, and $p = 100$ with $n_p = 10, 40, 100$, respectively.

Deterministic Σ_t . We consider two settings for deterministic time-varying covariance matrices Σ_t generated respectively by

$$b_{it} = 4 + 10(t/T), \quad h_{it} = 10 + 25(t/T), \quad (28)$$

$$b_{it} = 4 + 2 \sin(2\pi t/T)(1 + 2t/T), \quad h_{it} = 10 + 2 \sin(2\pi t/T)(1 + 2t/T). \quad (29)$$

In these settings, we set $d_i = 2$, $i = 1, \dots, p$. The first one is a linear trend, while the second is a sine function of time. All settings provide considerable, smooth change for Σ_t over time.

Random Σ_t . We consider one setting for a stochastic time-varying Σ_t which is generated by the following rescaled unit root processes:

$$b_{it} = (2.4|u_{it}/\sqrt{t}| + 0.04)(1 + 2t/T), \quad h_{it} = (9|u_{it}/\sqrt{t}| + 16)(1 + 2t/T), \quad (30)$$

where $(u_i.)$ are independent random walk processes: $u_{it} = u_{i,t-1} + \xi_{it}$ and ξ_{it} are i.i.d. $N(0, 1)$. In both settings, we assume that d_i are i.i.d. χ_2^2 random variables, and in (27) we assume (Σ_t) to be independent of $(\boldsymbol{\varepsilon}_t)$.

Finally, and for reference, we also consider a case where the covariance matrix Σ_t is actually constant and equal to the identity matrix. Notice that all the above considered simulation designs are in accordance with our theoretical formulations. For example the processes defined in (30) involve scaled random walk processes which satisfy Assumption H. Similar arguments can be used to verify that all the simulated processes we consider satisfy our theoretical assumptions. As expected, in such cases, forecast methods that allow for time-varying Σ_t outperform forecast methods based on non-time varying estimates of Σ .

3.3 Monte Carlo results

In this section we discuss performance of the one-step-ahead out-of-sample forecasts of Σ_T for various estimation methods, $m_\gamma(\Sigma_t)$, of a time-varying covariance matrix Σ_t using several Monte Carlo experiments.

For the forecasting experiment, given the sample \mathbf{y}_t , $t = 1, \dots, T$, we use the estimate of Σ_T as the forecast for Σ_{T+1} . Formally, we define the one-step-ahead out-of-sample forecast of Σ_{T+1} based on estimation method $m_\gamma(\widehat{\Sigma}_t)$ as $\widehat{\Sigma}_{T+1|T} = m_\gamma(\widehat{\Sigma}_{T,T})$. In our Monte Carlo experiments, we evaluate the forecast error using the Frobenius norm

$$\|\widehat{\Sigma}_{T|T-1} - \Sigma_T\|_F \quad (31)$$

and denote by $\text{frmse}(m_\gamma)$ its Monte Carlo average over 500 replications. To compare the quality of forecasts based on different estimation methods, as a benchmark, we choose the standard sample covariance estimate $m_{bench} = \widehat{\Sigma}$ which does not account for time variation. We report the relative frmse, $\text{frmse}(m_\gamma)/\text{frmse}(m_{bench})$. The smaller the latter is, the better is the performance of the method. Since in-sample forecasting of Σ_t by $\widehat{\Sigma}_{t|t-1}$ reduces to estimation of Σ_{t-1} by $m_\gamma(\widehat{\Sigma}_{t-1,t-1})$ using data \mathbf{y}_t , $t = 1, \dots, t-1$, parameter γ can be selected using cross-validation.

For each replication the tuning parameter γ in (31) is estimated by cross-validation method using objective function (24) over the last 24 observations ($T_a = T - 24$ in (24)). We denote by subscript "cv" tuning parameters γ_{cv} obtained by cross-validation, set $H_{opt} = T^{2/3}$ and $H_{opt} = T^{1/2}$ in estimation of a deterministic and stochastic Σ_t , respectively, whereas K_{opt} is as in section 3, and ρ_{opt} as above.

Tables 1-7 present simulation results on the out-of-sample forecast error of large covariance matrices for various models of Σ_t and its estimation methods. They report the average relative $\text{frmse}(m_\gamma)/\text{frmse}(m_{bench})$, over 500 replications. Smaller numbers indicate superior performance. The best performing method is bolded and the second best is underlined.

Table 1: Relative frmse of one step ahead forecasts for deterministic time varying Σ_t . $T = 400$, Σ_t generated by (29): sine function with a drift, $\varepsilon_t \text{ iid } N(0, 1)$

"dimension p "		10	10	10	50	50	50	100	100	100
"sparsity n_p "		3	5	10	5	20	50	10	40	100
Method	Tuning prmt	Relative frmse								
Forecasts based on non time-varying estimates of Σ										
Bickel-Levina	κ_{cv}	1	1	1.01	0.97	1	1	0.97	1	1
Cai-Liu	$\delta = 2$	0.99	1	1	0.94	0.99	1	0.93	0.99	1
	δ_{cv}	1	1	1	0.97	1	1	0.97	1	1
POET	κ_{cv}, K_{opt}	1	0.95	0.96	0.94	0.98	0.99	0.93	0.99	1
Ledoit-Wolf	ρ_{opt}	1.03	1.02	1.01	1.09	1.02	1.01	1.11	1.02	1.01
	ρ_{cv}	1.01	1	1.01	1	1	1.01	1	1	1
Forecasts based on time-varying estimates of Σ_t										
$\widehat{\Sigma}_t$	H_{cv}	<u>0.83</u>	<u>0.71</u>	0.62	1.36	<u>0.78</u>	0.61	1.41	<u>0.79</u>	0.59
Tv-Ledoit-Wolf	H_{cv}, ρ_{cv}	0.86	0.72	<u>0.62</u>	1.33	0.78	0.62	1.41	0.79	<u>0.59</u>
Tv-Bickel-Levina	κ_{cv}, H_{opt}	0.71	0.69	0.82	0.81	0.77	<u>0.61</u>	0.81	0.88	0.89
	κ_{cv}, H_{opt}	0.83	0.85	0.85	<u>0.89</u>	0.86	0.85	<u>0.91</u>	0.87	0.95

Table 2: Relative frmse of one step ahead forecasts for deterministic time varying Σ_t . $T = 400$, Σ_t generated by (28): trend function, $\varepsilon_t \text{ iid } N(0, 1)$

"dimension p "		10	10	10	50	50	50	100	100	100
"sparsity n_p "		3	5	10	5	20	50	10	40	100
Method	Tuning prmt	Relative frmse								
Forecasts based on non time-varying estimates of Σ										
Bickel-Levina	κ_{cv}	1	1	1	0.97	1	1	0.96	1	1
Cai-Liu	$\delta = 2$	0.99	1	1	0.94	0.99	1	0.93	0.99	1
	δ_{cv}	1	1	1	0.97	1	1	0.96	1	1
POET	κ_{cv}, K_{opt}	1	0.95	0.97	0.94	0.98	0.99	0.93	0.99	1
Ledoit-Wolf	ρ_{opt}	1.03	1.02	1.01	1.09	1.02	1.01	1.11	1.02	1.01
	ρ_{cv}	1.01	1.01	1.01	1	1	1.01	1	1	1.01
Forecasts based on time-varying estimates of Σ_t										
$\widehat{\Sigma}_t$	H_{cv}	0.62	0.57	<u>0.5</u>	1.07	0.6	<u>0.5</u>	1.11	0.6	<u>0.5</u>
Tv-Ledoit-Wolf	ρ_{cv}, H_{cv}	0.68	0.61	0.52	1.09	0.61	0.51	1.12	0.61	0.52
Tv-Bickel-Levina	κ_{cv}, H_{cv}	<u>0.55</u>	<u>0.55</u>	0.59	<u>0.59</u>	0.58	0.6	<u>0.52</u>	<u>0.55</u>	0.59
	κ_{cv}, H_{opt}	0.53	0.47	0.49	0.57	0.49	0.5	0.51	0.43	0.44

Table 3: Relative frmse of one step ahead forecasts for stochastic time varying Σ_t . $T = 400$, Σ_t stochastic generated by (30), $\varepsilon_t \text{ iid } N(0, 1)$

"dimension p "		10	10	10	50	50	50	100	100	100
"sparsity n_p "		3	5	10	5	20	50	10	40	100
Method	Tuning prmt	Relative frmse								
Forecasts based on non time-varying estimates of Σ										
Bickel-Levina	κ_{cv}	1	1	1.01	0.96	1	1	0.95	1	1
Cai-Liu	$\delta = 2$	0.99	1	1.01	0.93	1.01	1.03	0.91	1.02	1.03
	δ_{cv}	1	1	1.01	0.96	1	1	0.95	1	1
POET	κ_{cv}, K_{opt}	1.01	0.99	0.98	0.97	0.99	0.99	0.93	0.99	1
Ledoit-Wolf	ρ_{opt}	1.02	1.01	1.01	1.01	1.02	1.01	1.02	1.02	1.01
	ρ_{cv}	1.01	1	1	0.99	0.99	1	1	1	1
Forecasts based on time-varying estimates of Σ_t										
$\widehat{\Sigma}_t$	H_{cv}	0.66	<u>0.5</u>	0.58	1.13	<u>0.61</u>	0.63	1.23	<u>0.74</u>	0.54
Tv-Ledoit-Wolf	ρ_{cv}, H_{cv}	0.66	0.52	<u>0.6</u>	0.98	0.61	<u>0.64</u>	1.12	0.75	<u>0.54</u>
Tv-Bickel-Levina	κ_{cv}, H_{cv}	0.61	0.5	0.63	0.70	0.60	0.74	0.77	0.7	0.8
	κ_{cv}, H_{opt}	<u>0.65</u>	0.48	0.66	<u>0.75</u>	0.63	0.71	<u>0.79</u>	0.78	0.81

Tables 1-6 provide comparisons of the quality of out-of-sample forecasting of a time-varying covariance matrix Σ_t based on the sample covariances $\widehat{\Sigma}$, $\widehat{\Sigma}_t$, and their regularized versions by Ledoit-Wolf shrinkage and Bickel-Levina thresholding methods. For the latter, the threshold λ is adapted to time variation via bandwidth H as described in (15) and (20), while tuning parameters are selected by cross-validation as described in Section 3.1. For comparison reasons we also include the Cai-Liu adaptive thresholding estimator and the POET estimator of Σ_t . We use cross-validation to choose the tuning parameter, δ , in Cai-Liu adaptive thresholding estimator, as well as the theoretically optimal value ($\delta = 2$). In POET estimator we use information criteria to choose the optimal number of factors K denoted by K_{opt} .

Monte Carlo results, for deterministic Σ_t , reported in Tables 1-2, 4-5 and for stochastic Σ_t , reported in Tables 3, 6 clearly show what to expect from the considered methods. In general, they indicate a significant impact from deterministic or stochastic change of Σ_t on the quality of forecasting of Σ_t by various methods.

A number of interesting conclusions can be drawn from the tables. It is unlikely that the time invariant sample covariance estimate $\widehat{\Sigma}$ will produce good forecasts of Σ_t for all t . We use it as a benchmark. Clearly, estimation and forecasting need to be adapted to both time variation and sparsity of Σ_t . The relative forecasting error, reported in the tables shows

Table 4: Relative frmse of one step ahead forecasts for deterministic time varying Σ_t . $T = 400$, Σ_t generated by (29): sine function with a drift, $\varepsilon_t \text{ iid } t(12)$

"dimension p "		10	10	10	50	50	50	100	100	100
"sparsity n_p "		3	5	10	5	20	50	10	40	100
Method	Tuning prmt	Relative frmse								
Forecasts based on non time-varying estimates of Σ										
Bickel-Levina	κ_{cv}	0.99	1	1.03	0.93	1	1	0.91	0.99	1.02
Cai-Liu	$\delta = 2$	0.98	0.99	1	0.88	0.99	1	0.87	0.99	1
	δ_{cv}	0.99	1	1	0.94	1	1	0.92	0.99	1
POET	κ_{cv}, K_{opt}	1	0.93	0.95	0.89	0.97	0.99	0.88	0.98	0.99
Ledoit-Wolf	ρ_{opt}	1.04	1.03	1.01	1.12	1.03	1.01	1.13	1.03	1.01
	ρ_{cv}	1	1	1.03	1	1.01	1	1	1	1.02
Forecasts based on time-varying estimates of Σ_t										
$\widehat{\Sigma}_t$	H_{cv}	0.8	0.69	0.52	1.27	0.71	<u>0.54</u>	1.07	0.67	<u>0.5</u>
Tv-Ledoit-Wolf	ρ_{cv}, H_{cv}	0.79	0.7	0.58	1.26	0.72	0.56	1.07	0.69	0.54
Tv-Bickel-Levina	κ_{cv}, H_{cv}	<u>0.71</u>	<u>0.65</u>	0.61	<u>0.76</u>	<u>0.61</u>	0.55	<u>0.59</u>	<u>0.55</u>	0.53
	κ_{cv}, H_{opt}	0.62	0.56	<u>0.57</u>	0.74	0.58	0.53	0.57	0.49	0.48

Table 5: Relative frmse of one step ahead forecasts for deterministic time varying Σ_t . $T = 400$, Σ_t generated by (28): trend function, $\varepsilon_t \text{ iid } t(12)$

"dimension p "		10	10	10	50	50	50	100	100	100
"sparsity n_p "		3	5	10	5	20	50	10	40	100
Method	Tuning prmt	Relative frmse								
Forecasts based on non time-varying estimates of Σ										
Bickel-Levina	κ_{cv}	0.99	1	1	0.93	1	1	0.93	0.99	1
Cai-Liu	$\delta = 2$	0.98	0.99	1	0.89	0.99	1	0.88	0.99	1
	δ_{cv}	0.99	1	1	0.93	1	1	0.93	1	1
POET	κ_{cv}, K_{opt}	1	0.93	0.95	<u>0.89</u>	0.97	0.99	<u>0.88</u>	0.98	0.99
Ledoit-Wolf	ρ_{opt}	1.04	1.02	1.01	1.12	1.03	1.01	1.14	1.03	1.01
	ρ_{cv}	1	1	1.01	1	1	1.01	1	1	1
Forecasts based on time-varying estimates of Σ_t										
$\widehat{\Sigma}_t$	H_{cv}	1	0.88	0.78	1.18	0.94	0.8	1.15	0.96	0.74
Tv-Ledoit-Wolf	ρ_{cv}, H_{cv}	1.01	0.88	0.79	1.17	0.95	0.82	1.14	0.97	0.75
Tv-Bickel-Levina	κ_{cv}, H_{cv}	<u>0.95</u>	<u>0.83</u>	<u>0.78</u>	0.92	<u>0.87</u>	<u>0.8</u>	0.95	<u>0.86</u>	<u>0.74</u>
	κ_{cv}, H_{opt}	0.86	0.78	0.82	0.95	0.85	0.8	0.89	0.84	0.78

Table 6: Relative frmse of one step ahead forecasts for stochastic time varying Σ_t . $T = 400$, Σ_t stochastic generated by (30), $\varepsilon_t \text{ iid } t(12)$

"dimension p "	10	10	10	50	50	50	100	100	100	
"sparsity n_p "	3	5	10	5	20	50	10	40	100	
Method	Relative frmse									
Forecasts based on non time-varying estimates of Σ										
Bickel-Levina	κ_{cv}	0.99	1.02	1.01	0.93	1.02	1	0.97	1	1
Cai-Liu	$\delta = 2$	<u>0.97</u>	1	1.02	<u>0.88</u>	1.02	1.02	0.93	1.03	1.05
	δ_{cv}	0.98	1.01	1.01	0.93	1.01	1	0.96	1	1
POET	κ_{cv}, K_{opt}	1.03	0.99	0.97	0.95	0.97	1	0.96	0.98	1
Ledoit-Wolf	ρ_{opt}	1.02	1.01	1.02	0.98	1.03	1.01	1.01	1.03	1.01
	ρ_{cv}	1	1.03	1	1	1.01	0.99	0.99	1	1
Forecasts based on time-varying estimates of Σ_t										
$\widehat{\Sigma}_t$	H_{cv}	1.01	0.73	<u>0.67</u>	1.03	<u>0.9</u>	<u>0.57</u>	0.85	<u>0.84</u>	<u>0.71</u>
Tv-Ledoit-Wolf	ρ_{cv}, H_{cv}	1.01	0.76	0.66	1.04	0.9	0.56	0.84	0.84	0.69
Tv-Bickel-Levina	κ_{cv}, H_{cv}	0.93	<u>0.73</u>	0.73	0.77	0.88	0.59	0.65	0.81	0.72
	κ_{cv}, H_{opt}	1.19	0.79	0.82	0.9	1.04	0.58	<u>0.7</u>	0.88	0.73

Table 7: Relative frmse of one step ahead forecasts for fixed deterministic Σ . $T = 400$, $\Sigma = I$, $\varepsilon_t \text{ iid } t(12)$

"dimension p "	10	50	100	
"sparsity n_p "	1	1	1	
Method	Tuning prmt	Relative frmse		
Forecasts based on non time-varying estimates of Σ				
Bickel-Levina	κ_{cv}	0.9	0.7	0.61
Cai-Liu	$\delta = 2$	<u>0.78</u>	<u>0.47</u>	<u>0.35</u>
	δ_{cv}	0.88	0.68	0.58
POET	κ_{cv}, K_{opt}	1.16	0.86	0.71
Ledoit-Wolf	ρ_{opt}	0.7	0.42	0.31
	ρ_{cv}	0.98	1	1
Forecasts based on time-varying estimates of Σ_t				
$\widehat{\Sigma}_t$	H_{cv}	1.29	1.15	1.08
Tv-Ledoit-Wolf	ρ_{cv}, H_{cv}	1.02	1.09	1.05
Tv-Bickel-Levina	κ_{cv}, H_{cv}	1.05	0.72	0.58
	κ_{cv}, H_{opt}	1.18	0.87	0.68

that regularization of the sample covariance estimate $\widehat{\Sigma}$ by Ledoit-Wolf shrinkage, Bickel-Levina, Cai-Liu and POET thresholding methods is not effective. However, the time varying sample covariance estimate $\widehat{\Sigma}_t$ seems to be able to adapt to time variation effectively: in some cases it significantly improves the quality of forecasting of Σ_t . Regularization of $\widehat{\Sigma}_t$ by Ledoit-Wolf shrinkage does not materially improve forecasting quality, while regularization of $\widehat{\Sigma}_t$ by Bickel-Levina thresholding method significantly improves it in all experiments. As expected, improvement tends to be *slightly* stronger when Σ_t is sparse, however, the effect of time variation in Σ_t on forecasting quality is more severe than that of sparsity.

The thresholding procedure $T_{\lambda_{opt}}(\widehat{\Sigma}_t)$ with λ_{opt} based on the theoretically optimal H_{opt} and cross-validated κ and thresholding procedure $T_{\lambda}(\widehat{\Sigma}_t)$ where both κ and H are chosen using cross-validation, produce similar improvements of the forecasting quality. Hence, the choice of theoretically optimal H_{opt} given in Theorem 2 and Theorem 4, is a decent alternative to cross-validated H and the $T_{\lambda_{opt}}(\widehat{\Sigma}_t)$ thresholding method can be recommended for empirical work.

Finally, in Table 7, we consider a base case scenario in which the true covariance matrix $\Sigma = I$ is the identity matrix and does not vary with t . Now, the sample variance estimate $\widehat{\Sigma}$ performs better, especially for small p . The methods that do not account for time variation, outperform the time varying methods, as expected, while the Ledoit-Wolf shrinkage estimator is the best performing method in all experiments. This is not surprising given that the Ledoit-Wolf method actually shrinks the estimate towards the identity matrix. As expected, our proposed time-varying forecasting methods suffer somewhat when p is small, while when p increases, they can deliver benefits over sample variance estimate $\widehat{\Sigma}$. This indicates that the sparsity feature can be captured over the methods in this case too.

3.4 Empirical application

In this section we present an empirical demonstration of the potential gains from using the proposed covariance matrix estimators to design minimum variance portfolios. The illustration is motivated by the availability of large datasets on stock returns, and the increased demand for portfolios with lower risk exposure.

The literature on portfolio allocation is well grounded on the mean variance efficient portfolio frontier proposed by Markowitz (1952). The minimum variance portfolio offers a suitable device for examining the possible superiority of portfolio weights based on the proposed large time-varying covariance estimators, as it neutralizes the impact of the estimated expected return parameters, focusing solely on the covariance matrix estimation (see e.g. Jagannathan and Ma (2003)).

Noting that Σ_t is the $p \times p$ covariance matrix of p -dimensional vector of returns, collected in \mathbf{y}_t , the minimum variance portfolio is designed to minimize investors' exposure to risk, regardless of preferences. The vector of optimal weights given by

$$w_t(\Sigma_t) = \frac{\Sigma_t^{-1} \mathbf{1}_p}{\mathbf{1}_p' \Sigma_t^{-1} \mathbf{1}_p},$$

where $\mathbf{1}_p$ is a p -dimensional vector of ones. In practice, one has to choose an estimator for Σ_t and routinely compute the minimum variance weights. In our analysis, we compute portfolio weights, employing our large covariance matrix estimators, i.e. $\hat{w}_t = w_t(\hat{\Sigma}_t)$.

Since our proposals involve sparse covariance estimators, we choose our sample to accommodate this assumption as much as possible. To this end, we obtained data from the Center for Research in Security Prices database and focus on daily returns from 10 different industry sectors. The sectors are: Consumer NonDurables, Utilities, Healthcare-Medical Equipment-Drugs, Telephone-Television Transmission, Business Equipment, Oil-Gas-Coal Extraction, Manufacturing, Consumer Durables, Consumer NonDurables, Other. The smallest portfolio examined is a 10 stocks portfolio, comprised of 1 stock from each of the 10 sectors. The 20 stocks portfolio is comprised of 2 stocks from each of the 10 sectors and so forth until the 100 stocks portfolio which is the largest portfolio examined. With this selection of stocks, our aim is to accommodate a block diagonal structure for the large covariance matrix of returns. The sample period starts at 10-Jun-2005 and ends at 23-Aug-2019 ($T = 3967$ daily observations).

To examine the out of sample risk performance, we will proceed as follows: every five days portfolio weights are selected as a function of the large covariance matrix estimate $m_\gamma(\hat{\Sigma}_t)$. These are kept constant for the subsequent 5 trading days while the corresponding portfolio returns are computed as $\mathbf{y}_s^{port} = \sum_{i=1}^p \hat{w}_{i,s} y_{i,s}$ for $s = t + 1, \dots, t + 5$. Then the large covariance matrix is reestimated and the portfolio weights are updated (see Table 8 Panels A, B, C), or it is not reestimated and the portfolio weights remain unchanged (see Table 8 Panel D). The above steps are repeated until the end of the sample, i.e. for the last 775 trading days of our sample (period from 28-Jul-2016 until 23-Aug-2019). At the end, we use the 775 out of sample portfolio returns \mathbf{y}_s^{port} , $s = 1, \dots, 775$, to compute the associated variance

$$PortfolioVariance = var(\widehat{y^{port}}).$$

The method that provides the minimum out of sample portfolio variance is considered as the best performing one.

To accommodate our minimum risk objective we modify the cross-validation criteria developed in Section 3.1 as follows: we choose tuning parameter(s) γ , used to compute \hat{w}_t , that

minimize the objective function

$$Q_{t,\gamma}^{emp} := \frac{1}{80} \sum_{s=t-79}^t (\widehat{w}_s \mathbf{y}_s - \frac{1}{80} \sum_{s=t-79}^t \widehat{w}_s \mathbf{y}_s)^2$$

where \mathbf{y}_t is the p -vector of portfolio returns at time t . The tuning parameter ranges and positive definiteness constraints discussed in Section 3.1 also apply here.

To account for non sparse covariance matrices, we further extend our proposals following the approach developed in Fan, Liao, and Mincheva (2013). To this end, we first extract a number of factors from the portfolio returns and then regularize the remaining part by our proposed estimator. More formally, we consider the approximate factor model, for returns \mathbf{y}_t ,

$$\mathbf{y}_t = \mathbf{b}\mathbf{f}_t + \mathbf{u}_t$$

where \mathbf{b} is the $p \times K$ matrix of factor loadings, \mathbf{f}_t is the $K \times 1$ vector of unobserved factors, with $K \ll p$, and \mathbf{u}_t is the idiosyncratic error. We can then conclude that

$$\boldsymbol{\Sigma}_t = \mathbf{b}'\boldsymbol{\Sigma}_f\mathbf{b} + \boldsymbol{\Sigma}_{u,t} \quad (32)$$

where $\boldsymbol{\Sigma}_f$, $\boldsymbol{\Sigma}_{u,t}$ are defined as $\boldsymbol{\Sigma}_f = \text{var}(\mathbf{f}_t)$, and $\boldsymbol{\Sigma}_{u,t} = \text{var}(\mathbf{u}_t)$. Assuming that $\boldsymbol{\Sigma}_{u,t}$ is sparse and possesses significant time variation, a natural estimate of (32) is

$$\widehat{\boldsymbol{\Sigma}}_t = \widehat{\mathbf{b}}'\widehat{\boldsymbol{\Sigma}}_f\widehat{\mathbf{b}} + T_\lambda(\widehat{\boldsymbol{\Sigma}}_{u,t}) \quad (33)$$

with $\widehat{\mathbf{b}}$, $\widehat{\mathbf{f}}_t$ estimated by PCA, $\widehat{\boldsymbol{\Sigma}}_f = T^{-1}\widehat{\mathbf{f}}\widehat{\mathbf{f}}'$, and $T_\lambda(\widehat{\boldsymbol{\Sigma}}_{u,t})$ is the time varying regularized large covariance estimator of the residuals $\widehat{\mathbf{u}}_t = \mathbf{y}_t - \widehat{\mathbf{b}}\widehat{\mathbf{f}}_t$. In practice, the number of factors K can be chosen by information criteria (see e.g. Bai and Ng (2002)) or set equal to a fixed number. In our empirical exercise we estimate K according to

$$\widehat{K} = \arg \min_{0 \leq K \leq M} \left\{ \ln \left(\frac{1}{p} \sum_{i=1}^p \widehat{\sigma}_i^2 \right) + K \left(\frac{p+T}{pT} \right) \ln \left(\frac{pT}{p+T} \right) \right\} \quad (34)$$

where $\widehat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^T \widehat{\mathbf{u}}_{it}^2$, and $\widehat{\mathbf{u}}_{it}$ is the i -th element of \mathbf{u}_t .

In addition to all the above mentioned models we consider two more methods for estimating the large covariance matrices. The first model is due to Ledoit and Wolf (2015) who extend the previously discussed linear shrinkage estimator to the non linear case. To this end, let u be an eigenvector of $\boldsymbol{\Sigma} = \text{var}(\mathbf{y}_t)$ and $\widehat{\boldsymbol{\Sigma}}$ be an associated estimate of $\boldsymbol{\Sigma}$. The corresponding sample eigenvalue is equal to $u'\widehat{\boldsymbol{\Sigma}}u$. Nonlinear shrinkage replaces this quantity with a consistent estimator of $u'\boldsymbol{\Sigma}u$. Recovering the population eigenvalues from the sample eigenvalues requires inverting the Marchenko-Pastur equation (see Theorem 1 of Marchenko and Pastur (1968)).

Table 8: Performance of out of sample minimum variance portfolio forecast

"dimension p "	10	20	30	40	50	70	100	
"dimension T "	3967	3967	3967	3967	3967	3967	3967	
Method	Tuning prmt	Relative Variance						
Panel A: Forecasts based on non time-varying estimates of Σ (with reestimation)								
1/p		1.166	1.636	1.486	1.572	2.175	1.993	1.76
rolling 6mth		0.96	1.075	0.927	0.931	0.82	1.014	1.953
rolling 12mth		0.893	0.952	0.812	0.823	0.654	0.658	0.73
Bickel-Levina	κ_{cv}	0.934	0.978	0.975	0.961	1.047	1.014	1.019
Cai-Liu	$\kappa = 2$	1	0.999	0.993	1.01	1.041	1.196	1.238
	κ_{cv}	0.932	0.984	0.971	0.96	0.986	0.982	1.03
POET	κ_{cv}, K_{opt}	1.011	1.012	1.006	1.016	1.014	1.019	1.011
Ledoit-Wolf	ρ_{opt}	0.993	0.978	0.982	0.992	1.012	1.009	1.007
	ρ_{cv}	0.929	0.925	0.843	0.893	1.023	0.989	0.969
Ledoit-Wolf(non linear)		1	0.995	0.993	0.991	0.999	1.002	0.998
Panel B: Forecasts based on time-varying estimates of Σ_t								
TV	H_{cv}	0.919	0.959	0.814	0.746	0.625	0.637	0.771
TV-LW	H_{cv}, ρ_{cv}	0.879	0.959	0.822	0.799	0.659	0.619	0.599
Tv-Bickel-Levina ^S	κ_{cv}, H_{cv}	0.877	0.981	0.877	0.831	0.616	0.605	0.597
	κ_{cv}, H_{opt}	0.936	1.073	0.994	0.929	0.829	0.958	0.951
Tv-Bickel-Levina-Factor ^S	κ_{cv}, H_{cv}	1.026	1.144	1.057	1.041	0.924	0.844	0.824
	κ_{cv}, H_{opt}	1.397	1.301	1.232	1.182	0.992	0.979	0.947
Tv-Bickel-Levina ^D	κ_{cv}, H_{cv}	0.87	0.987	0.87	0.832	0.61	0.604	0.597
	κ_{cv}, H_{opt}	0.884	0.899	0.755	0.789	0.648	0.641	0.634
Tv-Bickel-Levina-Factor ^D	κ_{cv}, H_{cv}	1.031	1.139	1.046	1.056	0.916	0.846	0.835
	κ_{cv}, H_{opt}	1.279	1.36	1.108	1.144	0.95	0.926	0.905
Panel C: Forecasts based on conditional covariance matrix estimates of $\Sigma_{t t-1}$								
RARCH		0.92	0.909	0.801	0.8	0.726	0.994	1
Panel D: Forecasts based on non time-varying estimates of Σ (no reestimation)								
rolling 6mth		1.113	1.028	0.929	1.142	1.223	1.718	2.494
rolling 12mth		0.966	0.938	0.964	1.376	1.1	1.197	1.215
Bickel-Levina	κ_{cv}	1.027	1.038	1.036	1.146	1.075	1.096	1.091
Cai-Liu	$\kappa = 2$	1.027	1.037	1.039	1.068	1.109	1.562	1.576
	κ_{cv}	1.128	1.077	0.995	1.097	1.125	1.152	1.092
POET	κ_{cv}, K_{opt}	1.027	1.05	1.065	1.055	1.131	1.163	1.124
Ledoit-Wolf	ρ_{opt}	1.018	1.008	1.029	1.047	1.09	1.104	1.098
	ρ_{cv}	0.948	0.942	0.833	0.894	1.192	1.109	1.038
Ledoit-Wolf(non linear)		1.028	1.032	1.042	1.049	1.078	1.101	1.09

Notes: Out of sample portfolio variance relative to the performance of the portfolio derived from the full sample covariance estimate. The method that provides the minimum out of sample risk exposure is marked in bold(blue). The second best is underlined(red). The examined period starts at 10-Jun-2005 and ends at 23-Aug-2019 ($T = 3967$ daily observations), and the out of sample evaluation period covers the last 775 observations of our sample. Data correspond to daily returns from 10 different industry sectors. The superscript "S" in "Tv-Bickel-Levina" and "Tv-Bickel-Levina-Factor" methods indicate thresholding a stochastic Σ_t using λ given in (20) while the superscript "D" corresponds to thresholding a deterministic Σ_t using λ as in (15).

Ledoit and Wolf (2015) introduced an effective numerical method for inverting the Marchenko-Pastur equation.

The second model stems from the large conditional covariance modelling literature. Noureldin, Shephard, and Sheppard (2014) propose a model that allows estimation of flexible GARCH type dynamics in moderately large dimensions. They refer to it as the multivariate rotated ARCH model. The main idea of this approach is to undertake a transformation (in particular, a rotation) of the raw returns, and then use a BEKK-type parametrization of the time-varying covariance matrix. Inference is computationally attractive and based on the quasi-maximum likelihood (QML).

For the p -dimensional returns $\mathbf{y}_t = \Sigma^{1/2}\varepsilon_t$, the unconditional covariance Σ is decomposed as $\Sigma = P\Lambda P'$ where P is the matrix of eigenvectors and Λ is the matrix of non negative eigenvalues. Then, since $\varepsilon_t = P(\Lambda)^{-1/2}P'y_t$, with $\text{var}(\varepsilon_t) = I_p$, the conditional variance of ε_t can be modelled as a BEKK type parametrization (see Engle and Kroner (1995))

$$\text{var}(\varepsilon_t|\varepsilon_{t-1}) = G_{t|t-1} = (I_p - AA' - BB') + A\varepsilon_{t-1}\varepsilon'_{t-1}A' + BG_{t-1|t-1}B', \quad G_0 = I_d. \quad (35)$$

In our empirical application we apply the scalar specification of (35) that assumes $A = a^{1/2}I_p$ and $B = b^{1/2}I_p$.

Finally, we also include in our analysis a number of portfolios which are commonly considered in the portfolio selection literature. The first is the equally weighted portfolio $\hat{w}_t = \mathbf{1}/p$, where $\mathbf{1}$ is a vector of ones of dimension p . The second is the rolling sample, covariance matrix estimate based portfolio. We consider rolling sample estimates of size equal to 6 and 12 months.

In Table 8 we report out of sample portfolio variance results, as a ratio to the benchmark portfolio which is the full sample estimate, reestimated every 5 days. From Table 8 we can draw some important conclusions about the performance of our approaches. First, our study suggests that there are significant advantages from using the proposed time-varying covariance estimators to derive minimum variance portfolios. In all large portfolios considered, the proposed time-varying methodologies (see Panel B) outperform the full sample fixed ones (see Panel A and D). This is true when we estimate the fixed large covariance only once, using data up to 28-Jul-2016 (see Panel D) or when we reestimate them every 5 days in our out of sample evaluation period (see Panel A).

The unregularized time-varying estimator $\hat{\Sigma}_t$, seems to perform well, while its performance is affected by the size of the portfolio p as one would expect. For the proposed regularized estimator, the results remain satisfactory for both H selected by cross-validation and fixed H_{opt} , indicating the effectiveness of the derived theoretical value, H_{opt} . In this case the deterministic estimator seems to outperform the stochastic one, providing larger risk improvements.

When we adjust our approach for non sparse covariance matrices, through the factor decomposition mechanism given in (33), the designed portfolios perform better than the fixed estimators (Panel A and D), especially when p is large. Having said that, it is important to emphasize that non sparse versions of our proposals deteriorate, compared to their sparse analogues, indicating that the fixed component of estimator (33) captured by the factor part dominates the sparse time varying component, while the latter proves more important for minimum risk portfolios. On the other hand, time varying regularization through shrinkage (TV-LW), proves important as this method is among the top performing ones. Finally, comparing our methods with the large conditional covariance model RARCH (see Panel C), it becomes apparent that this outperforms the fixed estimators (Panel A and D), for small and medium p , while for $p = 100$ this becomes equivalent to the full sample estimate.

4 Exponential inequalities

This section contains new results on Bernstein type inequalities for (weighted) sums of random variables (r.v.'s) (ξ_j) that are dependent, unbounded and have thin- or heavy-tailed distributions. We suppose that (ξ_j) satisfies the following α -mixing assumption.

Assumption A. (ξ_j) is such that $\xi_j - E\xi_j$ is an α -mixing (but not necessarily stationary) sequence with the mixing coefficients α_k such that for some $c_* > 0$ and $0 < \phi < 1$,

$$\alpha_k \leq c_* \phi^k, \quad k \geq 1. \quad (36)$$

In, addition we assume that variables, (ξ_j) , have thin- or heavy-tailed distributions.

The notation $(\xi_j) \in \mathcal{E}(s)$, $s > 0$ denotes thin tails and means that for some $a > 0$,

$$\max_j E \exp(a|\xi_j|^s) < \infty. \quad (37)$$

The notation $(\xi_j) \in \mathcal{H}(\theta)$, $\theta > 2$ corresponds to heavy tails and means that

$$\max_j E|\xi_j|^\theta < \infty. \quad (38)$$

Exponential inequalities for sums of r.v.'s, ξ_j with thin- and heavy-tailed distributions will be stated respectively using functions

$$f_t(\gamma_1, \gamma_2, c, \zeta) = c_0 \left\{ \exp(-c_1 \zeta^{\gamma_1}) + \exp\left(-c_2 \left(\frac{\zeta \sqrt{t}}{\log^2 t}\right)^{\gamma_2}\right) \right\}, \quad \zeta > 0, \quad t \geq 2, \quad (39)$$

$$g_t(\gamma_1, \theta, c, \zeta) = c_0 \left\{ \exp(-c_1 \zeta^\gamma) + \zeta^{-\theta} t^{-(\theta/2-1)} \right\},$$

where $\gamma_1 > 0$, $\gamma_2 > 0$, $\theta > 2$ and non-negative constants $c = (c_0, c_1, c_2)$ do not depend on ζ , t . Throughout the paper, we denote $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$, while C stands for generic constants.

4.1 Exponential inequalities for unbounded variables

First, we establish Bernstein type inequalities for sums

$$S_T = T^{-1/2} \sum_{k=1}^T (\xi_k - E\xi_k)$$

of α -mixing and bounded variables ξ_k .

Lemma 1, (40), below significantly improves the bound for $P(S_T \geq \zeta)$ obtained in Theorem 3.5 of White and Wooldridge (1991). Its proof uses exponential inequalities for α -mixing bounded random variables obtained in Theorem 2 of Merlevede *et al.* (2009).

Lemma 1 *Let the sequence (ξ_j) of r.v.'s satisfy Assumption A. Then, for all $\zeta > 0$, $T \geq 2$,*

$$P(|S_T| \geq \zeta) \leq \begin{cases} f_T(2, \gamma, c, \zeta) & \text{if } (\xi_j) \in \mathcal{E}(s), s > 0, \\ g_T(2, \theta', c, \zeta) & \text{if } (\xi_j) \in \mathcal{H}(\theta), \theta > 2 \end{cases} \quad (40)$$

with $\gamma = s/(s+1)$ and for any $2 < \theta' < \theta$ where c does not depend on ζ, T .

4.2 Exponential inequalities for weighted variables

Next we obtain Bernstein type inequalities for sums

$$S_{T,t} := H^{-1/2} \sum_{k=1}^T b_{H,|t-k|} (\xi_k - E\xi_k) \quad (42)$$

of weighted α -mixing variables ξ_k with thin- or heavy-tailed distributions. The weights $b_{H,k}$ are defined in (12) and (13). Under the assumption (13), they satisfy

$$b_{H,k} \leq C(1 + (k/H)^\nu)^{-1}, \quad |b_{H,k} - b_{H,k+1}| \leq CH^{-1}(1 + (k/H)^\nu)^{-1} \quad (43)$$

with $\nu > 3$. So, in (42), r.v.'s. ξ_k are strongly downweighted when k is distant from t .

In the next lemma we obtain exponential inequalities for $P(|S_{T,t}| \geq \zeta)$.

Lemma 2 *Let (ξ_j) satisfy Assumption A and (43) holds.*

Then for all $\zeta > 0$, $1 \leq t < T$,

$$P(|S_{T,t}| \geq \zeta) \leq \begin{cases} f_H(2, \gamma, c, \zeta) & \text{if } (\xi_j) \in \mathcal{E}(s), s > 0, \\ g_H(2, \theta', c, \zeta) & \text{if } (\xi_j) \in \mathcal{H}(\theta), \theta > 2 \end{cases} \quad (44)$$

with $\gamma = s/(s+1)$ and for any $\theta' \in (2, \theta)$ where c does not depend on ζ, t, H, T .

To obtain exponential inequalities for sums

$$\tilde{S}_{T,t} := H^{-1/2} \sum_{k=1}^T b_{H,|t-k|}(\xi_k - E\xi_t),$$

when the variables ξ_k are centered by $E\xi_t$, write

$$\tilde{S}_{T,t} = S_{T,t} + r_{T,t}, \quad r_{T,t} := H^{-1/2} \sum_{k=1}^T b_{H,|t-k|}(E\xi_k - E\xi_t).$$

The next lemma provides bounds for $P(|\tilde{S}_{T,t}| \geq \zeta)$ and $|r_{T,t}|$.

Lemma 3 (a) For any $|\zeta| > 2|r_{T,t}|$, $P(|\tilde{S}_{T,t}| \geq \zeta) \leq P(|S_{T,t}| \geq \zeta/2)$.

(b) If $|E\xi_k - E\xi_t| \leq C|t-k|(t \vee k)^{-1}$ for $t, k = 1, 2, \dots$, then

$$|r_{T,t}| \leq C_1 H^{3/2} (H \vee t)^{-1}. \quad (46)$$

If $|E\xi_k - E\xi_t| \leq C|t-k|T^{-1}$ for $t, k = 1, 2, \dots, T$, then

$$|r_{T,t}| \leq C_1 H^{3/2} T^{-1}. \quad (47)$$

Constants C, C_1 do not depend on k, t, H, T .

Exponential inequalities for sums $S_{T,t}$ allow to establish a bound for $\max_{t=1, \dots, T} |S_{T,t}|$ which is useful in applications. A bound for $\max_{t=1, \dots, T} |\tilde{S}_{T,t}|$ can be obtained using

$$\max_t |\tilde{S}_{T,t}| \leq \max_t |S_{T,t}| + \max_t |r_{T,t}|.$$

Corollary 6 Let (ξ_j) satisfy Assumption A and (43) hold. Assume that $(\xi_j) \in \mathcal{E}(s)$, $s > 0$ or $(\xi_j) \in \mathcal{H}(\theta)$, $\theta > 2$.

(a) Then, for any sequence $1 \leq t = t_T \leq T$, as $T \rightarrow \infty$,

$$S_{T,t} = O_P(1). \quad (48)$$

(b) If in addition,

$$cT^\delta \leq H \leq T \quad \text{for some } c > 0, \delta > 0, \quad (49)$$

then for any $\varepsilon > 0$,

$$\max_{1 \leq t \leq T} |S_{T,t}| = \begin{cases} O_P(\log^{1/2} T) & \text{if } (\xi_j) \in \mathcal{E}(s), s > 0, \\ O_P(\log^{1/2} T + (TH)^{1/\theta} H^{\varepsilon-1/2}) & \text{if } (\xi_j) \in \mathcal{H}(\theta), \theta > 2. \end{cases} \quad (50)$$

Finally, we establish uniform bounds for sums of weighted variables $|\xi_k|$.

Denote

$$v_{T,t} = H^{-1} \sum_{k=1}^T b_{H,|t-k|} \left(\frac{|t-k|}{H} \right)^\nu |\xi_k|, \quad \text{where } \nu \in [0, 1].$$

Corollary 7 *Let (ξ_j) satisfy Assumption A and (43), (49) hold. Then,*

$$\max_{s=1, \dots, T} v_{T,s} = \begin{cases} O_P(1) & \text{if } (\xi_j) \in \mathcal{E}(s), s > 0, \\ O_P(1) & \text{if } (\xi_j) \in \mathcal{H}(\theta), \theta > 2 \text{ and } \delta > 1/(\theta - 1) \text{ in (49)}. \end{cases} \quad (52)$$

4.3 Exponential inequalities with random scaling

In this section we discuss exponential inequalities for sums

$$\begin{aligned} S_{T,t}^{(h)} &:= H^{-1/2} \sum_{k=1}^T b_{H,|t-k|} h_k (\xi_k - E\xi_k), \\ \tilde{S}_{T,t}^{(h)} &:= H^{-1/2} \sum_{k=1}^T b_{H,|t-k|} (h_k \xi_k - h_t E\xi_t), \end{aligned}$$

for products $h_k \xi_k$ of an α -mixing process (ξ_k) and a random scaling factor (“volatility”) h_k . Differently from ARCH models, where h_k is a stationary process, here h_k is a persistent (non-stationary) process satisfying Assumption B below. It introduces smoothness and thin tail distribution restrictions on h_t , see Example 5. We impose no restrictions on the dependence between (h_k) and (ξ_k) . We assume that (ξ_k) are α -mixing variables with thin- or heavy-tailed distributions.

Assumption B. (h_t) satisfies one of the following two assumptions:

$$|h_t - h_k| \leq \begin{cases} \left(\frac{|t-k|}{t \vee k} \right)^{1/2} \xi_{tk}, & t, k = 1, 2, \dots \\ \left(\frac{|t-k|}{T} \right)^{1/2} \xi_{tk}. & 1 \leq t, k \leq T \end{cases} \quad (54)$$

where (h_k) and (ξ_{tk}) are such that for some $\alpha > 0$,

$$(h_k) \in \mathcal{E}(\alpha), \quad (\xi_{tk}) \in \mathcal{E}(\alpha). \quad (56)$$

Set

$$d_{Ht} = \begin{cases} (t \vee H)^{1/2} H^{-1} & \text{if (54) holds,} \\ T^{1/2} H^{-1} & \text{if (55) holds.} \end{cases} \quad (57)$$

Denote $\zeta' = \zeta d_{Ht}$ for $\zeta > 0$. Notice that $\zeta \wedge \zeta' = (1 \wedge d_{Ht})\zeta$.

First we establish exponential inequalities for sums where variables $h_k \xi_k$ are centered by $h_k E\xi_k$.

Lemma 4 Let (ξ_j) satisfy Assumption A, (h_j) satisfy Assumption B with $\alpha > 0$ and (43) holds.

Then there exists $c > 0$ such that for all $\zeta > 0$, $1 \leq t, H \leq T$, $T > 1$,

$$P(|S_{T,t}^{(h)}| \geq \zeta) \leq \begin{cases} f_H(\gamma_1, \gamma_2, c, \zeta \wedge \zeta') & \text{if } (\xi_j) \in \mathcal{E}(s), s > 0, \\ g_H(\gamma_1, \theta', c, \zeta \wedge \zeta') & \text{if } (\xi_j) \in \mathcal{H}(\theta), \theta > 2, \end{cases} \quad (58)$$

$$(59)$$

with $\gamma_1 = 2\alpha/(2 + \alpha)$, $\gamma_2 = \alpha s/(\alpha + s + 1)$ and for any $\theta' \in (2, \theta)$.

The same exponential inequalities hold for $\tilde{S}_{T,t}^{(h)}$, when $h_k \xi_k$ are centered by $h_t E \xi_t$. For that we need an additional assumption on $E \xi_k$.

Assumption B'. (h_k) satisfies Assumption B, and there exists $C < \infty$ such that for all t, k, T ,

$$|E \xi_t - E \xi_k| \leq C \begin{cases} \left(\frac{|t-k|}{t \vee k}\right)^{1/2}, & t, k = 1, 2, \dots \text{ if } (h_k) \text{ satisfies (54),} \\ \left(\frac{|t-k|}{T}\right)^{1/2}, & 1 \leq t, k \leq T \text{ if } (h_k) \text{ satisfies (55).} \end{cases} \quad (60)$$

$$(61)$$

Lemma 5 Let the assumptions of Lemma 4 hold and Assumption B' be satisfied. Then

$$P(|\tilde{S}_{T,t}^{(h)}| \geq \zeta) \leq \begin{cases} f_H(\gamma_1, \gamma_2, c, \zeta \wedge \zeta') & \text{if } (\xi_j) \in \mathcal{E}(s), s > 0, \\ g_H(\gamma_1, \theta', c, \zeta \wedge \zeta') & \text{if } (\xi_j) \in \mathcal{H}(\theta), \theta > 2, \end{cases}$$

where γ_1, γ_2 and θ' are the same as in Lemma 4.

Exponential inequalities allow to bound $S_{T,t}^{(h)}$ and $\max_{1 \leq s \leq T} |\tilde{S}_{T,s}^{(h)}|$.

Corollary 8 Let Assumptions A and B with $\alpha > 0$ hold and (43) be satisfied. Assume that $(\xi_j) \in \mathcal{E}(s)$, $s > 0$ or $(\xi_j) \in \mathcal{H}(\theta)$, $\theta > 2$.

(a) Then, for any sequence $1 \leq t = t_T \leq T$, as $T \rightarrow \infty$,

$$S_{T,t}^{(h)} = O_P(1 + d_{Ht}^{-1}). \quad (62)$$

In addition, if $E \xi_k$ satisfies Assumption B' then $\tilde{S}_{T,t}^{(h)} = O_P(1 + d_{Ht}^{-1})$.

(b) Let h_k satisfy (55), $E \xi_k$ satisfy (61), and $cT^\delta \leq H \leq T$ where $c, \delta > 0$ do not depend on H, T . Then,

$$\max_{1 \leq s \leq T} |\tilde{S}_{T,s}^{(h)}| = \begin{cases} O_P((1 + HT^{-1/2})(\log T)^{1/\gamma_1}) & \text{if } (\xi_j) \in \mathcal{E}(s), \\ O_P((1 + HT^{-1/2})\{(\log T)^{1/\gamma_1} + \frac{(TH)^{1/\theta}}{H^{-\varepsilon+1/2}}\}) & \text{if } (\xi_j) \in \mathcal{H}(\theta), \theta > 2 \end{cases} \quad (63)$$

$$(64)$$

with $\gamma_1 = 2\alpha/(2 + \alpha)$ and for any $\varepsilon > 0$.

The following auxiliary corollary provides bounds for sums of weighted variables $|h_k \xi_k|$. Denote

$$\Delta_{T,t} = H^{-1} \sum_{k=1}^T b_{H,|t-k|} |\beta_k - \beta_t| |h_k \xi_k|, \quad 1 \leq t \leq T.$$

To bound $\Delta_{T,t}$, we need an additional assumption on (β_k, h_k) .

Assumption C. (β_k, h_k) satisfy one of the following two assumptions:

$$|(\beta_k - \beta_t) h_k| \leq \begin{cases} C \frac{|t-k|}{T}, & 1 \leq t, k \leq T \\ \left(\frac{|t-k|}{T}\right)^{1/2} \nu_{tk} & 1 \leq t, k \leq T \end{cases} \quad (65)$$

$$\left(\frac{|t-k|}{T}\right)^{1/2} \nu_{tk} \quad 1 \leq t, k \leq T \quad (66)$$

where $C < \infty$ does not depend on t, k, T and $(\nu_{tk}) \in \mathcal{E}(\alpha)$ for some $\alpha > 0$.

Corollary 9 *Let Assumptions A, C and (43) be satisfied. Assume that $(\xi_j) \in \mathcal{E}(s)$, $s > 0$ or $(\xi_j) \in \mathcal{H}(\theta)$, $\theta > 2$.*

(a) *Then, for any $1 \leq t = t_T \leq T$, as $T \rightarrow \infty$,*

$$\Delta_{T,t} = \begin{cases} O_P(HT^{-1}) & \text{if (65) holds,} \\ O_P((H/T)^{1/2}) & \text{if (66) holds.} \end{cases} \quad (67)$$

$$O_P((H/T)^{1/2}) \quad \text{if (66) holds.} \quad (68)$$

(b) *Assume that $cT^\delta \leq H \leq T$ for some $c, \delta > 0$, and $\delta > 1/(\theta - 1)$ if $(\xi_j) \in \mathcal{H}(\theta)$. Then,*

$$\max_{1 \leq s \leq T} |\Delta_{T,s}| = \begin{cases} O_P(HT^{-1}) & \text{if (65) holds,} \\ O_P((H/T)^{1/2} (\log T)^{1/\alpha}) & \text{if (66) holds.} \end{cases} \quad (69)$$

$$O_P((H/T)^{1/2} (\log T)^{1/\alpha}) \quad \text{if (66) holds.} \quad (70)$$

5 Conclusion

The estimation of covariance matrices for large datasets has received considerable attention in recent years. Various regularization techniques for improved estimation of such matrices were developed, mainly for independent, identically distributed variables with exponentially declining probability tails. Dependence, heavy-tailed distributions and structural change are prevalent in large economic and financial data sets, and they may affect regularized estimation. The paper shows that the standard Bickel-Levina type thresholding procedure remains consistent for α -mixing stationary variables following various probability distributions. It takes a further step away from stationarity and allows for heteroscedasticity and stochastic change (volatility) of a very general form. It shows that the thresholding procedure can be aligned with kernel type estimates of time-varying covariances, and that the optimal threshold $\lambda_{opt} = \kappa \sqrt{(\log p)/H_{opt}}$ used in such a case is an intuitive adjustment for heteroscedasticity

compared to the threshold, $\lambda_{opt} = \kappa\sqrt{(\log p)/T}$, that is appropriate for stationary data. The paper shows that the thresholding procedure, adjusted for heteroscedasticity, is robust to dependence and the type of distribution of the data, and its tuning parameters can be selected by cross-validation. Its finite sample performance, illustrated by a detailed Monte Carlo study and an empirical application on designing minimum variance portfolios, provides a clear rationale for the proposed theoretical methods. Finally, the paper derives Bernstein type exponential inequalities for weighted sums of dependent random variables with thin or heavy-tailed distributions, that are of independent theoretical interest.

Supplementary Material

Dendramis, Y., Giraitis, L. and Kapetanios, G. (2020): Supplement to *Estimation of Time-varying Covariance Matrices for Large Datasets*, Econometric Theory Supplementary Material. To view, please visit: [[DOI will be inserted here by typesetter]]

References

- ABADIR, K. M., W. DISTASO, AND F. ZIKES (2014): “Design-free estimation of variance matrices,” *Journal of Econometrics*, 181, 165–180.
- BAI, J., AND S. NG (2002): “Determining the number of factors in approximate factor models,” *Econometrica*, 70(1), 191–221.
- BAILEY, N., G. KAPETANIOS, AND M. H. PESARAN (2016): “Exponent of cross-sectional dependence: estimation and inference,” *Journal of Applied Econometrics*, 31, 929–960.
- BICKEL, M., Y. WANG, AND H. ZHOU (2013): “Optimal sparse volatility matrix estimation for high-dimensional Itô processes with measurement errors,” *The Annals of Statistics*, 41, 1816–1864.
- BICKEL, P., AND E. LEVINA (2008): “Covariance regularization by thresholding,” *The Annals of Statistics*, 36, 2577–2604.
- CAI, T., AND W. LIU (2011): “Adaptive thresholding for sparse covariance matrix estimation,” *Journal of the American Statistical Association*, 106, 672–684.
- CHEN, X., M. XU, AND W. B. WU (2013): “Covariance and precision matrix estimation for high-dimensional time series,” *The Annals of Statistics*, 41, 2994–3021.

- DAHLHAUS, R. (1997): “Fitting time series models to nonstationary processes,” *Annals of Statistics*, 25, 1–37.
- ENGLE, R. (1982): “Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation,” *Econometrica*, 50, 987–1008.
- ENGLE, R. F., AND K. F. KRONER (1995): “Multivariate simultaneous generalized Arch,” *Econometric Theory*, 11(1), 122–150.
- FAN, J., Y. LIAO, AND H. LIU (2016): “An overview of the estimation of large covariance and precision matrices,” *Econometrics Journal*, 19, C1–C32.
- FAN, J., Y. LIAO, AND M. MINCHEVA (2013): “Large covariance estimation by thresholding principal orthogonal complements,” *Journal of the Royal Statistical Society: Series B*, 75, 603–680.
- FAN, J., W. WANG, AND Y. ZHONG (2016): “Robust covariance estimation for approximate factor models,” *arXiv:1602.00719v1*, pp. 1–31.
- GIRAITIS, L., G. KAPETANIOS, AND S. PRICE (2013): “Adaptive forecasting in the presence of recent and ongoing structural change,” *Journal of Econometrics*, 177(2), 153–170.
- GIRAITIS, L., G. KAPETANIOS, AND T. YATES (2014): “Inference on stochastic time-varying coefficient models,” *Journal of Econometrics*, 179, 46–65.
- (2018): “Inference on multivariate heteroscedastic time varying random coefficient models,” *Journal of Time Series Analysis*, 39, 129–149.
- HAN, F., AND H. LIU (2017): “Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution,” *Bernoulli*, 23, 23–57.
- JAGANNATHAN, R., AND T. MA (2003): “Risk reduction in large portfolios: why imposing the wrong constraints helps,” *Journal of Finance*, 58, 1651–1684.
- KAPETANIOS, G. (2010): “Testing for strict stationarity in financial variables,” *Journal of Banking and Finance*, 33, 2346–2362.
- KOLAR, M., AND E. XING (2011): “On time varying undirected graphs,” *In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS) 2011 (JMLR)*, 15, 407–415.

- LEDOIT, O., AND M. WOLF (2003): “Improved estimation of the covariance matrix of stock returns with an application to portfolio selection,” *Journal of Empirical Finance*, 10, 603–621.
- (2004): “A well-conditioned estimator for large-dimensional covariance matrices,” *Journal of Multivariate Analysis*, 88, 365–411.
- LEDOIT, O., AND M. WOLF (2015): “Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions,” *Journal of Multivariate Analysis*, 139, 360–384.
- MARCHENKO, V. A., AND L. A. PASTUR (1968): “Distribution of eigenvalues for some sets of random matrices,” *Math. USSR, Sb.*, 1, 457–483.
- MARKOWITZ, H. M. (1952): “Portfolio selection,” *Journal of Finance*, 7, 77–91.
- NOURELDIN, D., N. SHEPHARD, AND K. SHEPPARD (2014): “Multivariate rotated ARCH models,” *Journal of Econometrics*, 179(1), 16 – 30.
- POURAHMADI, M. (2013): *High-Dimensional Covariance Estimation*. Hoboken: John Wiley & Sons.
- STOCK, J. H., AND M. W. WATSON (2002a): “Forecasting using principal components from a large number of predictors,” *Journal of the American Statistical Association*, 97, 1167–1179.
- (2002b): “Macroeconomic forecasting using diffusion indexes,” *Journal of Business & Economic Statistics*, 20, 147–162.
- WEGKAMP, M., AND Y. ZHAO (2016): “Adaptive estimation of the copula correlation matrix for semiparametric elliptical copulas,” *Bernoulli*, 22, 1184–1226.
- WHITE, H., AND J. M. WOOLDRIDGE (1991): “Some results on sieve estimation with dependent observations,” in *Semiparametric and Nonparametric Methods in Econometrics and Statistics*, ed. by W. Barnett, J. Powell, and G. Tauchen, pp. 459–493. Cambridge University Press.
- ZHOU, S., J. LAFFERTY, AND L. WASSERMAN (2010): “Time varying undirected graphs,” *Machine Learning*, 80, 295–319.

Online Supplement to

”Estimation of time-varying covariance matrices for large datasets”

Yiannis Dendramis¹, Liudas Giraitis² and George Kapetanios³

¹Athens University of Economics and Business ²Queen Mary University of London

³King’s College, London

This Supplement provides proofs of the results given in the text of the main paper. It is organised as follows: Section A provides proofs of the main results on exponential inequalities of Section 4 of the main paper. Section B provides proofs of Theorems 1-3 of the main paper. Section C contains auxiliary technical lemmas.

Formula numbering in this supplement includes the section number, e.g. (A.1), and references to lemmas are signified as “Lemma A#”, “Lemma B#”, “Lemma C#”, e.g. Lemma A1. Equation, lemma and theorem references to the main paper do not include section number and are signified as “Equation (#)”, “Lemma #”, “Theorem #”, e.g. (1), Theorem 1.

In the proofs, C stands for a generic positive constant which may assume different values in different contexts, and we denote $a \vee b = \max(a, b)$, $a \wedge b = \min(a, b)$.

A. Exponential inequalities. Proofs

This section contains the proofs of the results of Section 4 on Bernstein inequalities for (weighted) sums of random variables ξ_j that are dependent, unbounded and have thin- or heavy-tailed distributions.

We shall frequently refer to the α -mixing Assumption A and property (36) of (ξ_j) of Section 4 of the main paper. To denote that r.v.’s (ξ_j) have thin- or heavy-tailed distributions, we use respectively notation $(\xi_j) \in \mathcal{E}(s)$, $s > 0$ and $(\xi_j) \in \mathcal{H}(\theta)$, $\theta > 2$ of Section 4 of the main paper, see (37) and (38).

Merlevede, Peligrad and Rio (2009) in their Theorem 2 obtained a Bernstein type inequality for bounded α -mixing random variables. The following lemma is a minor auxiliary generalization of their result to a sequence of truncated random variables.

Lemma A1 *Let the sequence (ξ_k) of zero mean random variables satisfy Assumption A. Set $\xi_{D,k} := \xi_k I(|\xi_k| \leq D)$ where $D > 0$. Suppose that*

$$m^* := \max_{k \geq 1} (E|\xi_k|^p)^{1/p} < \infty \quad \text{for some } p > 2.$$

Then, there exist $0 < c < \infty$ such that for all $\zeta > 0$, $D > 0$ and $T \geq 2$,

$$P\left(\left|\sum_{k=1}^T(\xi_{D,k} - E\xi_{D,k})\right| \geq \zeta\right) \leq \exp\left(-\frac{c\zeta^2}{\bar{v}^2 T + D^2 + \zeta D \log^2 T}\right), \quad (\text{A.1})$$

with $\bar{v}^2 = m^*(1 + 24 \sum_{j=1}^{\infty} \alpha_j^{1-2/p})$ where $c > 0$ depends only on c_* in (36) of Assumption A.

Proof of Lemma A1. By Theorem 14.1 of Davidson (1994), under Assumption A, the truncated process $(\xi_{D,t})$ is also α -mixing with mixing coefficients $\tilde{\alpha}_k \leq \alpha_k$. Hence, the bound (2.3) of Theorem 2 in Merlevede *et al.* (2009) implies

$$P\left(\left|\sum_{k=1}^T(\xi_{D,k} - E\xi_{D,k})\right| \geq \zeta\right) \leq \exp\left(-\frac{c\zeta^2}{v_D^2 T + D^2 + \zeta D \log^2 T}\right),$$

with

$$v_D^2 = \sup_{i>0} \left(\text{var}(\xi_{D,i}) + 2 \sum_{j>i} |\text{cov}(\xi_{D,i}, \xi_{D,j})| \right),$$

where c depends only on c_* in (36). We will show that $v_D^2 \leq \bar{v}^2$ which proves (A.1).

The conclusion (2.2) in Davydov (1968) applied with $p = q > 2$ gives

$$|\text{cov}(\xi_{D,i}, \xi_{D,j})| \leq 12(E|\xi_{D,i}|^p)^{1/p}(E|\xi_{D,j}|^p)^{1/p}\tilde{\alpha}_{|i-j|}^{1-2/p} \leq 12m^*\alpha_{|i-j|}^{1-2/p}.$$

Observe that $\text{var}(\xi_{D,i}) \leq E\xi_{D,i}^2 \leq (E|\xi_{D,i}|^p)^{2/p} \leq m^*$. Hence,

$$v_D^2 \leq m^* \left(1 + 24 \sum_{j=1}^{\infty} \alpha_j^{1-2/p}\right) = \bar{v}^2 < \infty$$

which completes the proof of the lemma. \square

The proof of Lemma 1 of the main paper combines the modified version, Lemma A1, of the exponential inequality for bounded random variables by Merlevede, Peligrad and Rio (2009), with a truncation argument employed in White and Wooldridge (1991).

Proof of Lemma 1. Without restriction of generality we prove the validity of (40), (41) for $\zeta \geq 1$. (The inequalities (40) and (41) can be extended to $0 < \zeta < 1$ by selecting large enough constant c_0 .) Recall that $S_T = T^{-1/2} \sum_{k=1}^T (\xi_k - E\xi_k)$.

We start with (40). We need to prove that

$$P(|S_T| > \zeta) \leq f_T(2, \gamma, c, \zeta) = c_0 \left\{ \exp(-c_1 \zeta^2) + \exp\left(-c_2 \left(\frac{\zeta \sqrt{T}}{\log^2 T}\right)^\gamma\right) \right\}, \quad (\text{A.2})$$

with $\gamma = s/(s+1)$ where positive constants c_0, c_1, c_2 do not depend on ζ, T . Denote by $D = D_{T,\zeta}$ the truncation constant depending on T, ζ which will be selected later. Write $\xi_k = w_k + v_k$ where $w_k = \xi_k I(|\xi_k| \leq D)$, $v_k = \xi_k I(|\xi_k| > D)$. Then,

$$\begin{aligned} S_T &= T^{-1/2} \sum_{k=1}^T (w_k - Ew_k) + T^{-1/2} \sum_{k=1}^T (v_k - Ev_k) \\ &=: s_{T,1} + s_{T,2} \end{aligned} \quad (\text{A.3})$$

and

$$P(|S_T| \geq \zeta) \leq P(|s_{T,1}| \geq \zeta/2) + P(|s_{T,2}| \geq \zeta/2).$$

Thus, to prove (A.2), it suffices to show that for some c , for all $\zeta \geq 1$, $T \geq 2$,

$$P(|s_{T,i}| \geq \zeta) \leq f_T(2, \gamma, c, \zeta), \quad i = 1, 2. \quad (\text{A.4})$$

By Assumption A, $(\xi_j - E\xi_j)$ is an α -mixing process which mixing coefficients α_k satisfy (36). Hence, by Theorem 14.1 in Davidson (1994), $(w_j - Ew_j)$ and $(v_j - Ev_j)$ are α -mixing sequences and their respective mixing coefficients $\alpha_{w,k}$ and $\alpha_{v,k}$ satisfy

$$\alpha_{w,k} \leq \alpha_k, \quad \alpha_{v,k} \leq \alpha_k, \quad k \geq 1. \quad (\text{A.5})$$

Thus, by Lemma A1, for all $T \geq 2$ and $D > 0$,

$$P(|s_{T,1}| \geq \zeta) \leq \exp\left(-\frac{c_1 \zeta^2 T}{\bar{v}^2 T + D^2 + \zeta T^{1/2} D \log^2 T}\right) \quad (\text{A.6})$$

where $c_1 > 0$ does not depend on T , D or ζ . Using, on the r.h.s. of (A.6), the inequality

$$-\frac{1}{|a| + |b| + |c|} \leq -\frac{1}{3 \max(|a|, |b|, |c|)},$$

with $a = \bar{v}^2 T$, $b = D^2$, $c = \zeta T^{1/2} D \log^2 T$, we obtain

$$P(|s_{T,1}| \geq \zeta) \leq \exp(-c'_1 \zeta^2) + \exp\left(-\frac{c'_2 \zeta^2 T}{D^2}\right) + \exp\left(-\frac{c'_2 \zeta T^{1/2}}{D \log^2 T}\right), \quad \zeta \geq 1 \quad (\text{A.7})$$

with $c'_1 = c_1/(3\bar{v}^2)$, $c'_2 = c_1/3$. Setting

$$\Delta_T = \frac{T^{1/2}}{\log^2 T},$$

(A.7) becomes

$$\begin{aligned} & P(|s_{T,1}| \geq \zeta) \\ & \leq \exp(-c'_1 \zeta^2) + \exp\left(-c'_2 \left(\frac{\zeta \Delta_T}{D}\right)^2 \log^4 T\right) + \exp\left(-\frac{c'_2 \zeta \Delta_T}{D}\right). \end{aligned} \quad (\text{A.8})$$

We select $D = D_{T,\zeta}$ such that $\zeta \Delta_T / D = D^s$. Then,

$$D = (\zeta \Delta_T)^{1/(s+1)}, \quad D^s = (\zeta \Delta_T)^{s/(s+1)} \quad \text{and} \quad \frac{\zeta \Delta_T}{D} = (\zeta \Delta_T)^{s/(s+1)}. \quad (\text{A.9})$$

For $\zeta \geq 1$, $T \geq 2$ it holds $\zeta \Delta_T \geq \Delta_T \geq 1$. This together with (A.9) implies $(\zeta \Delta_T)/D \geq 1$. Notice that $\log^4 T \geq \log^4 2 =: v > 0$ for $T \geq 2$, and $v \in (0, 1)$. Hence,

$$\begin{aligned} & \left(\frac{\zeta \Delta_T}{D}\right)^2 \log^4 T \geq \left(\frac{\zeta \Delta_T}{D}\right) v, \\ & \frac{\zeta \Delta_T}{D} \geq v \frac{\zeta \Delta_T}{D} = v (\zeta \Delta_T)^{s/(s+1)} = v \left(\frac{\zeta \sqrt{T}}{\log^2 T}\right)^{s/(s+1)}. \end{aligned}$$

Applying these relations in (A.8), we obtain

$$\begin{aligned}
P(|s_{T,1}| \geq \zeta) &\leq \exp(-c'_1 \zeta^2) + 2 \exp\left(-\frac{c'_2 v \zeta \Delta_T}{D}\right) \\
&\leq \exp(-c'_1 \zeta^2) + 2 \exp(-c'_2 v (\zeta \Delta_T)^{s/(s+1)}) \\
&\leq 2 \left(\exp(-c'_1 \zeta^2) + \exp\left(-c'_2 v \left(\frac{\zeta \sqrt{T}}{\log^2 T}\right)^{s/(s+1)}\right) \right) \\
&\leq f_T(2, \gamma, c, \zeta).
\end{aligned}$$

This proves (A.4) for $P(|s_{T,1}| \geq \zeta)$. Turning to $s_{T,2}$, by Markov inequality,

$$\begin{aligned}
P(|s_{T,2}| \geq \zeta) &\leq \zeta^{-2} T^{-1} E \left(\sum_{k=1}^T (v_k - E v_k) \right)^2 \\
&\leq \zeta^{-2} T^{-1} \sum_{j,k=1}^T \text{cov}(v_j, v_k).
\end{aligned} \tag{A.10}$$

Let $p, q > 1$, $1/p + 1/q < 1$. Assumption $(\xi_j) \in \mathcal{E}(s)$ implies $E|v_j|^p < \infty$, $E|v_j|^q < \infty$. Since $(v_j - E v_j)$ is α -mixing sequence with the mixing coefficients $\alpha_{v,k} \leq \alpha_k$, $k \geq 1$, then, by Conclusion 2.2 in Davydov (1968),

$$|\text{cov}(v_j, v_k)| \leq 12(E|v_j|^p)^{1/p} (E|v_j|^q)^{1/q} \alpha_{|j-k|}^{1-1/p-1/q}, \quad j \neq k. \tag{A.11}$$

In turn, for $j = k$, $\text{var}(v_j) \leq E v_j^2$. Setting

$$V_p := \max_{j \geq 1} (E|v_j|^p)^{1/p},$$

we obtain

$$\begin{aligned}
P(|s_{T,2}| \geq \zeta) &\leq \zeta^{-2} T^{-1} \left[\sum_{j=1}^T \text{var}(v_j) + \sum_{j,k=1: k \neq j}^T \text{cov}(v_j, v_k) \right] \\
&\leq \zeta^{-2} V_2^2 + \zeta^{-2} 12 V_p V_q (T^{-1} \sum_{j,k=1: k \neq j}^T \alpha_{|j-k|}^e)
\end{aligned}$$

where $e := 1 - 1/p - 1/q > 0$. By (36),

$$T^{-1} \sum_{j,k=1: j > k}^T \alpha_{|j-k|}^e = T^{-1} \sum_{s=1}^T \alpha_s^e (T-s) \leq \sum_{s=1}^{\infty} \alpha_s^e < \infty.$$

This implies that with some C that does not depend on T or D , it holds that

$$P(|s_{T,2}| \geq \zeta) \leq C \zeta^{-2} (V_2^2 + V_p V_q). \tag{A.12}$$

Set $p = q = 2 + \delta$ where $\delta > 0$ is a small number. Then, by (A.12),

$$P(|s_{T,2}| \geq \zeta) \leq C \zeta^{-2} (V_2^2 + V_p^2) \leq C \zeta^{-2} V_p^2 \tag{A.13}$$

because $V_2^2 = \max_j E v_j^2 \leq \max_j (E |v_j|^p)^{2/p} = V_p^2$. For $D > 0$, by (C.9) it holds that

$$E |v_j|^p = E [|\xi_j|^p I(|\xi_j| > D)] \leq c'_0 \exp(-c'_1 D^s)$$

for some $c'_0, c'_1 > 0$ which do not depend on j and D . This implies

$$V_p^2 \leq (c'_0)^{2/p} \exp(-(2/p)c'_1 D^s).$$

Thus, there exists $c_0 > 0, c_2 > 0$ such that for all $\zeta \geq 1, T \geq 2$, in view of (A.9),

$$\begin{aligned} P(|s_{T,2}| \geq \zeta) &\leq C \zeta^{-2} \exp(- (2/p)c'_2 D^s) \\ &\leq c_0 \exp(- c_2 (\zeta \Delta_T)^{s/(s+1)}) \\ &= c_0 \exp\left(- c_2 \left(\frac{\zeta \sqrt{T}}{\log^2 T}\right)^{s/(s+1)}\right) \\ &\leq f_T(2, \gamma, c, \zeta), \end{aligned}$$

which proves the bound (A.4) for $s_{T,2}$ and completes the proof of (A.2) and (40).

Proof of (41). Let $(\xi_j) \in \mathcal{H}(\theta)$. We need to prove that for any fixed $2 < \theta' < \theta$,

$$\begin{aligned} P(|S_T| > \zeta) &\leq g_T(2, \theta', c, \zeta) \\ &= c_0 \left\{ \exp(-c_1 \zeta^2) + \zeta^{-\theta'} T^{-(\theta'/2-1)} \right\}, \quad \zeta > 0, \quad T \geq 2. \end{aligned} \tag{A.14}$$

Write $S_T = s_{T,1} + s_{T,2}$ as in (A.3). To verify (A.14), it remains to show that

$$P(|s_{T,i}| \geq \zeta) \leq g_T(2, \theta', c, \zeta), \quad i = 1, 2 \quad \text{for some } c.$$

It suffices to consider the case $\zeta \geq 1$.

We start with the evaluation $P(|s_{T,1}| \geq \zeta)$. Set

$$D = \frac{a^{-1} \zeta \sqrt{T}}{\log^3(\zeta \sqrt{T})} \geq 1,$$

where $a > 0$ will be selected below. For $\zeta \geq 1$ it holds $\log(\zeta \sqrt{T}) \geq \log(\sqrt{T}) \geq \log(\sqrt{2}) =: b > 0$. Then, from (A.7) we obtain

$$\begin{aligned} P(|s_{T,1}| \geq \zeta) &\leq \exp(-c'_1 \zeta^2) + \exp(-c'_2 a^2 \log^6(\zeta \sqrt{T})) + \exp(-c'_2 a \log(\zeta \sqrt{T})) \\ &\leq \exp(-c'_1 \zeta^2) + \exp(-c'_2 a^2 b^5 \log(\zeta \sqrt{T})) + \exp(-c'_2 a \log(\zeta \sqrt{T})). \end{aligned}$$

Hence, selecting a such that $c'_2 a^2 b^5 \geq \theta', c'_2 a \geq \theta'$, we obtain

$$P(|s_{T,1}| \geq \zeta) \leq \exp(-c'_1 \zeta^2) + 2(\zeta \sqrt{T})^{-\theta'}.$$

This proves the bound (A.14) for $P(|s_{T,1}| \geq \zeta)$.

Next we turn to $P(|s_{T,2}| \geq \zeta)$. By (A.13),

$$P(|s_{T,2}| \geq \zeta) \leq C\zeta^{-2}V_p^2$$

with $p = 2 + \delta$. According to (C.10), we can bound

$$E|v_j|^p = E[|\xi_j|^p I(|\xi_j| > D)] \leq c'_0 D^{-(\theta-p)}$$

with some $c'_0 > 0$ which does not depend on D and j . This implies

$$V_p^2 \leq (c'_0)^{2/p} D^{-(\theta-p)(2/p)}.$$

Hence,

$$\begin{aligned} P(|s_{T,2}| \geq \zeta) &\leq C\zeta^{-2}D^{-(\theta-p)(2/p)} \\ &= C\zeta^{-2}(\zeta\sqrt{T})^{-(\theta'-2)}a_{T,\zeta}, \end{aligned} \tag{A.15}$$

where

$$a_{T,\zeta} := \frac{(\zeta\sqrt{T})^{\theta'-2}}{D^{(\theta-p)(2/p)}} = \frac{(a \log^3(\zeta\sqrt{T}))^{(\theta-p)(2/p)}}{(\zeta\sqrt{T})^\gamma}$$

and

$$\gamma = (\theta - p)(2/p) - (\theta' - 2) = \theta - \theta' - \theta(p - 2)/p = \theta - \theta' - \theta\delta/p > 0$$

when $\theta > \theta'$, $p = 2 + \delta$ and $\delta > 0$ is selected sufficiently small. Since $\zeta\sqrt{T} \geq \sqrt{2}$ for $\zeta \geq 1$, $T \geq 2$, this implies that $\sup_{\zeta \geq 1, T \geq 2} a_{T,\zeta} \leq C' < \infty$. Thus, (A.15) implies

$$P(|s_{T,2}| \geq \zeta) \leq C\zeta^{-\theta'}T^{-(\theta'/2-1)} \leq g_T(2, \theta', c, \zeta)$$

which proves the bound (A.14) for $P(|s_{T,2}| \geq \zeta)$.

This completes the proof of (41) and the lemma. \square

We start the proof of Lemma 2 with the following technical lemma.

Lemma A2 *Let x_{tk} , $k, t \geq 1$ be random variables such that $E|x_{tk}| < \infty$ and a_{tk} and $v_{tk} > 0$ be real numbers such that*

$$\max_{n \geq 1} \max_{1 \leq t \leq n} \sum_{1 \leq k \leq n}^n |a_{tk}|v_{tk} < \infty. \tag{A.16}$$

Then there exists $\varepsilon > 0$ such that for all $\zeta > 0$, $t \geq 1$,

$$P\left(\left|\sum_{k=1}^n a_{tk}x_{tk}\right| \geq \zeta\right) \leq \varepsilon^{-1} \max_{1 \leq k \leq n} E\left[\frac{|x_{tk}|}{\zeta v_{tk}} I\left(\frac{|x_{tk}|}{\zeta v_{tk}} \geq \varepsilon\right)\right]. \tag{A.17}$$

Proof of Lemma A2. By (A.16) there exists $\varepsilon > 0$ such that

$$\sum_{k=1}^n |a_{tk}v_{tk}| < 1/(2\varepsilon), \quad t \geq 1.$$

From

$$|x_{tk}/v_{tk}| = |x_{tk}/v_{tk}|(I(|x_{tk}/v_{tk}| \leq \varepsilon\zeta) + I(|x_{tk}/v_{tk}| > \varepsilon\zeta)) \leq \varepsilon\zeta + y_{tk},$$

where $y_{tk} := |x_{tk}/v_{tk}|I(|x_{tk}/v_{tk}| \geq \varepsilon\zeta)$, we obtain

$$\begin{aligned} \left| \sum_{k=1}^n a_{tk}x_{tk} \right| &\leq \sum_{k=1}^n |a_{tk}v_{tk}| \frac{|x_{tk}|}{v_{tk}} \leq \sum_{k=1}^n |a_{tk}v_{tk}|(\varepsilon\zeta) + \sum_{k=1}^n |a_{tk}v_{tk}|y_{tk} \\ &\leq \zeta/2 + \sum_{k=1}^n |a_{tk}v_{tk}|y_{tk}. \end{aligned}$$

Then, by Markov inequality,

$$\begin{aligned} P\left(\left|\sum_{k=1}^n a_{tk}x_{tk}\right| \geq \zeta\right) &\leq P\left(\sum_{k=1}^n |a_{tk}v_{tk}|y_{tk} \geq \zeta/2\right) \leq (\zeta/2)^{-1} \sum_{k=1}^n |a_{tk}v_{tk}|Ey_{tk} \\ &\leq (\zeta/2)^{-1} \left(\sum_{k=1}^n |a_{tk}v_{tk}|\right) \max_{1 \leq k \leq n} Ey_{tk} \\ &\leq (\zeta/2)^{-1} (2\varepsilon)^{-1} \max_{1 \leq k \leq n} Ey_{tk} \end{aligned}$$

which proves (A.17). \square

Proof of Lemma 2. Without restriction of generality assume that $\zeta \geq 1$. Notice that property (43) of $b_{H,k}$ implies

$$\max_{k=1, \dots, T} b_{H,k} \left(\frac{k \vee H}{H}\right)^{1/2} \leq C, \quad \sum_{k=1}^{T-1} |b_{H,k} - b_{H,k+1}| \left(\frac{k \vee H}{H}\right)^{1/2} \leq C, \quad (\text{A.18})$$

where $C < \infty$ does not depend on H, T .

Denote $\xi'_k := \xi_{t-k}$, $\xi''_k := \xi_{t+k}$ for $k \geq 0$. Write

$$\begin{aligned} S_{T,t} &= H^{-1/2} \sum_{k=1}^T b_{H,|t-k|} (\xi_k - E\xi_k) \\ &= H^{-1/2} \sum_{k=1}^{t-1} b_{H,k} (\xi'_k - E\xi'_k) + H^{-1/2} \sum_{k=0}^{T-t} b_{H,k} (\xi''_k - E\xi''_k) \\ &=: s_{T,t}^{(1)} + s_{T,t}^{(2)}. \end{aligned} \quad (\text{A.19})$$

To prove (44) for $P(|S_{T,t}| \geq \zeta)$, it suffices to verify that for $\ell = 1, 2$,

$$P(|s_{T,t}^{(\ell)}| \geq \zeta) \leq f_H(2, \gamma, c, \zeta) \quad \text{if } (\xi_j) \in \mathcal{E}(s), \quad s > 0. \quad (\text{A.20})$$

To prove (45) for $P(|S_{T,t}| \geq \zeta)$, it suffices to show that for $\ell = 1, 2$,

$$P(|s_{T,t}^{(\ell)}| \geq \zeta) \leq g_H(2, \theta', c, \zeta) \quad \text{if } (\xi_j) \in \mathcal{H}(\theta), \quad \theta > 2. \quad (\text{A.21})$$

We provide the proof for $s_{T,t}^{(1)}$. (For $s_{T,t}^{(2)}$ the proof is similar). Set

$$x_k = \sum_{i=1}^k (\xi'_i - E\xi'_i), \quad y_k = k^{-1/2}x_k, \quad \nu_k = \left(\frac{k \vee H}{k}\right)^{1/2} \quad \text{for } k = 1, \dots, t-1.$$

Using summation by parts, we can write $s_{T,t}^{(1)}$ as

$$\begin{aligned} s_{T,t}^{(1)} &= H^{-1/2} \sum_{k=1}^{t-2} (b_{H,k} - b_{H,k+1})x_k + H^{-1/2}b_{H,t-1}x_{t-1} \\ &= \sum_{k=1}^{t-1} a_{tk}x_k, \end{aligned} \tag{A.22}$$

where

$$a_{tk} = H^{-1/2}(b_{H,k} - b_{H,k+1}) \quad \text{for } k = 1, \dots, t-2, \quad a_{t,t-1} = H^{-1/2}b_{H,t-1}.$$

Subsequently, using notation y_k and ν_k introduced above, we can write

$$s_{T,t}^{(1)} = \sum_{k=1}^{t-1} a_{tk} (k \vee H)^{1/2} (y_k/\nu_k). \tag{A.23}$$

From (A.18) it follows

$$\sum_{k=1}^{t-1} a_{tk} (k \vee H)^{1/2} \leq C$$

where $C < \infty$ does not depend on t, H, T . Hence, by Lemma A2, there exists $\varepsilon > 0$ such that

$$p_{T,\zeta} = P(|s_{T,t}^{(1)}| \geq \zeta) \leq \varepsilon^{-1} \max_{1 \leq k < t} E\left[\frac{|y_k|}{\zeta \nu_k} I\left(\frac{|y_k|}{\zeta \nu_k} \geq \varepsilon\right)\right]. \tag{A.24}$$

Notice that $\nu_k \geq 1$.

Proof of (44). Suppose that $(\xi_j) \in \mathcal{E}(s)$. Then, (40) of Lemma 1 implies

$$P(|y_k| \geq \zeta) \leq f_k(2, \gamma, c, \zeta), \quad \zeta > 0, \quad k \geq 2.$$

Therefore, by (C.11) of Lemma C2(ii),

$$E[|y_k| I(|y_k| \geq \varepsilon \zeta \nu_k)] \leq f_k(2, \gamma, c', \varepsilon \zeta \nu_k)$$

for some c' which does not depend on k . Thus, (A.24) implies

$$p_{T,\zeta} \leq C \max_{1 \leq k < t} \nu_k^{-1} f_k(2, \gamma, c', \varepsilon \zeta \nu_k) \leq C \max_{1 \leq k < t} f_k(2, \gamma, c', \varepsilon \zeta \nu_k). \tag{A.25}$$

For $k \geq H$, it holds that $\nu_k = 1$, and we have

$$\begin{aligned} f_k(2, \gamma, c, \varepsilon \zeta \nu_k) &= c_0 \left\{ \exp(-c_1(\varepsilon \zeta)^2) + \exp\left(-c_2 \left(\frac{\varepsilon \zeta \sqrt{k}}{\log^2 k}\right)^{s/(s+1)}\right) \right\} \\ &\leq f_H(2, \gamma, c, \zeta). \end{aligned}$$

For $1 \leq k < H$, we have $\nu_k = (H/k)^{1/2} \geq 1$ and $\nu_k \sqrt{k} = \sqrt{H}$, which allows to conclude

$$\begin{aligned} f_k(2, \gamma, c, \varepsilon \zeta \nu_k) &= c_0 \left\{ \exp(-c_1(\varepsilon \zeta \nu_k)^2) + \exp\left(-c_2 \left(\frac{\varepsilon \zeta \nu_k \sqrt{k}}{\log^2 k}\right)^{s/(s+1)}\right) \right\} \\ &\leq c_0 \left\{ \exp(-c_1(\varepsilon \zeta)^2) + \exp\left(-c_2 \left(\frac{\varepsilon \zeta \sqrt{H}}{\log^2 H}\right)^{s/(s+1)}\right) \right\} \\ &= f_H(2, \gamma, c, \zeta). \end{aligned}$$

Together with (A.24), this yields $p_{T,\zeta} \leq f_H(2, \gamma, c, \zeta)$ which proves (A.20).

Proof of (45). Assume that $(\xi_j) \in \mathcal{H}(\theta)$ and let $\theta' \in (2, \theta)$. By (41) of Lemma 1,

$$P(|y_k| \geq \zeta) \leq g_k(2, \theta', c, \zeta)$$

for $k \geq 2$, and by (C.12) of Lemma C2(iii),

$$E[|y_k| I(|y_k| \geq \varepsilon \zeta \nu_k)] \leq \max(\varepsilon \zeta \nu_k, 1) g_k(2, \theta', c, \varepsilon \zeta \nu_k) \quad (\text{A.26})$$

for some c which does not depend on k . Notice that $\zeta \nu_k \geq 1$. Then,

$$(\varepsilon \zeta \nu_k)^{-1} \max(\varepsilon \zeta \nu_k, 1) \leq \max(1, (\varepsilon \zeta \nu_k)^{-1}) \leq 1 + \varepsilon^{-1}.$$

Thus, by (A.26) and (A.24),

$$p_{T,\zeta} \leq C \max_{1 \leq k < t} g_k(2, \theta', c', \varepsilon \zeta \nu_k) \quad (\text{A.27})$$

where C depends on ε . For $k \geq H$ we have $\nu_k = 1$, and therefore

$$\begin{aligned} g_k(2, \theta', c, \varepsilon \zeta \nu_k) &= g_k(2, \theta', c, \varepsilon \zeta) = c_0 \left\{ \exp(-c_1(\varepsilon \zeta)^2) + (\varepsilon \zeta)^{-\theta'} k^{-(\theta'/2-1)} \right\} \\ &\leq g_H(2, \theta', c, \varepsilon \zeta). \end{aligned}$$

For $k \leq H$, we have $\nu_k = (H/k)^{1/2} \geq 1$ and therefore

$$\begin{aligned} (\zeta \nu_k)^{-\theta'} k^{-(\theta'/2-1)} &= (\zeta (H/k)^{1/2})^{-\theta'} k^{-\theta'/2} k = (\zeta H^{1/2})^{-\theta'} k \\ &\leq \zeta^{-\theta'} H^{-(\theta'/2-1)} \end{aligned}$$

which allows to conclude

$$\begin{aligned} g_k(2, \theta', c, \varepsilon \zeta \nu_k) &= c_0 \left\{ \exp(-c_1(\varepsilon \zeta \nu_k)^2) + (\varepsilon \zeta \nu_k)^{-\theta'} k^{-(\theta'/2-1)} \right\} \\ &\leq c_0 \left\{ \exp(-c_1(\varepsilon \zeta)^2) + (\varepsilon \zeta)^{-\theta'} H^{-(\theta'/2-1)} \right\} \\ &= g_H(2, \theta', c, \varepsilon \zeta). \end{aligned}$$

Together with (A.27), this implies $p_{T,\zeta} \leq g_H(2, \theta', c, \zeta)$ which proves (A.21). \square

Proof of Lemma 3. (a) Write $\tilde{S}_{T,t} = S_{T,t} + r_{T,t}$. Assumption $\zeta > 2|r_{T,t}|$ implies $\zeta - |r_{T,t}| \geq \zeta/2$. Therefore

$$P(|\tilde{S}_{T,t}| \geq \zeta) \geq P(|S_{T,t}| \geq \zeta - |r_{T,t}|) \leq P(|S_{T,t}| \geq \zeta/2).$$

(b) If $|E\xi_k - E\xi_t| \leq C|k - t|/(t \vee k)$ for $k, t \geq 1$, then by (C.16) of Lemma C3,

$$|r_{T,t}| \leq CH^{-1/2} \sum_{k=1}^T b_{H,|t-k|} \left(\frac{|t-k|}{t \vee k} \right) \leq c_1 \frac{H^{3/2}}{H \vee t}$$

for some $c_1 > 0$ which proves (46).

If $|E\xi_k - E\xi_t| \leq C|k - t|/T$ for $k, t = 1, \dots, T$, then by (43),

$$\begin{aligned} |r_{T,t}| &\leq CH^{-1/2} \sum_{k=1}^T b_{H,|t-k|} \left(\frac{|t-k|}{T} \right) \\ &\leq C \left(H^{-1} \sum_{k=1}^T b_{H,|t-k|} \left(\frac{|t-k|}{H} \right) \right) \frac{H^{3/2}}{T} \leq c_2 \frac{H^{3/2}}{T} \end{aligned} \quad (\text{A.28})$$

for some $c_2 > 0$ which does not depend on t, H, T . This proves (47). \square

Proof of Corollary 6.

Proof of (a). The bounds (44)-(45) together with definition of f_t, g_t in (39) imply

$$P(|S_{T,t}| \geq b) \rightarrow 0, \quad T \rightarrow \infty, \quad b \rightarrow \infty.$$

Hence, $S_{T,t} = O_P(1)$ which proves (48).

Proof of (b). Assume that $(\xi_j) \in \mathcal{E}(s)$, $s > 0$. We will show that as $T \rightarrow \infty, b \rightarrow \infty$,

$$P\left(\max_{t=1, \dots, T} |S_{T,t}| > b\delta_{T,H}\right) = o_P(1), \quad \delta_{T,H} = (\log T)^{1/2} + \frac{(\log H)^2}{H^{1/2}} (\log T)^{1/\gamma} \quad (\text{A.29})$$

with $\gamma = s/(s+1)$. Let $b > 0$. Then by (44) and (39),

$$\begin{aligned} P\left(\max_{t=1, \dots, T} |S_{T,t}| \geq b\delta_{T,H}\right) &\leq \sum_{t=1}^T P(|S_{T,t}| \geq b\delta_{T,H}) \\ &\leq \sum_{t=1}^T f_H(2, \gamma, c, b\delta_{T,H}) \\ &\leq Tc_0 \left\{ \exp(-c_1(b\delta_{T,H})^2) + \exp\left(-c_2 \left(\frac{b\delta_{T,H}\sqrt{H}}{\log^2 H}\right)^\gamma\right) \right\} \\ &\leq Tc_0 \left\{ \exp(-c_1 b^2 \log T) + \exp(-c_2 b^\gamma \log T) \right\} \\ &\leq 2T^{-1} \rightarrow 0 \end{aligned}$$

for b such that $c_1 b^2 \geq 2, c_2 b^\gamma \geq 2$. This proves (A.29). Under assumption $cT^\delta \leq H \leq T$ it holds $\delta_{T,H} = O(\log^{1/2} T)$. Hence, (A.29) implies (50):

$$P\left(\max_{t=1, \dots, T} |S_{T,t}| > b \log^{1/2} T\right) \rightarrow 0, \quad T \rightarrow \infty, \quad b \rightarrow \infty.$$

Next, assume that $(\xi_j) \in \mathcal{H}(\theta)$, $\theta > 2$. Let $\theta' \in (2, \theta)$. We will show that, as $T \rightarrow \infty$, $b \rightarrow \infty$,

$$P\left(\max_{t=1, \dots, T} |S_{T,t}| > b\delta_{T,H}\right) = o_P(1), \quad \delta_{T,H} = (\log T)^{1/2} + H^{1/2} \left(\frac{T}{H^{\theta'-1}}\right)^{1/\theta'}. \quad (\text{A.30})$$

For $b > 0$, by (45) and definition of g_t , (39),

$$\begin{aligned} & P(\max_{t=1, \dots, T} |S_{T,t}| \geq b\delta_{T,H}) \\ & \leq \sum_{t=1}^T P(|S_{T,t}| \geq b\delta_{T,H}) \leq \sum_{t=1}^T g_H(2, \theta', c, b\delta_{T,H}) \\ & \leq Tc_0 \left\{ \exp(-c_1(b\delta_{T,H})^2) + (b\delta_{T,H})^{-\theta'} H^{-(\theta'/2-1)} \right\} \\ & \leq c_0 \left\{ T \exp(-c_1 b^2 \log T) + b^{-\theta'} \left(\frac{T}{H^{\theta'/2-1}}\right)^{-1} T H^{-(\theta'/2-1)} \right\} \\ & \leq c_0 \{T^{-1} + b^{-\theta'}\} \rightarrow 0 \end{aligned}$$

as $T \rightarrow \infty$ and $b \rightarrow \infty$. This proves (A.30). To prove that (A.30) implies (51), it suffices to show that for any $\varepsilon > 0$ there exists $2 < \theta' < \theta$ and $a > 0$ such that

$$\log^{1/2} T + (HT)^{1/\theta} H^{\varepsilon-1/2} \geq a\delta_{T,H}, \quad \delta_{T,H} = \log^{1/2} T + (TH)^{1/\theta'} H^{-1/2}. \quad (\text{A.31})$$

Write

$$(HT)^{1/\theta} H^{\varepsilon-1/2} = (HT)^{1/\theta'} H^{-1/2} v_H, \quad v_H := (HT)^{1/\theta-1/\theta'} H^\varepsilon.$$

We will show that $v_H \geq a > 0$ for some $1 > a > 0$ which proves (A.31). By the assumption of the corollary, $cT^\delta \leq H \leq T$. Then,

$$v_H = \frac{H^\varepsilon}{(HT)^{1/\theta'-1/\theta}} \geq \frac{(cT^\delta)^\varepsilon}{T^{2(1/\theta'-1/\theta)}} = c^\varepsilon T^b, \quad b := \delta\varepsilon - 2(1/\theta' - 1/\theta).$$

If $b \geq 0$, this implies $v_H \geq c^\varepsilon$. Clearly, $b \geq 0$ if θ' is selected sufficiently close to θ . \square

Proof of Corollary 7. Let $0 \leq \nu \leq 1$. Write

$$v_{T,t} := H^{-1} \sum_{k=1}^T \tilde{b}_{H,|t-k|} |\xi_k|, \quad \tilde{b}_{H,|t-k|} := b_{H,|t-k|} (|t-k|/H)^\nu. \quad (\text{A.32})$$

By (43), $\tilde{b}_{H,|t-k|} \leq C(1 + (k/H)^{\nu-1})^{-1}$. It is easy to see that $\tilde{b}_{H,k}$ satisfies (43) with parameter $\nu - 1$. Since under assumptions of corollary,

$$\max_{k \geq 1} E|\xi_k| \leq C < \infty, \quad \max_{t \geq 1} H^{-1} \sum_{k=1}^T \tilde{b}_{H,|t-k|} \leq C,$$

then

$$\max_{1 \leq t \leq T} |v_{T,t}| \leq \max_{1 \leq t \leq T} |v'_{T,t}| + C, \quad v'_{T,t} = H^{-1} \sum_{k=1}^T \tilde{b}_{H,|t-k|} (|\xi_k| - E|\xi_k|).$$

Since (ξ_k) satisfies Assumption A, then by Theorem 14.1 in Davidson (1994), $(|\xi_k|)$ also satisfies Assumption A. To prove the claim (52)-(53) of the corollary, it remains to show that

$$\max_{1 \leq t \leq T} |v'_{T,t}| = O_P(1). \quad (\text{A.33})$$

Let $(\xi_j) \in \mathcal{E}(s)$, $s > 0$. Then, by (50) of Corollary 6 and assumption (49) on H ,

$$\max_{1 \leq t \leq T} |v'_{T,t}| = O(H^{-1/2} \log^{1/2} T) = o_P(1).$$

Let $(\xi_j) \in \mathcal{H}(\theta)$, $\theta > 2$. Then, by (51) of Corollary 6 and assumption (49), for any $\varepsilon > 0$,

$$\max_{1 \leq t \leq T} |v'_{T,t}| = O_P(H^{-1/2} \log^{1/2} T + (TH)^{1/\theta} H^{\varepsilon-1}).$$

By assumption, $H \geq cT^\delta$ with $\delta > 1/(\theta - 1)$ which implies that $(TH)^{1/\theta} H^{\varepsilon-1} = o(1)$ when ε is selected sufficiently small. This proves (A.33) and completes the proof of the corollary. \square

Proof of Lemma 4. Without restriction of generality assume that $\zeta \geq 1$.

Proof of (58)-(59) for $P(|S_{T,t}^{(h)}| \geq \zeta)$. Denote

$$h'_k := h_{t-k}, \xi'_k := \xi_{t-k}, h''_k := h_{t+k}, \xi''_k := \xi_{t+k} \text{ for } k \geq 0.$$

As in (A.19) write $S_{T,t}^{(h)}$ as

$$\begin{aligned} S_{T,t}^{(h)} &= H^{-1/2} \sum_{k=1}^T b_{H,|t-k|} h_k (\xi_k - E\xi_k) \\ &= H^{-1/2} \sum_{k=1}^{t-1} b_{H,k} h'_k \xi'_k + H^{-1/2} \sum_{k=0}^{T-t} b_{H,k} h''_k \xi''_k \\ &=: s_{T,t;1}^{(h)} + s_{T,t;2}^{(h)}. \end{aligned}$$

Proof of (58)-(59) for $S_{T,t}^{(h)}$ reduces to verification of these bounds for $s_{T,t;1}^{(h)}$ and $s_{T,t;2}^{(h)}$:

$$P(|s_{T,t;\ell}^{(h)}| \geq \zeta) \leq \begin{cases} f_H(\gamma_1, \gamma_2, c, \zeta \wedge \zeta') & \text{if } (\xi_j) \in \mathcal{E}(s), \\ g_H(\gamma_1, \theta', c, \zeta \wedge \zeta') & \text{if } (\xi_j) \in \mathcal{H}(\theta), \theta > 2 \end{cases} \quad (\text{A.34})$$

$$P(|s_{T,t;\ell}^{(h)}| \geq \zeta) \leq \begin{cases} f_H(\gamma_1, \gamma_2, c, \zeta \wedge \zeta') & \text{if } (\xi_j) \in \mathcal{E}(s), \\ g_H(\gamma_1, \theta', c, \zeta \wedge \zeta') & \text{if } (\xi_j) \in \mathcal{H}(\theta), \theta > 2 \end{cases} \quad (\text{A.35})$$

for $\ell = 1, 2$. We start with $s_{T,t;1}^{(h)}$. Denote

$$x_k = \sum_{i=1}^k (\xi'_i - E\xi'_i), \quad y_k = k^{-1/2} x_k, \quad y'_k = h'_k y_k, \quad \nu_k := \left(\frac{k \vee H}{k} \right)^{1/2}, \quad k \geq 1. \quad (\text{A.36})$$

Then as in (A.22), summation by parts yields

$$\begin{aligned} s_{T,t;1}^{(h)} &= H^{-1/2} \sum_{k=1}^{t-2} (b_{H,k} h'_k - b_{H,k+1} h'_{k+1}) x_k + H^{-1/2} b_{H,t-1} h'_{t-1} \\ &= \left\{ H^{-1/2} \sum_{k=1}^{t-2} (b_{H,k} - b_{H,k+1}) (h'_k x_k) + H^{-1/2} b_{H,t-1} (h'_{t-1} x_{t-1}) \right\} \\ &\quad + H^{-1/2} \sum_{k=1}^{t-2} b_{H,k} (h'_k - h'_{k+1}) x_k \\ &=: s_{T,t;1}^{(1)} + s_{T,t;1}^{(2)}. \end{aligned} \quad (\text{A.37})$$

Hence, it suffices to verify the bounds (A.34)-(A.35) for $s_{T,t;1}^{(1)}$ and $s_{T,t;1}^{(2)}$.

First, we evaluate $P(|s_{T,t;1}^{(1)}| \geq \zeta)$. The sum $s_{T,t;1}^{(1)}$ can be obtained from $s_{T,t}^{(1)}$ in (A.22) by replacing x_k by $h'_k x_k$. Therefore, the same argument as in the proof of (A.24) implies that there exists $\varepsilon > 0$ such that

$$P(|s_{T,t;1}^{(1)}| \geq \zeta) \leq \varepsilon^{-1} \max_{k=1,\dots,T} (\zeta \nu_k)^{-1} E[|y'_k| I(|y'_k| \geq \varepsilon \zeta \nu_k)], \quad \zeta \geq 1, \quad T \geq 2. \quad (\text{A.38})$$

We now show that for all $\zeta > 0$, $k \geq 2$,

$$P(|y'_k| \geq \zeta) \leq f_k(\gamma_1, \gamma_2, c, \zeta) \quad \text{if } (\xi_k) \in \mathcal{E}(s), \quad (\text{A.39})$$

$$P(|y'_k| \geq \zeta) \leq g_k(\gamma_1, \theta', c, \zeta) \quad \text{if } (\xi_j) \in \mathcal{H}(\theta), \quad (\text{A.40})$$

with γ_1 , γ_2 and θ' as in (58)-(59). Recall that $y'_k = h'_k y_k$ where $(h'_k) \in \mathcal{E}(\alpha)$ by assumption (56). Moreover, (40) and (41) imply that

$$P(|y_k| \geq \zeta) \leq f_k(2, \gamma, c, \zeta) \quad \text{if } (\xi_k) \in \mathcal{E}(s),$$

$$P(|y_k| \geq \zeta) \leq g_k(2, \theta', c, \zeta) \quad \text{if } (\xi_j) \in \mathcal{H}(\theta).$$

So, (A.39) and (A.40) follow from Lemma C1 (iii) and (iv), respectively.

As shown in the proof of (44) and (45), the relations (A.38)-(A.40) imply

$$P(|s_{T,t;1}^{(1)}| \geq \zeta) \leq f_H(\gamma_1, \gamma_2, c, \zeta) \quad \text{if } (\xi_k) \in \mathcal{E}(s), \quad (\text{A.41})$$

$$P(|s_{T,t;1}^{(1)}| \geq \zeta) \leq g_H(\gamma_1, \theta', c, \zeta) \quad \text{if } (\xi_k) \in \mathcal{H}(\theta), \quad (\text{A.42})$$

which verifies (A.34)-(A.35) for $s_{T,t;1}^{(1)}$. Next we show that setting $\zeta' = \zeta d_{Ht}$,

$$P(|s_{T,t;1}^{(2)}| \geq \zeta) \leq f_H(\gamma_1, \gamma_2, c, \zeta') \quad \text{if } (\xi_k) \in \mathcal{E}(s), \quad (\text{A.43})$$

$$P(|s_{T,t;1}^{(2)}| \geq \zeta) \leq g_H(\gamma_1, \theta', c, \zeta') \quad \text{if } (\xi_k) \in \mathcal{H}(\theta). \quad (\text{A.44})$$

Together with (A.37)-(A.40), the latter proves (A.34)-(A.35) for $s_{T,t;1}^{(h)}$.

We now prove (A.43)-(A.44). We have

$$P(|s_{T,t;1}^{(2)}| \geq \zeta) = P(d_{Ht} |s_{T,t;1}^{(2)}| \geq d_{Ht} \zeta). \quad (\text{A.45})$$

In view of definition of h'_k , by assumptions (54)-(55),

$$h'_k - h'_{k+1} = h_{t-k} - h_{t-k-1} = \delta_{tk}^{-1/2} \xi_{tk}, \quad \text{for } k = 1, \dots, t-2,$$

and $\delta_{tk} = t - k$ if (54) holds; $\delta_{tk} = T$ if (55) holds, while $(\xi_{tk}) \in \mathcal{E}(\alpha)$ by assumption (56).

Then, with ν_k and y_k as in (A.36), setting $y''_k = \xi_{tk} x_k k^{-1/2} = \xi_{tk} y_k$, we can write

$$(h'_k - h'_{k+1}) x_k = \left(\frac{k}{\delta_{tk}}\right)^{1/2} \frac{\xi_{tk} x_k}{k^{1/2}} = \left(\frac{k \vee H}{\delta_{tk}}\right)^{1/2} \frac{y''_k}{\nu_k}.$$

Hence,

$$\begin{aligned} d_{Ht}s_{T,t;1}^{(2)} &= \sum_{k=1}^{t-2} \frac{d_{Ht}}{H^{1/2}} b_{H,k} (h'_k - h'_{k+1}) x_k \\ &= \sum_{k=1}^{t-2} \tilde{a}_{tk} \left(\frac{y''_k}{\nu_k} \right), \quad \tilde{a}_{tk} = \frac{d_{Ht}}{H^{1/2}} b_{H,k} \left(\frac{k \vee H}{\delta_{tk}} \right)^{1/2}. \end{aligned} \quad (\text{A.46})$$

Next we show that for all t, H, T ,

$$s_{Ht} := \sum_{k=1}^{t-2} |\tilde{a}_{tk}| \leq C < \infty. \quad (\text{A.47})$$

Let (54) hold. Then, by definition, $d_{Ht} = (H \vee t)^{1/2} H^{-1}$, $\delta_{Ht} = |t - k|$, and by (C.17),

$$s_{Ht} = d_{Ht} \sum_{k=1}^{t-2} \frac{b_{H,k}}{H^{1/2}} \left(\frac{k \vee H}{t - k} \right)^{1/2} \leq \left(\frac{H}{H \vee t} \right)^{-1/2} \left(\sum_{j=1}^{t-1} \frac{b_{H,|t-j|}}{H} \left(\frac{|t-j| \vee H}{j} \right)^{1/2} \right) \leq C, \quad t \geq 2.$$

Let (55) holds. Then, $d_{Ht} = T^{1/2} H^{-1}$, $\delta_{Ht} = T$, and by property (43) of $b_{H,k}$,

$$s_{Ht} = \sum_{k=1}^{t-2} \frac{b_{H,k}}{H} \left(\frac{k \vee H}{H} \right)^{1/2} \leq C, \quad t \geq 2, T \geq 2.$$

From (A.45)–(A.47) and Lemma A2 it follows that there exists $\varepsilon > 0$ such that

$$P(|s_{T,t;1}^{(2)}| \geq \zeta) = P(|d_{Ht}s_{T,t;1}^{(2)}| \geq \zeta') \leq \varepsilon^{-1} \max_{1 \leq k \leq t-2} E \left[\frac{|y''_k|}{\zeta' \nu_k} I \left(\frac{|y''_k|}{\zeta' \nu_k} \geq \varepsilon \right) \right]. \quad (\text{A.48})$$

This bound is of the same type as (A.38) for $P(|s_{T,t;1}^{(1)}| \geq \zeta)$. Recall that $y''_k = \xi_{tk} y_k$ and by (56), variables ξ_{tk} have the property $(\xi_{tk}) \in \mathcal{E}(\alpha)$. Hence, (A.48) implies (A.43)–(A.44) by the same argument as in the proof of (A.41)–(A.42) for $s_{T,t;1}^{(1)}$.

The proof of the bounds (A.34)–(A.35) for $s_{T,t;2}^{(h)}$ can be obtained using similar arguments as above for $s_{T,t;1}^{(h)}$. This completes the proof of (A.34)–(A.35) which imply (58)–(59) of Lemma 4 for $P(|S_{T,t}^{(h)}| \geq \zeta)$. \square

Proof of Lemma 5. It suffices to verify (A.34)–(A.35) for $P(|\tilde{S}_{T,t}^{(h)}| \geq \zeta)$. Write

$$\tilde{S}_{T,t}^{(h)} = S_{T,t}^{(h)} + r_{T,t}, \quad r_{T,t} := H^{-1/2} \sum_{k=1}^T b_{H,|t-k|} (h_k E \xi_k - h_t E \xi_t).$$

Since by Lemma 4, $P(|S_{T,t}^{(h)}| \geq \zeta)$ satisfies (58)–(59) and thus (A.34)–(A.35), to establish the corresponding bounds for $P(|\tilde{S}_{T,t}^{(h)}| \geq \zeta)$, it suffices to show that $P(|r_{T,t}| \geq \zeta)$ satisfies (A.34)–(A.35) as well. We will prove that there exists $c_0 > 0$ and $c_1 > 0$ such that

$$P(|r_{T,t}| \geq \zeta) \leq c_0 \exp(-c_1 \zeta^\alpha), \quad \zeta > 0, T \geq 2. \quad (\text{A.49})$$

Since $\alpha > \gamma_1 = 2\alpha/(2 + \alpha)$, (A.49) together with definition (39) of f_t and g_t , implies (A.34)-(A.35) for $P(|r_{T,t}| \geq \zeta)$.

Proof of (A.49). Write

$$P(|r_{T,t}| \geq \zeta) = P(d_{Ht}|r_{T,t}| \geq \zeta'), \quad \zeta' = d_{Ht}\zeta.$$

Case 1. Suppose that $E|\xi_k - E\xi_t|$ and $|h_k - h_t|$ satisfy assumptions ((60), (54)). Then,

$$\begin{aligned} |h_k E\xi_k - h_t E\xi_t| &\leq |h_k(E\xi_k - E\xi_t)| + |E\xi_t| |h_k - h_t| \\ &\leq C \left(\frac{|t-k|}{t \vee k}\right)^{1/2} z_k, \quad z_k = |h_k| + |\xi_{tk}|. \end{aligned} \quad (\text{A.50})$$

Under (54), by definition (57), $d_{Ht} = (t \vee H)^{1/2} H^{-1}$. Hence,

$$d_{Ht}|r_{T,t}| \leq C \sum_{k=1}^T a_{tk} z_k, \quad a_{tk} = \left(\frac{t \vee H}{H}\right)^{1/2} \frac{b_{H,|t-k|}}{H} \left(\frac{|t-k|}{t \vee k}\right)^{1/2}.$$

Applying (C.16) with $\gamma = 1/2$, we get

$$\max_{t=1, \dots, T} \sum_{k=1}^T a_{tk} \leq C < \infty.$$

Hence, by Lemma A2, there exists $\varepsilon > 0$ such that

$$P(d_{Ht}|r_{T,t}| \geq \zeta') \leq \varepsilon^{-1} \max_{1 \leq k \leq T} E \left[\frac{|z_k|}{\zeta'} I\left(\frac{|z_k|}{\zeta'} \geq \varepsilon\right) \right]. \quad (\text{A.51})$$

By assumption (56), $(z_k) \in \mathcal{E}(\alpha)$. Hence, by Lemma C2(i),

$$E[|z_k| I(|z_k| \geq \varepsilon \zeta')] \leq c'_0 \exp(-c'_1 \zeta'^\alpha)$$

which together with (A.51) implies

$$P(d_{Ht}|r_{T,t}| \geq \zeta') \leq c'_0 \zeta'^{-1} \exp(-c'_1 \zeta'^\alpha).$$

Therefore,

$$P(|r_{T,t}| \geq \zeta) \leq c'_0 \exp(-c'_1 \zeta'^\alpha) \quad \text{for } \zeta' \geq 1. \quad (\text{A.52})$$

This bound remains valid for $0 < \zeta' < 1$ if c'_0 is selected such that $c'_0 \exp(-c'_1) \geq 1$. Then,

$$c'_0 \exp(-c'_1 \zeta'^\alpha) \geq c'_0 \exp(-c'_1) \geq 1 \quad \text{for } 0 < \zeta' < 1$$

and, thus, (A.52) holds. This proves (A.49).

Case 2. Suppose that $E|\xi_k - E\xi_t|$ and $|h_k - h_t|$ satisfy assumptions ((61), (55)). Then, instead of (A.50), we have the bound

$$|h_k E\xi_k - h_t E\xi_t| \leq C \left(\frac{|t-k|}{T}\right)^{1/2} z_k, \quad z_k = |h_k| + |\xi_{tk}|. \quad (\text{A.53})$$

Under (55), by definition (57), $d_{Ht} = T^{1/2}H^{-1}$ for $t = 1, \dots, T$. Hence,

$$d_{Ht}|r_{T,t}| \leq C \sum_{k=1}^T a_{tk}^* z_k, \quad a_{tk}^* = \left(\frac{T}{H}\right)^{1/2} \frac{b_{H,|t-k|}}{H} \left(\frac{|t-k|}{T}\right)^{1/2} = \frac{b_{H,|t-k|}}{H} \left(\frac{|t-k|}{H}\right)^{1/2}.$$

By the same argument as in (A.28) it follows that

$$\max_{t=1, \dots, T} \sum_{k=1}^T a_{tk}^* \leq C < \infty.$$

Hence, as above, by Lemma A2, there exists $\varepsilon > 0$ such that (A.51) holds, and using the same argument as in Case 1, we obtain (A.49).

Thus, $P(|r_{T,t}| \geq \zeta)$ satisfies (A.49) which completes the proof of the lemma. \square

Proof of Corollary 8. (a) Recall that $\zeta \wedge \zeta' = \zeta(1 \wedge d_{Ht})$. The bounds (58)-(59) together with definitions of f_t, g_t in (39) imply

$$P((1 \wedge d_{Ht})|S_{T,t}^{(h)}| \geq b) \rightarrow 0, \quad b \rightarrow \infty.$$

This proves (62):

$$S_{T,t}^{(h)} = O_P((1 \wedge d_{Ht})^{-1}) = O_P(1 + d_{Ht}^{-1}).$$

The same argument implies $\tilde{S}_{T,t}^{(h)} = O_P(1 + d_{Ht}^{-1})$, since by Lemma 5, $P(|\tilde{S}_{T,t}^{(h)}| \geq \zeta)$ satisfies the same bounds (58)-(59).

(b) Under assumption (55), $d_{Ht} = T^{1/2}H^{-1}$. Set $z_{T,t} := (1 \wedge d_{Ht})\tilde{S}_{T,t}^{(h)}$.

Assume that $(\xi_j) \in \mathcal{E}(s)$, $s > 0$. We will show that as $T \rightarrow \infty$, $b \rightarrow \infty$,

$$P\left(\max_{t=1, \dots, T} |z_{T,t}| > b\delta_{T,H}\right) = o_P(1), \quad \delta_{T,H} = (\log T)^{1/\gamma_1} + \frac{(\log H)^2}{H^{1/2}} (\log T)^{1/\gamma_2}, \quad (\text{A.54})$$

where γ_1 and γ_2 are the same as in (58) of Lemma 4.

For $b > 0$, by (58), definition of f_t , (39), and equality $(\zeta \wedge \zeta')(1 \wedge d_{Ht})^{-1} = \zeta$,

$$\begin{aligned} P(\max_{t=1, \dots, T} |z_{T,t}| \geq b\delta_{T,H}) &\leq \sum_{t=1}^T P((1 \wedge d_{Ht})|S_{T,t}| \geq b\delta_{T,H}) \\ &\leq \sum_{t=1}^T f_H(\gamma_1, \gamma_2, c, b\delta_{T,H}) \\ &\leq Tc_0 \left\{ \exp(-c_1(b\delta_{T,H})^{\gamma_1}) + \exp\left(-c_2\left(\frac{(b\delta_{T,H})\sqrt{H}}{\log^2 H}\right)^{\gamma_2}\right) \right\} \\ &\leq Tc_0 \left\{ \exp(-c_1 b^{\gamma_1} \log T) + \exp(-c_2 b^{\gamma_2} \log T) \right\} \\ &\leq 2c_0 T^{-1} \rightarrow 0 \end{aligned}$$

for b such that $c_1 b^{\gamma_1} \geq 2$, $c_2 b^{\gamma_2} \geq 2$. This proves (A.54). Since for $cT^\delta \leq H \leq T$ it holds $\delta_{T,H} = O(\log^{1/\gamma_1} T)$, (A.54) implies:

$$\max_{t=1,\dots,T} |z_{T,t}| = O_P(\delta_{T,H}) = O_P(\log^{1/\gamma_1} T).$$

This together with definition $z_{T,t} := (1 \wedge d_{Ht}) \tilde{S}_{T,t}^{(h)}$, where $d_{Ht} = T^{1/2} H^{-1}$, and inequality $(1 \wedge T^{1/2} H^{-1})^{-1} \leq 1 + HT^{-1/2}$, implies (63):

$$\max_{t=1,\dots,T} |\tilde{S}_{T,t}^{(h)}| = O_P((1 \wedge T^{1/2} H^{-1})^{-1} \log^{1/\gamma_1} T) = O_P((1 + HT^{-1/2}) \log^{1/\gamma_1} T).$$

Next, consider the case $(\xi_j) \in \mathcal{H}(\theta)$, $\theta > 2$. First we show that for any $\theta' \in (2, \theta)$, as $T \rightarrow \infty$,

$$\max_{t=1,\dots,T} |z_{T,t}| = O_P(\delta_{T,H}), \quad \delta_{T,H} = (\log T)^{1/\gamma_1} + H^{1/2} \left(\frac{T}{H^{\theta'-1}} \right)^{1/\theta'}. \quad (\text{A.55})$$

For $b > 0$, by (59), definition of g_t , (39) and equality $(\zeta \wedge \zeta')(1 \wedge d_{Ht})^{-1} = \zeta$, we obtain

$$\begin{aligned} & P(\max_{t=1,\dots,T} |z_{T,t}| \geq b\delta_{T,H}) \\ & \leq \sum_{t=1}^T P((1 \wedge d_{Ht}) |S_{T,t}| \geq b\delta_{T,H}) \\ & \leq \sum_{t=1}^T g_H(\gamma_1, \theta', c, b\delta_{T,H}) \\ & \leq T c_0 \left\{ \exp(-c_1 (b\delta_{T,H})^{\gamma_1}) + (b\delta_{T,H})^{-\theta'} H^{-(\theta'/2-1)} \right\} \\ & \leq c_0 \left\{ T \exp(-c_1 b^{\gamma_1} \log T) + b^{-\theta'} \left(\frac{T}{H^{\theta'/2-1}} \right)^{-1} T H^{-(\theta'/2-1)} \right\} \\ & \leq c_0 \{ T^{-1} + b^{-\theta'} \} \rightarrow 0 \end{aligned}$$

as $T \rightarrow \infty$ and $b \rightarrow \infty$. This proves (A.55). The same argument as in the proof of (51) of Corollary 6 shows that validity of (A.55) for any $2 < \theta' < \theta$ implies that for any $\varepsilon > 0$,

$$\max_{t=1,\dots,T} |z_{T,t}| = O_P(\tilde{\delta}_{T,H}), \quad \tilde{\delta}_{T,H} = (\log T)^{1/\gamma_1} + H^{\varepsilon-1/2} (TH)^{1/\theta}. \quad (\text{A.56})$$

In turn, since $d_{Ht} = T^{1/2} H^{-1}$, this yields

$$\begin{aligned} \max_{t=1,\dots,T} |\tilde{S}_{T,t}^{(h)}| &= O_P((1 \wedge H^{-1} T^{1/2})^{-1} \tilde{\delta}_{T,H}) \\ &= O_P((1 + HT^{-1/2}) \{ (\log T)^{1/\gamma_1} + H^{\varepsilon-1/2} (TH)^{1/\theta} \}) \end{aligned}$$

which proves (64) and completes the proof of the corollary. \square

Proof of Corollary 9. Denote

$$v_{T,t,\nu} = H^{-1} \sum_{k=1}^T b_{H,|t-k|} \left| \frac{t-k}{H} \right|^\nu |\xi_k|, \quad 0 \leq \nu \leq 1. \quad (\text{A.57})$$

Proof of (67) and (69). Let (65) hold. Then,

$$|\Delta_{T,t}| \leq CH^{-1} \sum_{k=1}^T b_{H,|t-k|} \left| \frac{t-k}{T} \right| |\xi_k| = C(H/T)v_{T,t,1}. \quad (\text{A.58})$$

Since under assumptions of lemma, $\max_k E|\xi_k| < \infty$, together with (43) this implies

$$Ev_{T,t,1} \leq C(\max_k E|\xi_k|)H^{-1} \sum_{k=1}^T b_{H,|t-k|} \left| \frac{t-k}{H} \right| = O(1).$$

Hence, $v_{T,t,1} = O_P(1)$ which together with (A.58) proves (67):

$$|\Delta_{T,t}| = C(H/T)O_P(1) = O_P(H/T).$$

Notice that by Corollary 7, under the assumptions of Corollary 9(b),

$$\max_{1 \leq s \leq T} |v_{T,s,\nu}| = O_P(1), \quad 0 \leq \nu \leq 1. \quad (\text{A.59})$$

This together with (A.58) proves (69):

$$\max_{1 \leq s \leq T} |\Delta_{T,t}| \leq C(H/T) \max_{1 \leq s \leq T} |v_{T,s,1}| = O_P(H/T).$$

Proof of (68) and (70). Let (66) hold. Then

$$|\Delta_{T,t}| \leq CH^{-1} \sum_{k=1}^T b_{H,|t-k|} \left(\frac{t-k}{T} \right)^{1/2} |\nu_{tk}\xi_k|. \quad (\text{A.60})$$

By (66), $(\nu_{tk}) \in \mathcal{E}(\alpha)$, $\alpha > 0$, while by assumption of corollary, $(\xi_j) \in \mathcal{E}(s)$, $s > 0$ or $(\xi_j) \in \mathcal{H}(\theta)$, $\theta > 2$. Thus, from Lemma C1 (i)-(ii) it follows $\max_{tk} E|\nu_{tk}\xi_k| < \infty$. Hence,

$$E|\Delta_{T,t}| \leq C(H/T)^{1/2} (\max_{tk} E|\nu_{tk}\xi_k|) H^{-1} \sum_{k=1}^T b_{H,|t-k|} \left(\frac{t-k}{H} \right)^{1/2} \leq C(H/T)^{1/2},$$

where $C > 0$ does not depend on t, H, T . This proves (68), $\Delta_{T,t} = O_P((H/T)^{1/2})$.

Next, by (A.60),

$$\max_{1 \leq t \leq T} |\Delta_{T,t}| \leq C(H/T)^{1/2} (\max_{1 \leq k, t \leq T} |\nu_{tk}|) (\max_{1 \leq t \leq T} v_{T,s,1/2}), \quad (\text{A.61})$$

where $v_{T,s,1/2}$ is defined by (A.57). Since $(\nu_{tk}) \in \mathcal{E}(\alpha)$, (C.3) of Lemma C1 implies:

$$\max_{1 \leq k, t \leq T} |\nu_{tk}| = O_P((\log T)^{1/\alpha}).$$

By (A.59),

$$\max_{1 \leq t \leq T} |v_{T,s,1/2}| = O_P(1)$$

which together with (A.61) proves (70):

$$\max_{1 \leq t \leq T} |\Delta_{T,t}| = O_P((H/T)^{1/2} (\log T)^{1/\alpha}).$$

This completes the proof of the corollary. \square

B. Proofs of Theorems 1-3.

For convenience of the proof of Theorems 1-3, we include Lemma B1 which summarizes the key steps of the proof of Theorem 1, Bickel and Levina (2008) and adjusts them to our setting.

Recall notation of $p \times p$ covariance matrix $\Sigma_t = [\sigma_{ij,t}]$, sample covariance estimator $\widehat{\Sigma}_t = [\widehat{\sigma}_{ij,t}]$ of Σ_t , (10), and the regularized sample covariance estimate defined in (11):

$$T_\lambda(\widehat{\Sigma}_t) = (\widehat{\sigma}_{ij,t} I(|\widehat{\sigma}_{ij,t}| > \lambda)).$$

Denote

$$M = \max_{i,j=1,\dots,p} |\widehat{\sigma}_{ij,t} - \sigma_{ij,t}|, \quad N = \max_{i=1,\dots,p} \sum_{j=1}^p I(|\widehat{\sigma}_{ij,t} - \sigma_{ij,t}| > \lambda/2).$$

Recall the definition of the sparsity parameter n_p of covariance matrix Σ_t which is the maximum number of non-zero elements in a row of Σ_t , see, e.g., (8).

Lemma B1 (see Bickel and Levina (2008, proof of Theorem 1)). For any $\lambda > 0$,

$$\|T_\lambda(\widehat{\Sigma}_t) - \Sigma_t\| \leq 2MN + Mn_p + 2\lambda n_p. \quad (\text{B.1})$$

Moreover, if λ is such that as $T \rightarrow \infty$,

$$\max_{i,j=1,\dots,p} P(|\widehat{\sigma}_{ij,t} - \sigma_{ij,t}| > \lambda/2) = o(p^{-2}), \quad (\text{B.2})$$

then

$$\|T_\lambda(\widehat{\Sigma}_t) - \Sigma_t\| = O_P(n_p \lambda). \quad (\text{B.3})$$

In addition, if (B.3) holds, $n_p \lambda = o(1)$ and $\|\Sigma_t\| \geq c > 0$, then

$$\|T_\lambda(\widehat{\Sigma}_t)^{-1} - \Sigma_t^{-1}\| = O_P(n_p \lambda). \quad (\text{B.4})$$

Proof. Verification of (B.1) follows closely the steps of the proof of Theorem 1, pp. 2582-2584 in Bickel and Levina (2008). For clarity, we include the details of the proof.

We have

$$T_\lambda(\widehat{\Sigma}_t) - \Sigma_t = [\delta_{ij,t}]_{i,j=1,\dots,p}, \quad \delta_{ij,t} = \widehat{\sigma}_{ij,t} I(|\widehat{\sigma}_{ij,t}| > \lambda) - \sigma_{ij,t}.$$

By the well-known property of the spectral norm of a symmetric matrix,

$$\|T_\lambda(\widehat{\Sigma}_t) - \Sigma_t\| \leq \max_{i=1,\dots,p} \left(\sum_{j=1}^p |\delta_{ij,t}| \right). \quad (\text{B.5})$$

Write

$$\begin{aligned}\delta_{ij,t} &= \{\widehat{\sigma}_{ij,t}I(|\widehat{\sigma}_{ij,t}| > \lambda) - \sigma_{ij,t}I(|\sigma_{ij,t}| > \lambda/2)\} + \{-\sigma_{ij,t}I(|\sigma_{ij,t}| \leq \lambda/2)\} \\ &= \delta_{ij,t}^{(1)} + \delta_{ij,t}^{(2)}.\end{aligned}$$

Notice that

$$|\delta_{ij,t}^{(2)}| \leq |\sigma_{ij,t}|I(|\sigma_{ij,t}| \leq \lambda/2) \leq (\lambda/2)I(|\sigma_{ij,t}| \neq 0).$$

On the other hand,

$$\begin{aligned}\delta_{ij,t}^{(1)} &= \widehat{\sigma}_{ij,t}(I(|\widehat{\sigma}_{ij,t}| > \lambda) - I(|\sigma_{ij,t}| > \lambda/2)) + (\widehat{\sigma}_{ij,t} - \sigma_{ij,t})I(|\sigma_{ij,t}| > \lambda/2) \\ &= \{\widehat{\sigma}_{ij,t}I(|\widehat{\sigma}_{ij,t}| > \lambda, |\sigma_{ij,t}| \leq \lambda/2)\} + \{-\widehat{\sigma}_{ij,t}I(|\widehat{\sigma}_{ij,t}| \leq \lambda, |\sigma_{ij,t}| > \lambda/2)\} \\ &\quad + \{(\widehat{\sigma}_{ij,t} - \sigma_{ij,t})I(|\sigma_{ij,t}| > \lambda/2)\} = v_{ij,t}^{(1)} + v_{ij,t}^{(2)} + v_{ij,t}^{(3)}.\end{aligned}$$

Notice that for $|\widehat{\sigma}_{ij,t}| > \lambda$, $|\sigma_{ij,t}| \leq \lambda/2$ it holds

$$\begin{aligned}|\widehat{\sigma}_{ij,t}| &\leq 2(|\widehat{\sigma}_{ij,t}| - |\sigma_{ij,t}|) \leq 2|\widehat{\sigma}_{ij,t} - \sigma_{ij,t}|, \\ |\widehat{\sigma}_{ij,t} - \sigma_{ij,t}| &\geq |\widehat{\sigma}_{ij,t}| - |\sigma_{ij,t}| > \lambda/2.\end{aligned}$$

Hence,

$$\begin{aligned}|v_{ij,t}^{(1)}| &\leq 2|\widehat{\sigma}_{ij,t} - \sigma_{ij,t}|I(|\widehat{\sigma}_{ij,t} - \sigma_{ij,t}| > \lambda/2), \\ |v_{ij,t}^{(2)}| &\leq \lambda I(|\sigma_{ij,t}| \neq 0), \\ |v_{ij,t}^{(3)}| &\leq |\widehat{\sigma}_{ij,t} - \sigma_{ij,t}|I(|\sigma_{ij,t}| \neq 0).\end{aligned}$$

Therefore,

$$\begin{aligned}|\delta_{ij,t}| &\leq |\delta_{ij,t}^{(1)}| + |\delta_{ij,t}^{(2)}| \leq |v_{ij,t}^{(1)}| + |v_{ij,t}^{(2)}| + |v_{ij,t}^{(3)}| + |\delta_{ij,t}^{(2)}| \\ &\leq 2|\widehat{\sigma}_{ij,t} - \sigma_{ij,t}|I(|\widehat{\sigma}_{ij,t} - \sigma_{ij,t}| > \lambda/2) + |\widehat{\sigma}_{ij,t} - \sigma_{ij,t}|I(|\sigma_{ij,t}| \neq 0) + 2\lambda I(|\sigma_{ij,t}| \neq 0).\end{aligned}$$

Note that by definition of the sparsity parameter, $\max_{i=1,\dots,p} \sum_{j=1}^p I(|\sigma_{ij,t}| \neq 0) = n_p$. Applying this in (B.5), we obtain

$$\begin{aligned}\|T_\lambda(\widehat{\Sigma}_t) - \Sigma_t\| &\leq \max_{i=1,\dots,p} \left(\sum_{j=1}^p |\delta_{ij,t}| \right) \\ &\leq 2 \left\{ \max_{i,j=1,\dots,p} |\widehat{\sigma}_{ij,t} - \sigma_{ij,t}| \right\} \left\{ \max_{i=1,\dots,p} \sum_{j=1}^p I(|\widehat{\sigma}_{ij,t} - \sigma_{ij,t}| > \lambda/2) \right\} \\ &\quad + \left\{ \max_{i,j=1,\dots,p} |\widehat{\sigma}_{ij,t} - \sigma_{ij,t}| \right\} \left\{ \max_i \sum_{j=1}^p I(|\sigma_{ij,t}| \neq 0) \right\} + 2\lambda \left\{ \max_i \sum_{j=1}^p I(|\sigma_{ij,t}| \neq 0) \right\} \\ &\leq 2M N + M n_p + 2\lambda n_p\end{aligned}$$

which proves (B.1).

Proof of (B.3). If (B.2) holds then

$$P(M \geq \lambda/2) \leq \sum_{i,j=1}^p P(|\hat{\sigma}_{ij,t} - \sigma_{ij,t}| > \lambda/2) \leq p^2 \max_{ij} P(|\hat{\sigma}_{ij,t} - \sigma_{ij,t}| > \lambda/2) = o(1).$$

In turn,

$$P(N > 1) \leq P(N > 0) \leq P(M > \lambda/2) \rightarrow 0.$$

This shows that $M = O_P(\lambda)$ and $N = O_P(1)$ which together with (B.1) proves (B.3).

To prove (B.4), set $B := T_\lambda(\hat{\Sigma}_t)$, $A := \Sigma_t$. By assumption, $\|A\| \geq c > 0$ and $n_p \lambda = o(1)$. By (B.3), $\|B - A\| = O_P(n_p \lambda) = o_P(1)$. Thus,

$$\|B\| \geq \|A + (B - A)\| \geq \|A\| - \|B - A\| \geq c - o_P(1) \geq c(1 + o_P(1)).$$

This implies $\|B^{-1}\| = O_P(1)$. Hence,

$$\begin{aligned} \|B^{-1} - A^{-1}\| &= \|A^{-1}(A - B)B^{-1}\| \leq \|A^{-1}\| \|A - B\| \|B^{-1}\| \\ &\leq c^{-1} O_P(n_p \lambda) O_P(1) = O_P(n_p \lambda) \end{aligned}$$

which proves (B.4). \square

Proof of Theorem 1. Recall that

$$\lambda = \kappa(T^{-1} \log p)^{1/2} \tag{B.6}$$

has property $\lambda \rightarrow 0$ as $T \rightarrow \infty$ in view of (9). By assumption of the theorem, (\mathbf{y}_t) is a stationary sequence, the sample covariance matrix $\hat{\Sigma} = (\hat{\sigma}_{ij})$ given by (5) is the estimate of $\Sigma = (\sigma_{ij}) = \text{var}(\mathbf{y}_t)$ and σ_{ij} does not depend on t .

By Lemma B1, in view of definition (B.6), to show (B.3) and thus, the claim (6) of Theorem 1, it suffices to prove that for sufficiently large κ ,

$$\max_{i,j=1,\dots,p} P(|\hat{\sigma}_{ij} - \sigma_{ij}| > 2\lambda) = o(p^{-2}). \tag{B.7}$$

(Notice that (B.7) implies that (B.2) holds for sufficiently large κ which in turn proves (B.3).)

Fix (i, j) and set $z_k = y_{ik}y_{jk}$. Because of stationarity assumption, $E y_{ik} = E y_{i1}$, $E y_{jk} = E y_{j1}$ and $E z_k = \sigma_{ij}$ do not depend on k . Observe that

$$\sigma_{ij} = \text{cov}(y_{ik}, y_{jk}) = E z_k - E y_{ik} E y_{jk}.$$

Then we can write

$$\begin{aligned}\widehat{\sigma}_{ij} - \sigma_{ij} &= T^{-1} \sum_{k=1}^T y_{ik} y_{jk} - \bar{y}_i \bar{y}_j - \sigma_{ij} \\ &= s_{T,ij} - \bar{y}_i \bar{y}_j + E y_{i1} E y_{j1}, \quad s_{T,ij} := T^{-1} \sum_{k=1}^T (z_k - E z_k), \quad \bar{y}_i = T^{-1} \sum_{k=1}^T y_{ik}.\end{aligned}$$

Observe that

$$\bar{y}_i \bar{y}_j - E y_{i1} E y_{j1} = (\bar{y}_i - E y_{i1})(\bar{y}_j - E y_{j1}) + E[y_{i1}](\bar{y}_j - E y_{j1}) + E[y_{j1}](\bar{y}_i - E y_{i1}).$$

Both assumptions $(y_{ik}) \in \mathcal{E}(s)$ and $(y_{ik}) \in \mathcal{H}(\theta)$ imply that $m = \max_{i,k} E|y_{ik}| < \infty$.

Therefore,

$$|\widehat{\sigma}_{ij} - \sigma_{ij}| \leq |s_{T,ij}| + |\bar{y}_i - E y_{i1}| |\bar{y}_j - E y_{j1}| + m |\bar{y}_j - E y_{j1}| + m |\bar{y}_i - E y_{i1}|.$$

So, we obtain

$$\begin{aligned}P(|\widehat{\sigma}_{ij} - \sigma_{ij}| > 4\lambda) &\leq P(|s_{T,ij}| > \lambda) + P(|\bar{y}_i - E y_{i1}| |\bar{y}_j - E y_{j1}| > \lambda) \\ &\quad + P(m |\bar{y}_j - E y_{j1}| > \lambda) + P(m |\bar{y}_i - E y_{i1}| > \lambda).\end{aligned}\tag{B.8}$$

Since $\lambda = o(1)$ as $T \rightarrow \infty$, then $\sqrt{\lambda} \geq \lambda$ for $\lambda \leq 1$. Hence,

$$\begin{aligned}P(|\bar{y}_i - E y_{i1}| |\bar{y}_j - E y_{j1}| > \lambda) & \\ &\leq P(|\bar{y}_i - E y_{i1}| > \sqrt{\lambda}) + P(|\bar{y}_j - E y_{j1}| > \sqrt{\lambda}) \\ &\leq P(|\bar{y}_i - E y_{i1}| > \lambda) + P(|\bar{y}_j - E y_{j1}| > \lambda).\end{aligned}\tag{B.9}$$

Therefore, to prove (B.7), it suffices to show that uniformly in i, j , as $T \rightarrow \infty$,

$$\max_{i,j=1,\dots,p} P(|s_{T,ij}| > \lambda) = o(p^{-2}), \quad \max_{i=1,\dots,p} P(|\bar{y}_i - E y_{i1}| > \lambda) = o(p^{-2}),\tag{B.10}$$

$$\max_{i=1,\dots,p} P(m |\bar{y}_i - E y_{i1}| > \lambda) = o(p^{-2}),\tag{B.11}$$

when κ is selected sufficiently large. We will prove (B.10), while (B.11) can be shown using the same argument as in the proof of the second claim in (B.10).

Denote

$$S_{T,ij}^* = T^{1/2} s_{T,ij}, \quad S_{T,i}^* = T^{1/2} \bar{y}_i = T^{-1/2} \sum_{k=1}^T (y_{ik} - E y_{ik}).$$

Then, with $\eta = T^{1/2} \lambda$,

$$\begin{aligned}P(|s_{T,ij}| > \lambda) &\leq P(|S_{T,ij}^*| > \eta), \\ P(|\bar{y}_i - E y_{i1}| > \lambda) &\leq P(|S_{T,i}^*| \geq \eta).\end{aligned}\tag{B.12}$$

By Assumption M, the process $(\mathbf{y}_k - E\mathbf{y}_k)$ is α -mixing, and therefore processes $(z_k - Ez_k)$, $(y_{ik} - Ey_{ik})$ are also α -mixing with mixing coefficients satisfying (1).

(i) Let $(y_{ik}) \in \mathcal{E}(s)$. Then $(z_k) \in \mathcal{E}(s/2)$, and $Ez_k = \sigma_{ij}$ does not depend on k . Hence, (40) of Lemma 1 implies that with $\gamma = (s/2)(1 + s/2)$,

$$\begin{aligned} P(|S_{T,ij}^*| > \eta) &\leq f_T(2, \gamma, c, \eta), \\ P(|S_{T,i}^*| > \eta) &\leq f_T(2, \gamma, c, \eta). \end{aligned} \tag{B.13}$$

Notice that $\eta = \kappa(\log p)^{1/2}$. Then, by definition of f_t in (39),

$$\begin{aligned} f_T(2, \gamma, c, \eta) &= c_0 \left\{ \exp(-c_1 \eta^2) + \exp\left(-c_2 \left(\frac{\eta T^{1/2}}{\log^2 T}\right)^\gamma\right) \right\} \\ &= c_0 \left\{ \exp(-c_1 \kappa^2 \log T) + \exp\left(-c_2 \left(\frac{\kappa(\log T)^{1/2} T^{1/2}}{\log^2 T}\right)^\gamma\right) \right\} \\ &= o(p^{-2}) \end{aligned}$$

because $c_1 \kappa^2 > 2$ when κ is chosen large enough, and under assumption (9), $T \geq cp^\varepsilon$,

$$\log p = o\left(\left(\frac{(\log T)^{1/2} T^{1/2}}{\log^2 T}\right)^\gamma\right).$$

This together with (B.13) and (B.12) proves (B.10).

(ii) Let $(y_{ik}) \in \mathcal{H}(\theta)$. Then, $(z_k) \in \mathcal{H}(\theta/2)$, and (41) of Lemma 1 implies

$$\begin{aligned} P(|S_{T,ij}^*| > \eta) &\leq g_T(2, \theta', c, \eta), \quad 2 < \theta' < \theta/2, \\ P(|S_{T,i}^*| > \eta) &\leq g_T(2, \theta', c, \eta). \end{aligned} \tag{B.14}$$

Recall that $\eta = \kappa(\log p)^{1/2}$. Then the function g_T given in (39) has property

$$\begin{aligned} g_T(2, \theta', c, \eta) &= c_0 \left\{ \exp(-c_1 \eta^2) + \eta^{-\theta} T^{-(\theta'/2-1)} \right\} \\ &= c_0 \left\{ \exp(-c_1 \kappa^2 \log p) + (\kappa(\log p)^{1/2})^{-\theta'} T^{-(\theta'/2-1)} \right\} \\ &= o(p^{-2}) \end{aligned}$$

because $c_1 \kappa^2 > 2$ for large enough κ , and since under assumption (9) of the theorem,

$$p^2 = o(T^{\theta'/2-1}) \tag{B.15}$$

if $\theta' \in (2, \theta/2)$ is selected close enough to $\theta/2$. Indeed, then $T \geq c_0 p^\varepsilon$, $\varepsilon > 8/(\theta - 4)$ which implies $p^2 = o(T^{\theta'/2-1})$ if θ' is selected close enough to $\theta/2$.

This, together with (B.14) and (B.12) proves (B.10) which completes the proof of (6).

Property (7) follows using (B.4) of Lemma B1. \square

Proof of Theorem 2. Recall that in Theorem 2

$$\lambda = \kappa(\log p)^{1/2} \max(H^{-1/2}, H/T). \quad (\text{B.16})$$

By Lemma B1, to prove (B.3) which is equivalent to the claim (15) of theorem, it suffices to verify validity of (B.2) when parameter κ is selected sufficiently large. For notational simplicity, instead of (B.2) we will show that for sufficiently large κ ,

$$\max_{i,j=1,\dots,p} P(|\hat{\sigma}_{ij,t} - \sigma_{ij,t}| > 4\lambda) = o(p^{-2}). \quad (\text{B.17})$$

Since κ can be arbitrary selected, (B.17) implies (B.2).

Recall that $\mathbf{y}_t = (y_{1t}, \dots, y_{pt})'$. Set $z_k = y_{ik}y_{jk}$. Notice that

$$\sigma_{ij,k} = \text{cov}(y_{ik}, y_{jk}) = Ez_k - Ey_{ik}Ey_{jk}.$$

Then,

$$\begin{aligned} \hat{\sigma}_{ij,t} - \sigma_{ij,t} &= K_t^{-1} \sum_{k=1}^T b_{H,|t-k|} y_{ik}y_{jk} - \bar{y}_{it}\bar{y}_{jt} - \sigma_{ij,t} \\ &= s_{T,ij,t} - \bar{y}_{it}\bar{y}_{jt} + Ey_{it}Ey_{jt}, \\ s_{T,ij,t} &= K_t^{-1} \sum_{k=1}^T b_{H,|t-k|} (z_k - Ez_t), \quad \bar{y}_{it} = K_t^{-1} \sum_{k=1}^T b_{H,|t-k|} y_{ik}. \end{aligned} \quad (\text{B.18})$$

Notice that

$$\bar{y}_{it}\bar{y}_{jt} - Ey_{it}Ey_{jt} = (\bar{y}_{it} - Ey_{it})(\bar{y}_{jt} - Ey_{jt}) + E[y_{it}](\bar{y}_{jt} - Ey_{jt}) + E[y_{jt}](\bar{y}_{it} - Ey_{it}).$$

Under assumption $(y_{ik}) \in \mathcal{E}(s)$ or $(y_{ik}) \in \mathcal{H}(\theta)$, $\max_{i,t} |Ey_{it}| \leq m < \infty$. Hence,

$$|\hat{\sigma}_{ij,t} - \sigma_{ij,t}| \leq |s_{T,ij,t}| + |\bar{y}_{it} - Ey_{it}| |\bar{y}_{jt} - Ey_{jt}| + m |\bar{y}_{jt} - Ey_{jt}| + m |\bar{y}_{it} - Ey_{it}|.$$

Therefore,

$$\begin{aligned} P(|\hat{\sigma}_{ij,t} - \sigma_{ij,t}| > 4\lambda) &\leq P(|s_{T,ij,t}| > \lambda) + P(|\bar{y}_{it} - Ey_{it}| |\bar{y}_{jt} - Ey_{jt}| > \lambda) \\ &\quad + P(m |\bar{y}_{jt} - Ey_{jt}| > \lambda) + P(m |\bar{y}_{it} - Ey_{it}| > \lambda). \end{aligned} \quad (\text{B.19})$$

Notice that $\lambda = o(1)$ as $T \rightarrow \infty$ by (14). Hence, $\sqrt{\lambda} \geq \lambda$ for $\lambda \leq 1$. So,

$$\begin{aligned} P(|\bar{y}_{it} - Ey_{it}| |\bar{y}_{jt} - Ey_{jt}| > \lambda) & \\ &\leq P(|\bar{y}_{it} - Ey_{it}| > \sqrt{\lambda}) + P(|\bar{y}_{jt} - Ey_{jt}| > \sqrt{\lambda}) \\ &\leq P(|\bar{y}_{it} - Ey_{it}| > \lambda) + P(|\bar{y}_{jt} - Ey_{jt}| > \lambda). \end{aligned} \quad (\text{B.20})$$

Therefore, to prove (B.17), it suffices to show that uniformly in i, j , as $T \rightarrow \infty$,

$$\max_{i,j=1,\dots,p} P(|s_{T,ij,t}| > \lambda) = o(p^{-2}), \quad \max_{i=1,\dots,p} P(|\bar{y}_{it} - Ey_{it}| > \lambda) = o(p^{-2}), \quad (\text{B.21})$$

$$\max_{i=1,\dots,p} P(m|\bar{y}_{it} - Ey_{it}| > \lambda) = o(p^{-2}). \quad (\text{B.22})$$

We will prove (B.21). ((B.22) can be shown using the same argument as in the proof of the second claim in (B.21)). Write

$$\begin{aligned} s_{T,ij,t} &= H^{1/2} K_t^{-1} \left(H^{-1/2} \sum_{k=1}^T b_{H,|t-k|} (z_k - Ez_k) + H^{-1/2} \sum_{k=1}^T b_{H,|t-k|} (Ez_k - Ez_t) \right) \\ &=: H^{1/2} K_t^{-1} (s_{T,ij,t}^* + r_{T,ij,t}), \\ \bar{y}_{it} - Ey_{it} &= H^{1/2} K_t^{-1} \left(H^{-1/2} \sum_{k=1}^T b_{H,|t-k|} (y_{ik} - Ey_{ik}) + H^{-1/2} \sum_{k=1}^T b_{H,|t-k|} (Ey_{ik} - Ey_{it}) \right) \\ &=: H^{1/2} K_t^{-1} (s_{T,i,t}^* + r_{T,i,t}). \end{aligned}$$

Observe that there exists $a_1, a_2 > 0$ such that for all $1 \leq t \leq T$, $T \geq 1$,

$$a_1 H \leq K_t \leq a_2 H.$$

Then

$$(K_t/H^{1/2})\lambda \geq a_1 H^{1/2} \lambda =: \eta. \quad (\text{B.23})$$

Therefore

$$\begin{aligned} P(|s_{T,ij,t}| > \lambda) &\leq P(|s_{T,ij,t}^* + r_{T,ij,t}| > \eta) \leq P(|s_{T,ij,t}^*| > \eta - |r_{T,ij,t}|), \\ P(|\bar{y}_{it} - Ey_{it}| > \lambda) &\leq P(|s_{T,i,t}^* + r_{T,i,t}| > \eta) \leq P(|s_{T,i,t}^*| > \eta - |r_{T,i,t}|). \end{aligned}$$

First we show that, as $p \rightarrow \infty$,

$$|r_{T,ij,t}| \leq \eta/2, \quad |r_{T,i,t}| \leq \eta/2 \quad (\text{B.24})$$

which implies

$$\begin{aligned} P(|s_{T,ij,t}| > \lambda) &\leq P(|s_{T,ij,t}^*| \geq \eta/2), \\ P(|\bar{y}_{it} - Ey_{it}| > \lambda) &\leq P(|s_{T,i,t}^*| \geq \eta/2). \end{aligned} \quad (\text{B.25})$$

To verify (B.24), we use the equality $Ez_t = E[y_{it}y_{jt}] = \text{cov}(y_{it}, y_{jt}) + Ey_{it}Ey_{jt}$ which together with assumption (2) implies that uniformly in i, t, s ,

$$\begin{aligned} |Ey_{it}| &\leq C, \quad |Ey_{it} - Ey_{is}| \leq C \frac{|t-s|}{t \vee s}, \\ |Ez_t| &\leq C, \quad |Ez_t - Ez_s| \leq C \frac{|t-s|}{t \vee s}. \end{aligned} \quad (\text{B.26})$$

This together with (46) of Lemma 3 and the assumption of the theorem, $\delta T \leq t \leq T$, yields

$$\begin{aligned} |r_{T,ij,t}| &\leq C_* \frac{H^{3/2}}{H \vee t} \leq C_* \frac{H^{3/2}}{H \vee \delta T} \leq C_* \frac{H^{3/2}}{\delta T}, \\ |r_{T,i,t}| &\leq C_* \frac{H^{3/2}}{\delta T} \end{aligned} \quad (\text{B.27})$$

because $H = o(T)$ by the assumption (14). Since

$$\lambda\sqrt{H} \geq \kappa(\log p)^{1/2} H^{3/2}/T,$$

this implies

$$\frac{\eta}{|r_{T,ij,t}|} \geq \frac{a_1 \kappa (\log p)^{1/2}}{C_* / \delta} > 2, \quad \frac{\eta}{|r_{T,i,t}|} > 2$$

when $\kappa(\log p)^{1/2}$ is sufficiently large. This proves (B.24) and (B.25).

By Assumption M, the process $(\mathbf{x}_t - E\mathbf{x}_t)$ is α -mixing, and therefore $(z_t - Ez_t)$ is also α -mixing with mixing coefficients satisfying (1).

(i) Let $(\mathbf{y}_k) \in \mathcal{E}(s)$. Then, $(z_k) \in \mathcal{E}(s/2)$ and $(y_{ik}) \in \mathcal{E}(s/2)$. So, applying (44) of Lemma 2 we obtain

$$\begin{aligned} P(|s_{T,ij,t}^*| > \eta/2) &\leq f_H(2, \gamma, c, \eta/2), \quad \gamma = (s/2)(1 + s/2), \\ P(|s_{T,i,t}^*| > \eta/2) &\leq f_H(2, \gamma, c, \eta/2). \end{aligned} \quad (\text{B.28})$$

The function

$$f_H(\gamma_1, \gamma_2, c, \zeta) \leq c_0 \left\{ \exp(-c_1 \zeta^{\gamma_1}) + \exp\left(-c_2 \left(\frac{\zeta H^{1/2}}{\log^2 H}\right)^{\gamma_2}\right) \right\} \quad (\text{B.29})$$

given in (39) is non-increasing in ζ . By (B.23),

$$\eta/2 \geq (a_1/2)\kappa(\log p)^{1/2}. \quad (\text{B.30})$$

Thus,

$$\begin{aligned} f_H(2, \gamma, c, \eta/2) & \\ &\leq c_0 \left\{ \exp\left(-c_1 (a_1/2)^2 \kappa^2 \log p\right) + \exp\left(-c_2 \left((a_1/2)\kappa \log^{1/2} p \frac{H^{1/2}}{\log^2 H}\right)^\gamma\right) \right\} = o(p^{-2}) \end{aligned} \quad (\text{B.31})$$

because $c_1(a_1/2)^2 \kappa^2 > 2$ when κ is chosen large enough, and by (14), $H \geq c_0 p^\varepsilon$, which implies

$$\log p = o\left(\left(\log^{1/2} p \frac{H^{1/2}}{\log^2 H}\right)^\gamma\right).$$

This together with (B.28) and (B.25) proves (B.21).

(ii) Let $(\mathbf{y}_k) \in \mathcal{H}(\theta)$. Then, $(z_k) \in \mathcal{H}(\theta/2)$ and $(y_{ik}) \in \mathcal{H}(\theta/2)$, and using (45) of Lemma 2, we obtain

$$\begin{aligned} P(|s_{T,ij,t}^*| > \eta/2) &\leq g_H(2, \theta', c, \eta/2), \quad 2 < \theta' < \theta/2, \\ P(|s_{T,i,t}^*| > \eta/2) &\leq g_H(2, \theta', c, \eta/2). \end{aligned} \quad (\text{B.32})$$

The function

$$g_H(\gamma, \theta', c, \zeta) = c_0 \left\{ \exp(-c_1 \zeta^\gamma) + \zeta^{-\theta} t^{-(\theta'/2-1)} \right\} \quad (\text{B.33})$$

given in (39) is non-increasing in ζ . Again, using the bound $\eta/2 \geq (a_1/2)\kappa(\log p)^{1/2}$, we obtain

$$\begin{aligned} g_H(2, \theta', c, \eta/2) & \\ &\leq c_0 \left\{ \exp(-c_1(a_1/2)^2 \kappa^2 \log p) + ((a_1/2)\kappa(\log p)^{1/2})^{-\theta'} H^{-(\theta'/2-1)} \right\} \\ &= o(p^{-2}) \end{aligned} \quad (\text{B.34})$$

because $c_1(a_1/2)^2 \kappa^2 > 2$ for large enough κ and because $p^2 = o(H^{\theta'/2-1})$ under the assumption (14) of the theorem if $\theta' \in (2, \theta/2)$ is selected close enough to $\theta/2$, see the proof of (B.15). Clearly, (B.34), (B.32) and (B.25) prove (B.21).

This completes the proof of the claim (15) of theorem.

The claim (16) of the theorem is shown in (B.4) of Lemma B1.

The bandwidth $H_{opt} = T^{2/3}$ minimizes $\max(H^{-1/2}, (H/T))$, so

$$\lambda = \kappa(\log p)^{1/2} \max(H^{-1/2}, (H/T)) \geq \lambda_{opt} = \kappa(\log p)^{1/2} T^{-1/3}$$

which proves the last claim of the theorem. \square

Proof of Theorem 4. In this theorem,

$$\lambda = \kappa(\log p)^\nu \max(H^{-1/2}, (H/T)^{1/2}), \quad \nu = \frac{\alpha + 4}{2\alpha}. \quad (\text{B.35})$$

Notice that by (19), $\lambda = o(1)$. As in Theorem 2, to prove the main result (20) of this theorem, it suffices to verify (B.17), i.e. to show that uniformly in i, j , for sufficiently large κ it holds:

$$P(|\widehat{\sigma}_{ij,t} - \sigma_{ij,t}| > 4\lambda) = o(p^{-2}). \quad (\text{B.36})$$

We will rewrite $\widehat{\sigma}_{ij,t} - \sigma_{ij,t}$ as follows. Observe that

$$\begin{aligned} \mathbf{y}_k &= \mathbf{H}_k \mathbf{x}_t = (y_{1k}, \dots, y_{pk})', \quad \text{where } y_{ik} = \sum_{u=1}^p h_{iu,k} x_{uk}, \\ \boldsymbol{\Sigma}_t &= \mathbf{H}_t \boldsymbol{\Sigma}_t^{(x)} \mathbf{H}_t' = (\sigma_{ij,t}), \quad \text{where } \sigma_{ij,t} = \sum_{u,v=1}^p h_{iu,t} h_{ju,t} \sigma_{uv,t}^{(x)} \end{aligned}$$

and $y_{ik}y_{jk} = \sum_{u,v=1}^p h_{iu,k}h_{jv,k}x_{uk}x_{vk}$. Since $\sigma_{uv,t}^{(x)} = E[x_{ut}x_{vt}] - E[x_{ut}]E[x_{vt}]$, then

$$\sigma_{ij,t} = \sum_{u,v=1}^p h_{iu,t}h_{ju,t}(E[x_{ut}x_{vt}] - E[x_{ut}]E[x_{vt}]).$$

So,

$$\begin{aligned} \widehat{\sigma}_{ij,t} - \sigma_{ij,t} &= K_t^{-1} \sum_{k=1}^T b_{H,|t-k|} y_{ik}y_{jk} - \bar{y}_{it}\bar{y}_{jt} - \sigma_{ij,t} \\ &= \sum_{u,v=1}^p \pi_{ij,uv,t}, \quad \text{where} \\ \pi_{ij,uv,t} &= \widetilde{s}_{ij,uv,t} - \bar{y}_{iu,t}\bar{y}_{jv,t} + (h_{iu,t}E[x_{ut}])(h_{jv,t}E[x_{vt}]), \\ \widetilde{s}_{ij,uv,t} &= K_t^{-1} \sum_{k=1}^T b_{H,|t-k|} (h_{iu,k}h_{jv,k}x_{uk}x_{vk} - h_{iu,t}h_{jv,t}E[x_{ut}x_{vt}]), \\ \bar{y}_{iu,t} &= K_t^{-1} \sum_{k=1}^T b_{H,|t-k|} h_{iu,k}x_{uk}. \end{aligned} \tag{B.37}$$

By assumption of the theorem, the sparsity parameter n_H of \mathbf{H}_t is finite and fixed, and does not depend on t, p, T . Therefore, for any fixed (i, j) the sum $\sum_{u,v=1}^p [\dots]$ in (B.37) includes no more than n_H^2 of non-zero terms. Without restriction of generality, assume that

$$\widehat{\sigma}_{ij,t} - \sigma_{ij,t} = \sum_{u,v=1}^{n_H} \pi_{ij,uv,t}.$$

Hence, to verify (B.36), it suffices to show that uniformly in i, j, u, v , for sufficiently large κ it holds:

$$P(|\pi_{ij,uv,t}| > 4\lambda') = o(p^{-2}), \quad \lambda' = \lambda/n_H^2. \tag{B.38}$$

Set $s_{iu,t} = \bar{y}_{iu,t} - h_{iu,t}E[x_{ut}]$, $v_{iu,t} = h_{iu,t}E[x_{ut}]$. Then,

$$\pi_{ij,uv,t} = \widetilde{s}_{ij,uv,t} - (s_{iu,t}s_{jv,t} + v_{iu,t}s_{jv,t} + v_{jv,t}s_{iu,t}).$$

Thus, similarly to (B.19),

$$\begin{aligned} \pi_{ij,uv,t} &\leq P(|\widetilde{s}_{ij,uv,t}| > \lambda') + P(|s_{iu,t}s_{jv,t}| > \lambda') \\ &\quad + P(|v_{iu,t}s_{jv,t}| > \lambda') + P(|v_{jv,t}s_{iu,t}| > \lambda'). \end{aligned} \tag{B.39}$$

Since $\lambda \rightarrow 0$, assume that $\lambda \leq 1$. Then, $\lambda' < 1$, and similarly to (B.20),

$$P(|s_{iu,t}s_{jv,t}| > \lambda') \leq P(|s_{iu,t}| > \lambda') + P(|s_{jv,t}| > \lambda').$$

Therefore, to prove (B.38), it suffices to show that uniformly in u, v, i, j , as $T \rightarrow \infty$, for sufficiently large κ it holds

$$\begin{aligned} P(|\widetilde{s}_{ij,uv,t}| > \lambda') &= o(p^{-2}), \quad P(|s_{jv,t}| > \lambda') = o(p^{-2}), \\ P(|v_{iu,t}s_{jv,t}| > \lambda') &= o(p^{-2}). \end{aligned} \tag{B.40}$$

Let i, j, u, v be fixed. Define $z_k := x_{uk}x_{vk}$, $\tilde{h}_k := h_{iu,k}h_{jv,k}$. By Assumption M, the process $(\mathbf{x}_k - E\mathbf{x}_k)$ is α -mixing, and therefore the process $(z_k - Ez_k)$ is also α -mixing with mixing coefficients satisfying (1). Moreover, as in the proof of Theorem 2, Ez_k satisfies (B.26). By Assumption H, $(h_{iu,k})$ satisfies (18) with parameter α and (\tilde{h}_k) with parameter $\alpha/2$.

We can write

$$\begin{aligned}\tilde{s}_{ij,uv,t} &= K_t^{-1}H^{1/2}q_{ij,uv,t}, & q_{ij,uv,t} &= H^{-1/2}\sum_{k=1}^T b_{H,|t-k|}(\tilde{h}_k z_k - \tilde{h}_t E z_t), \\ s_{iu,t} &= K_t^{-1}H^{1/2}q_{iu,t}, & q_{iu,t} &= H^{-1/2}\sum_{k=1}^T b_{H,|t-k|}(h_{iu,k}x_{uk} - h_{iu,t}E x_{ut}).\end{aligned}$$

This together with (B.23), setting $\eta = a_1 H^{1/2} \lambda'$, implies

$$\begin{aligned}P(|\tilde{s}_{ij,uv,t}| > \lambda') &\leq P(|q_{ij,uv,t}| > \eta), \\ P(|s_{iu,t}| > \lambda') &\leq P(|q_{iu,t}| \geq \eta), \\ P(|v_{iu,t}s_{jv,t}| > \lambda') &\leq P(|v_{iu,t}q_{jv,t}| \geq \eta).\end{aligned}\tag{B.41}$$

In addition, set $L = b(\log p)^{1/\alpha} > 1$, where $b > 0$ will be selected below. Then,

$$\begin{aligned}P(|v_{iu,t}q_{jv,t}| \geq \eta) &\leq P(|v_{iu,t}| \geq L) + P(L|q_{jv,t}| \geq \eta) \\ &= P(|v_{iu,t}| \geq L) + P(|q_{jv,t}| \geq L^{-1}\eta), \\ P(|q_{iu,t}| \geq \eta) &\leq P(|q_{iu,t}| \geq L^{-1}\eta).\end{aligned}\tag{B.42}$$

We will show that there exist sufficiently large $b > 0$ and $\kappa > 0$ such that

$$P(|v_{iu,t}| \geq L) = o(p^{-2}),\tag{B.43}$$

$$P(|q_{ij,uv,t}| \geq \eta) = o(p^{-2}), \quad P(|q_{jv,t}| \geq \eta/L) = o(p^{-2})\tag{B.44}$$

which together with (B.42), (B.41) implies (B.40) which completes the proof of (B.36).

Proof of (B.43). By assumption, $(\mathbf{x}_t) \in \mathcal{E}(s)$ or $(\mathbf{x}_t) \in \mathcal{H}(\theta)$ which implies $\max_{i,t} |E x_{it}| \leq m < \infty$. Therefore, $|v_{iu,t}| = |h_{iu,t}E[x_{ut}]| \leq m|h_{iu,t}|$. By Assumption H, $(h_{iu,t}) \in \mathcal{E}(\alpha)$. Therefore, $(v_{iu,t}) \in \mathcal{E}(\alpha)$ which implies that for some $c_0, c_1 \geq 0$,

$$P(|v_{iu,t}| \geq \zeta) \leq c_0 \exp(-c_1|\zeta|^\alpha), \quad \zeta > 0.$$

Using this bound with $\zeta = L = b(\log p)^{1/\alpha}$, we obtain

$$P(|v_{iu,t}| \geq L) \leq c_0 \exp(-c_1 b^\alpha \log p) = o(p^{-2})$$

when b is selected such that $c_1 b^\alpha > 2$. This proves (B.43).

Proof of (B.44).

(i) Let $(x_{ik}) \in \mathcal{E}(s)$. Recall that $q_{ij,uv,t}$ is a weighted sum of variables \tilde{h}_{k,z_k} , and by the assumptions of the theorem, $(h_{iu,k}) \in \mathcal{E}(\alpha)$. Thus, $(\tilde{h}_k) \in \mathcal{E}(\alpha/2)$ and $(z_k) \in \mathcal{E}(s/2)$. On the other hand, $q_{iu,t}$ is a weighted sum of variables $h_{iu,k}x_{uk}$, where $(h_{iu,k}) \in \mathcal{E}(\alpha)$ and $(x_{uk}) \in \mathcal{E}(s)$. Hence, by the claim (58) of Lemma 5,

$$\begin{aligned} P(|q_{ij,uv,t}| > \eta) &\leq f_H(\gamma_1, \gamma_2, c, \eta(1 \wedge d_{Ht})), \\ P(|q_{iu,t}| \geq \eta/L) &\leq f_H(\gamma'_1, \gamma'_2, c, (\eta/L)(1 \wedge d_{Ht})), \end{aligned} \quad (\text{B.45})$$

where

$$\begin{aligned} \gamma_1 &= \frac{2(\alpha/2)}{\alpha/2 + 2} = \frac{2\alpha}{\alpha + 4}, & \gamma_2 &= \frac{(\alpha/2)(s/2)}{\alpha/2 + s/2 + 1} = \frac{\alpha s}{2\alpha + 2s + 4}, \\ \gamma'_1 &= \frac{2\alpha}{\alpha + 2}, & \gamma'_2 &= \frac{\alpha s}{\alpha + s + 1}. \end{aligned} \quad (\text{B.46})$$

By assumption of the theorem, $\delta T \leq t \leq T$. We will show below that

$$\begin{aligned} \eta(1 \wedge d_{Ht}) &\geq a_\delta \kappa (\log p)^{1/\gamma_1}, & a_\delta &= \delta^{1/2} (a_1/n_H^2), \\ (\eta/L)(1 \wedge d_{Ht}) &\geq a'_\delta \kappa (\log p)^{1/\gamma'_1}, & a'_\delta &= b^{-1} \delta^{1/2} (a_1/n_H^2). \end{aligned} \quad (\text{B.47})$$

The function $f_H(\gamma_1, \gamma_2, c, \zeta)$, see (B.29), is non-increasing in ζ . So,

$$\begin{aligned} f_H(\gamma_1, \gamma_2, c, \eta(1 \wedge d_{Ht})) &\leq f_H(\gamma_1, \gamma_2, c, a_\delta \kappa (\log p)^{1/\gamma_1}), \\ f_H(\gamma'_1, \gamma'_2, c, (\eta/L)(1 \wedge d_{Ht})) &\leq f_H(\gamma'_1, \gamma'_2, c, a'_\delta \kappa (\log p)^{1/\gamma'_1}). \end{aligned}$$

Notice that,

$$\begin{aligned} &f_H(\gamma_1, \gamma_2, c, a_\delta \kappa (\log p)^{1/\gamma_1}) \\ &\leq c_0 \left\{ \exp(-c_1 (a_\delta \kappa)^{\gamma_1} \log p) + \exp\left(-c_2 (a_\delta \kappa (\log p)^{1/\gamma_1} \frac{H^{1/2}}{\log^2 H})^{\gamma_2}\right) \right\} = o(p^{-2}) \end{aligned}$$

because $c_1 (a_\delta \kappa)^{\gamma_1} > 2$ when κ is selected sufficiently large, and because by the assumption (19), $H \geq c_0 p^\varepsilon$, which implies

$$\log p = o\left(\left(\log^{1/\gamma_1} p \frac{H^{1/2}}{\log^2 H}\right)^{\gamma_2}\right).$$

The same argument implies, that for sufficiently large κ ,

$$f_H(\gamma'_1, \gamma'_2, c, a'_\delta \kappa (\log p)^{1/\gamma'_1}) = o(p^{-2}).$$

Together with (B.45) and (B.41) this proves (B.40).

Proof of (B.47). Notice that $\nu = (\alpha + 4)/(2\alpha)$ in (B.35) has property:

$$\nu\gamma_1 = 1, \quad (\nu - \alpha^{-1})\gamma'_1 = 1. \quad (\text{B.48})$$

By definition (57), $d_{Ht} = (t \vee H)^{1/2}H^{-1}$. By assumption, $\delta T \leq t \leq T$ and $H = o(T)$. Therefore,

$$\begin{aligned} d_{Ht} &\geq (\delta T \wedge H)^{1/2}H^{-1} \geq (\delta T)^{1/2}H^{-1}, \\ 1 \wedge d_{Ht} &\geq \delta^{1/2}(1 \wedge T^{1/2}H^{-1}). \end{aligned}$$

Since for any $e > 0$, $(1 \vee e)(1 \wedge e^{-1}) = 1$, we obtain

$$\begin{aligned} \eta(1 \wedge d_{Ht}) &= (a_1/n_H^2)\kappa(\log p)^\nu \{(H^{-1/2} \vee (H/T)^{1/2})H^{1/2}\}(1 \wedge d_{Ht}) \\ &\geq (a_1/n_H^2)\kappa(\log p)^\nu \delta^{1/2}(1 \vee HT^{-1/2})(1 \wedge T^{1/2}H^{-1}) \\ &= (a_1/n_H^2)\delta^{1/2}\kappa(\log p)^\nu = a_\delta\kappa(\log p)^\nu. \end{aligned}$$

Since by (B.48), $\nu = 1/\gamma_1$ this proves the first claim in (B.47).

On the other hand, $L^{-1} = b^{-1}(\log p)^{-1/\alpha}$, and therefore,

$$(\eta/L)(1 \wedge d_{Ht}) \geq b^{-1}(a_1/n_H^2)\delta^{1/2}\kappa(\log p)^{\nu-1/\alpha} = a'_\delta\kappa(\log p)^{1/\gamma'_1}$$

by (B.48) which completes the proof of (B.47).

(ii) Let $(x_{it}) \in \mathcal{H}(\theta)$. Then $q_{ij,uv,t}$ is a weighted sum of variables $\tilde{h}_k z_k$ where $(\tilde{h}_k) \in \mathcal{E}(\alpha/2)$ and $(z_k) \in \mathcal{H}(\theta/2)$. In turn, $q_{iu,t}$ is a weighted sum of variables $h_{iu,k} x_{uk}$ where $(h_{iu,k}) \in \mathcal{E}(\alpha)$ and $(x_{uk}) \in \mathcal{H}(\theta)$. Thus, by the claim (59) of Lemma 5,

$$\begin{aligned} P(|q_{ij,uv,t}| > \eta) &\leq g_H(\gamma_1, \theta', c, \eta(1 \wedge d_{Ht})), \quad \theta' \in (2, \theta/2), \\ P(|q_{iu,t}| \geq \eta/L) &\leq g_H(\gamma'_1, \theta', c, (\eta/L)(1 \wedge d_{Ht})), \end{aligned} \quad (\text{B.49})$$

where γ_1 and γ'_1 are the same as in (B.46). Since $g_H(\gamma_1, \gamma_2, c, \zeta)$, (B.33), is a non-increasing function in ζ , by (B.47) we can bound

$$\begin{aligned} g_H(\gamma_1, \theta', c, \eta(1 \wedge d_{Ht})) &\leq g_H(\gamma_1, \theta', c, a_\delta\kappa(\log p)^{1/\gamma_1}), \\ g_H(\gamma'_1, \theta', c, (\eta/L)(1 \wedge d_{Ht})) &\leq g_H(\gamma'_1, \theta', c, a'_\delta\kappa(\log p)^{1/\gamma'_1}). \end{aligned}$$

Notice that

$$\begin{aligned} &g_H(\gamma_1, \theta', c, a_\delta\kappa(\log p)^{1/\gamma_1}) \\ &\leq c_0 \left\{ \exp(-c_1(a_\delta\kappa)^{\gamma_1} \log p) + \frac{1}{(a_\delta\kappa(\log p)^{1/\gamma_1})^{\theta'}} \frac{1}{H^{\theta'/2-1}} \right\} = o(p^{-2}) \end{aligned}$$

when κ is selected such that $c_1(a_\delta\kappa)^{\gamma_1} > 2$, and $\theta' \in (2, \theta/2)$ is selected close enough to $\theta/2$, see the proof of (B.34). Similarly, it can be shown that for sufficiently large κ ,

$$g_H(\gamma'_1, \theta', c, a'_\delta\kappa(\log p)^{1/\gamma'_1}) = o(p^{-2}).$$

Together with (B.49) this implies (B.44). This completes the proof of the claim (20) of Theorem 4.

The claim (21) of the theorem is shown in (B.4) of Lemma B1.

The bandwidth $H_{opt} = T^{1/2}$ minimizes $\max(H^{-1/2}, (H/T)^{1/2})$ which implies

$$\lambda = \kappa(\log p)^\nu \max(H^{-1/2}, (H/T)^{1/2}) \geq \lambda_{opt} = \kappa(\log p)^\nu T^{-1/4}$$

which proves the last claim of the theorem. \square

C. Auxiliary results

This section contains auxiliary results used in the proofs.

Recall definition of functions f_t and g_t , (39).

Lemma C1 (i) Let $x \in \mathcal{E}(\alpha)$, $y \in \mathcal{E}(\alpha')$ where $\alpha > 0$, $\alpha' > 0$. Then $xy \in \mathcal{E}(\tilde{\alpha})$ where $\tilde{\alpha} = \alpha\alpha'/(\alpha + \alpha')$.

Moreover, $x + y \in \mathcal{E}(\min(\alpha, \alpha'))$ and $|z| \leq |x|$ implies $z \in \mathcal{E}(\alpha)$.

(ii) Let $x \in \mathcal{E}(\alpha)$, $y \in \mathcal{H}(\theta)$ where $\alpha > 0$, $\theta > 0$. Then $xy \in \mathcal{H}(\theta')$ for any $0 < \theta' < \theta$.

(iii) Let $(x_t) \in \mathcal{E}(\alpha)$, $\alpha > 0$ and $P(|y_t| \geq \zeta) \leq f_t(\gamma_1, \gamma_2, c, \zeta)$, $\zeta > 0$, $t \geq 2$ with $\gamma_1, \gamma_2 > 0$. Then

$$P(|x_t y_t| \geq \zeta) \leq f_t(\tilde{\gamma}_1, \tilde{\gamma}_2, c', \zeta), \quad \zeta > 0, \quad t \geq 2, \quad (\text{C.1})$$

where $\tilde{\gamma}_1 = \alpha\gamma_1/(\alpha + \gamma_1)$, $\tilde{\gamma}_2 = \alpha\gamma_2/(\alpha + \gamma_2)$ and c' does not depend on t, ζ .

(iv) Let $(x_t) \in \mathcal{E}(\alpha)$, $\alpha > 0$ and $P(|y_t| \geq \zeta) \leq g_t(\gamma, \theta, c, \zeta)$, $\zeta > 0$, $t \geq 2$ where $\gamma > 0$, $\theta > 2$. Then for any $\theta' \in (2, \theta)$,

$$P(|x_t y_t| \geq \zeta) \leq g_t(\tilde{\gamma}, \theta', c', \zeta), \quad \zeta > 0, \quad t \geq 2, \quad (\text{C.2})$$

where $\tilde{\gamma} = \alpha\gamma/(\alpha + \gamma)$ and c' does not depend on t, ζ .

(v) If $(x_t) \in \mathcal{E}(\alpha)$, $(x_{tk}) \in \mathcal{E}(\alpha)$ for some $\alpha > 0$ then as $T \rightarrow \infty$,

$$\max_{1 \leq t \leq T} |x_t| = O_P((\log T)^{1/\alpha}), \quad \max_{1 \leq t, k \leq T} |x_{tk}| = O_P((\log T)^{1/\alpha}). \quad (\text{C.3})$$

Proof.

(i) Let $x \in \mathcal{E}(\alpha)$, $y \in \mathcal{E}(\alpha')$ where $\alpha > 0$, $\alpha' > 0$ and let $\tilde{\alpha} = \alpha\alpha'/(\alpha + \alpha')$. Then for some $a > 0$,

$$E \exp(a|x|^\alpha) < \infty, \quad E \exp(a|y|^{\alpha'}) < \infty.$$

To prove (i), we will show that $E \exp(a|xy|^{\tilde{\alpha}}) < \infty$.

Set $p = (\alpha + \alpha')/\alpha'$, $q = (\alpha + \alpha')/\alpha$. Then $p > 1$, $q > 1$, $1/p + 1/q = 1$ and $\tilde{\alpha}p = \alpha$, $\tilde{\alpha}q = \alpha'$. Hence, for $k = 1, 2, \dots$ by Hölder's inequality,

$$\begin{aligned} E|xy|^{\tilde{\alpha}k} &= E[|x|^{\tilde{\alpha}k}|y|^{\tilde{\alpha}k}] \leq (E|x|^{k\tilde{\alpha}p})^{1/p} (E|y|^{k\tilde{\alpha}q})^{1/q} = (E|x|^{k\alpha})^{1/p} (E|y|^{k\alpha'})^{1/q} \\ &\leq (\max(E|x|^{k\alpha}, E|y|^{k\alpha'})^{1/p+1/q} = \max(E|x|^{k\alpha}, E|y|^{k\alpha'}) \\ &\leq E|x|^{k\alpha} + E|y|^{k\alpha'}. \end{aligned}$$

Therefore,

$$\begin{aligned} E \exp(a|xy|^{\tilde{\alpha}}) &\leq \sum_{k=0}^{\infty} \frac{a^k E|xy|^{\tilde{\alpha}k}}{k!} \leq \sum_{k=0}^{\infty} \frac{a^k (E|x|^{k\alpha} + E|y|^{k\alpha'})}{k!} \\ &\leq E \exp(a|x|^\alpha) + E \exp(a|y|^{\alpha'}) < \infty. \end{aligned}$$

(ii) Let $x \in \mathcal{E}(\alpha)$, $y \in \mathcal{H}(\theta)$ where $\alpha > 0$, $\theta > 0$. Then, for some $a > 0$,

$$E \exp(a|x|^\alpha) < \infty, \quad E|y|^\theta < \infty.$$

The latter implies that $E|x|^b < \infty$ for any $b > 0$.

Let $\theta' \in (0, \theta)$. To prove (ii), we will show that $E|xy|^{\theta'} < \infty$. Set $p = \theta/\theta'$ and let $q > 1$ be defined by equality $1/p + 1/q = 1$. Then, by Hölder inequality,

$$E|xy|^{\theta'} \leq (E|x|^{\theta'q})^{1/q} (E|y|^{\theta'p})^{1/p} = (E|x|^{\theta'})^{1/q} (E|y|^\theta)^{1/p} < \infty.$$

This completes the proof of (ii).

Before proceeding to the proof of (iii)-(iv), we obtain the following two auxiliary results. First, consider the function

$$f(x) = x^\alpha + c(v/x)^{\alpha'}, \quad x > 0$$

where $\alpha > 0$, $\alpha' > 0$, $v > 0$, $c > 0$. It achieves its unique minimum at

$$x_0 = (c\alpha'/\alpha)^{1/(\alpha+\alpha')} v^{\alpha'/(\alpha+\alpha')}$$

because x_0 is a unique solution of equation $f'(x) = \alpha x^{\alpha-1} - c\alpha'(v/x)^{\alpha'} x^{-1} = 0$ and $f''(x_0) = x_0^{\alpha-2} \alpha(\alpha + \alpha') > 0$. Thus,

$$f(x) \geq f(x_0) = c'v^{\tilde{\alpha}}, \quad x \geq 0 \tag{C.4}$$

where $\tilde{\alpha} = \alpha\alpha'/(\alpha + \alpha')$ and $c' = (c\alpha'/\alpha)^{\alpha/(\alpha+\alpha')}(1 + \alpha/\alpha')$.

Second, we obtain the upper bound for $P(|xy| \geq \zeta)$ for the product of r.v.'s x and y when $x \in \mathcal{E}(\alpha)$, $\alpha > 0$. Let $p, q > 1$, $1/p + 1/q = 1$. Then

$$\begin{aligned} P(|xy| \geq \zeta) &= \sum_{k=0}^{\infty} P(\{|x| \in [k, k+1)\} \cap \{|xy| \geq \zeta\}) \\ &\leq \sum_{k=0}^{\infty} P^{1/p}(|x| \in [k, k+1)) P^{1/q}(|y| \geq \zeta/(k+1)). \end{aligned}$$

Since $x \in \mathcal{E}(\alpha)$, then for $k \geq 0$,

$$P(|x| \in [k, k+1)) \leq P(|x| \geq k) \leq c'_0 \exp(-2c'_1 k^\alpha), \quad k \geq 0$$

for some $c'_0 > 0$, $c'_1 > 0$. Denote

$$g_{k\zeta} := \exp(-c'_1 k^\alpha) P^{1/q}(|y| \geq \zeta/k).$$

Then,

$$\begin{aligned} P(|xy| \geq \zeta) &\leq C \sum_{k=0}^{\infty} \exp\{-2c'_1 k^\alpha + c'_1 (k+1)^\alpha\} g_{k+1, \zeta} \\ &\leq C \max_{k \geq 1} g_{k\zeta} \sum_{k=0}^{\infty} \{-2c'_1 k^\alpha + c'_1 (k+1)^\alpha\} \\ &\leq C \max_{k \geq 1} g_{k\zeta}. \end{aligned} \tag{C.5}$$

We use this result to evaluate $P(|xy| \geq \zeta)$ in parts (iii)-(iv) of the lemma.

(iii) Without restriction of generality, we assume that $\zeta \geq 1$. By (C.5),

$$P(|x_t y_t| \geq \zeta) \leq C \max_{k \geq 1} g_{k\zeta}. \tag{C.6}$$

Under assumptions of (iii), $g_{k\zeta} = \exp(-c'_1 k^\alpha) f_t^{1/q}(2, \gamma, c, \zeta/k)$. To evaluate $f_t^{1/q}(2, \gamma, c, \zeta/k)$, denote $\zeta_t = \zeta\sqrt{t}/\log^2 t$. Using the definition of function f_t , (39), and inequality

$$(a+b)^{1/q} \leq a^{1/q} + b^{1/q}, \quad a, b > 0, \tag{C.7}$$

we obtain

$$\begin{aligned} f_t^{1/q}(\gamma_1, \gamma_2, c, \zeta/k) &\leq C \left(\exp(-c_1(\zeta/k)^{\gamma_1}) + \exp(-c_2(\zeta_t/k)^{\gamma_2}) \right)^{1/q} \\ &\leq C \left(\exp(-(c_1/q)(\zeta/k)^{\gamma_1}) + \exp(-(c_2/q)(\zeta_t/k)^{\gamma_2}) \right). \end{aligned}$$

Hence, there exist constants $c_1'', c_2'' > 0$ such that

$$g_{k,\zeta} \leq C \{ \exp(-c_1''(k^\alpha + (\zeta/k)^{\gamma_1})) + \exp(-c_2''(k^\alpha + (\zeta_t/k)^{\gamma_2})) \}.$$

Next, using (C.4) to bound $f(k) := k^\alpha + (\zeta/k)^{\gamma_1}$, $f(k) := k^\alpha + (\zeta_t/k)^{\gamma_2}$ from below, we obtain

$$g_{k,\zeta} \leq c_0^* \left(\exp(-c_1^* \zeta^{\tilde{\gamma}_1}) + \exp(-c_2^* \zeta_t^{\tilde{\gamma}_2}) \right) = f_t(\tilde{\gamma}_1, \tilde{\gamma}_2, c^*, \zeta), \quad k \geq 1,$$

with $\tilde{\gamma}_1 = \alpha\gamma_1/(\alpha + \gamma_1)$, $\tilde{\gamma}_2 = \alpha\gamma_2/(\alpha + \gamma_2)$. Thus, (C.6) implies

$$P(|x_t y_t| \geq \zeta) \leq f_t(\tilde{\gamma}_1, \tilde{\gamma}_2, c'', \zeta)$$

which proves (iii).

(iv) Let $\zeta \geq 1$. Under assumptions of (iv), (C.6) holds with

$$g_{k\zeta} = \exp(-c_1' k^\alpha) g_t^{1/q}(\gamma, \theta, c, \zeta/k).$$

Next we evaluate $g_t^{1/q}(\gamma, \theta, c, \zeta/k)$. Let $2 < \theta' < \theta$. Then, $\theta/\theta' > 1$ and $(\theta - 2)/(\theta' - 2) > 1$. Let $q > 1$ be such that $\min(\theta/\theta', (\theta - 2)/(\theta' - 2)) > q$. By (C.7) and definition of g_t , (39),

$$\begin{aligned} g_t^{1/q}(\gamma, \theta', c, \zeta/k) &\leq C \left(\exp\{-c_1(\zeta/k)^\gamma\} + (\zeta/k)^{-\theta} t^{-(\theta/2-1)} \right)^{1/q} \\ &\leq C \{ \exp\{-(c_1/q)(\zeta/k)^\gamma\} + \zeta^{-\theta/q} t^{-(\theta/2-1)/q} k^{\theta/q} \}. \end{aligned} \quad (\text{C.8})$$

Definition of $q > 1$ implies $\theta/q > \theta'$ and $(\theta/2 - 1)/q > \theta'/2 - 1$. This together with (C.8) yields

$$g_t^{1/q}(\gamma, \theta', c, \zeta/k) \leq C \left(\exp\{-(c_1/q)(\zeta/k)^\gamma\} + \zeta^{-\theta'} t^{-(\theta'/2-1)} k^{\theta'/q} \right), \quad \zeta \geq 1, t \geq 1.$$

Hence,

$$\begin{aligned} \max_{k \geq 1} g_{k\zeta} &\leq C \max_{k \geq 1} \exp\{-c_1''(k^\alpha + (\zeta/k)^\gamma)\} + C \zeta^{-\theta'} t^{-(\theta'/2-1)} \max_{k \geq 1} \{ \exp(-c_2' k^\alpha) k^{\theta'/q} \} \\ &\leq C \left(\max_{k \geq 1} \exp\{-c_1''(k^\alpha + (\zeta/k)^\gamma)\} + \zeta^{-\theta'} t^{-(\theta'/2-1)} \right). \end{aligned}$$

Applying to $f(k) := k^\alpha + (\zeta/k)^\gamma$ the bound (C.4), we obtain

$$\max_{k \geq 1} g_{k\zeta} \leq c_0^* \left(\exp(-c_1^* \zeta^{\tilde{\gamma}}) + \zeta^{-\theta'} t^{-(\theta'/2-1)} \right) = g_t(\tilde{\gamma}, \theta', c^*, \zeta)$$

with $\tilde{\gamma} = \alpha\gamma/(\alpha + \gamma)$. Then (C.6) implies $P(|x_t y_t| \geq \zeta) \leq g_t(\tilde{\gamma}, \theta', c^*, \zeta)$ which proves (iv).

(v) We need to show that, as $T \rightarrow \infty$, $b \rightarrow \infty$,

$$P\left(\max_{t=1, \dots, T} |x_t| \geq b(\log T)^{1/\alpha} \right) \rightarrow 0, \quad P\left(\max_{t,k=1, \dots, T} |x_{tk}| \geq b(\log T)^{1/\alpha} \right) \rightarrow 0.$$

By assumption, there exist $a > 0$ and $\alpha > 0$ such that

$$\max_{t \geq 1} E \exp(a|x_t|^\alpha) < \infty, \quad \max_{t,k \geq 1} E \exp(a|x_{tk}|^\alpha) < \infty.$$

Let b be such that $ab^\alpha \geq 2$. Then, as $T \rightarrow \infty$,

$$\begin{aligned} P\left(\max_{t=1,\dots,T} |x_t| \geq b(\log T)^{1/\alpha}\right) &\leq \sum_{t=1}^T P(|x_t| \geq b(\log T)^{1/\alpha}) \\ &\leq \sum_{t=1}^T \frac{E(\exp(a|x_t|^\alpha))}{\exp(ab^\alpha \log T)} \leq T^{-2} \sum_{t=1}^T C \rightarrow 0. \end{aligned}$$

Similarly,

$$\begin{aligned} P\left(\max_{t,k=1,\dots,T} |x_{tk}| \geq b(\log T)^{1/\alpha}\right) &\leq \sum_{t,k=1}^T P(|x_{tk}| \geq b(\log T)^{1/\alpha}) \\ &\leq \sum_{t,k=1}^T \frac{E(\exp(a|x_{tk}|^\alpha))}{\exp(ab^\alpha \log T)} \leq T^{-2} \sum_{t,k=1}^T C \rightarrow 0. \end{aligned}$$

This completes the proof of (v) and the lemma. \square

Lemma C2 *Let $\gamma > 0$.*

(i) *Let ξ be a zero mean random variable. Then for all $\zeta > 0$,*

$$E(|\xi|^\gamma I(|\xi| > \zeta)) \leq \begin{cases} c_0 \exp(-c_1 \zeta^s) & \text{if } \xi \in \mathcal{E}(s), s > 0 \\ c_0 \zeta^{\gamma-\theta} & \text{if } \xi \in \mathcal{H}(\theta), \gamma < \theta \end{cases} \quad (\text{C.9})$$

$$(\text{C.10})$$

for some $c_0 > 0$, $c_1 > 0$ which do not depend on ζ .

(ii) *Let s_t , $t \geq 1$ be zero mean random variables such that for some $\gamma_1 > 0$, $\gamma_2 > 0$ and c ,*

$$P(|s_t| \geq \zeta) \leq f_t(\gamma_1, \gamma_2, c, \zeta) \quad \text{for all } \zeta > 0, t \geq 2.$$

Then,

$$E[|s_t|^\gamma I(|s_t| > \zeta)] \leq f_t(\gamma_1, \gamma_2, c', \zeta), \quad \zeta > 0, t \geq 2, \quad (\text{C.11})$$

where c' does not depend on ζ , t .

(iii) *Let s_t , $t \geq 1$ be zero mean random variables such that for some $\theta > 0$, $\gamma_1 > 0$ and c ,*

$$P(|s_t| \geq \zeta) \leq g_t(\gamma_1, \theta, c, \zeta) \quad \text{for all } \zeta > 0, t \geq 2.$$

Then, for $0 < \gamma < \theta$,

$$E[|s_t|^\gamma I(|s_t| > \zeta)] \leq \max(\zeta^\gamma, 1) g_t(\gamma_1, \theta, c', \zeta), \quad \zeta > 0, t \geq 2, \quad (\text{C.12})$$

where c' does not depend on ζ , t .

Proof. Without restriction of generality let $\zeta \geq 1$. Denote $F(x) = P(|\xi| \geq x)$. Then

$$E[|\xi|^\gamma I(|\xi| > \zeta)] = - \int_\zeta^\infty x^\gamma dF(x) = -\zeta^\gamma F(\zeta) + \int_\zeta^\infty x^{\gamma-1} F(x) dx. \quad (\text{C.13})$$

(i) If $(\xi_k) \in \mathcal{E}(s)$, then $F(x) \leq c'_0 \exp(-2c'_1|x|^s)$ for some $c'_0, c'_1 > 0$. Notice that $F(x) \geq F(\zeta)$, $x \geq \zeta$. Applying these bounds in (C.13), we obtain (C.9):

$$\begin{aligned} E[|\xi|^\gamma I(|\xi| > \zeta)] &\leq F^{1/2}(\zeta) \{ \zeta^\gamma F^{1/2}(\zeta) + \int_1^\infty x^{\gamma-1} F^{1/2}(x) dx \} \\ &\leq C F^{1/2}(\zeta) \leq C \exp(-c'_1 \zeta^s). \end{aligned}$$

If $(\xi) \in \mathcal{H}(\theta)$, then $F(x) \leq c'_0|x|^{-\theta}$. Using this bound in (C.13), we obtain (C.10):

$$E[|\xi|^\gamma I(|\xi| > \zeta)] \leq C \{ \zeta^\gamma |\zeta|^{-\theta} + \int_\zeta^\infty x^{\gamma-1} x^{-\theta} dx \} \leq C \zeta^{\gamma-\theta}.$$

(ii) Let again $\zeta \geq 1$. Denote $F_t^*(x) = P(|S_t| \geq x)$. Then as in (C.13),

$$\begin{aligned} E[|s_t|^\gamma I(|s_t| > \zeta)] &= - \int_\zeta^\infty x^\gamma dF_t^*(x) = -\zeta^\gamma F_t^*(\zeta) + \int_\zeta^\infty x^{\gamma-1} F_t^*(x) dx, \quad (\text{C.14}) \\ E[|s_t|^\gamma I(|s_t| > \zeta)] &\leq F_t^{*1/2}(\zeta) \{ \zeta^\gamma F_t^{*1/2}(\zeta) + \int_1^\infty x^{\gamma-1} F_t^{*1/2}(x) dx \}. \end{aligned}$$

By assumption, $P(|s_t| \geq \zeta) \leq f_t(\gamma_1, \theta, c, \zeta)$. Definition (39) of f_t implies that

$$f_t(\gamma_1, \gamma_2, c, \zeta) \leq c_0 \exp(-2c_1 \zeta^{\min(\gamma_1, \gamma_2)}), \quad \zeta > 0, t \geq 2$$

for some $c_0, c_1 > 0$. Thus, by (C.14), for $\zeta \geq 1$

$$\begin{aligned} E[|s_t|^\gamma I(|s_t| > \zeta)] &\leq f_t^{1/2}(\gamma_1, \gamma_2, c, \zeta) (\zeta^\gamma f_t^{1/2}(\gamma_1, \gamma_2, c, \zeta) + \int_1^\infty x^{\gamma-1} f_t^{1/2}(\gamma_1, \gamma_2, c, x) dx) \\ &\leq C f_t^{1/2}(\gamma_1, \gamma_2, c, \zeta) \leq C f_t(\gamma_1, \gamma_2, c', \zeta) \end{aligned}$$

for some c' in view of (C.7). This proves (C.11).

(iii) Let $\zeta \geq 1$. Since $P(|s_t| \geq \zeta) \leq g_t(\gamma_1, \theta, c, \zeta)$, (C.14) implies

$$E[|s_t|^\gamma I(|s_t| > \zeta)] \leq \zeta^\gamma g_t(\gamma_1, \theta, c, \zeta) + \int_\zeta^\infty x^{\gamma-1} g_t(\gamma_1, \theta, c, x) dx. \quad (\text{C.15})$$

By definition (39), $g_t(\gamma_1, \theta, c, \zeta) \leq c_0 \{ \exp(-2c_1 \zeta^{\gamma_1}) + \zeta^{-\theta} t^{-(\theta/2-1)} \}$ for some $c_0, c_1 > 0$. Thus,

$$\begin{aligned} &\int_\zeta^\infty x^{\gamma-1} g_t(\gamma_1, \theta, c, x) dx \\ &\leq C \left(\exp(-c_1 \zeta^{\gamma_1}) \int_\zeta^\infty x^{\gamma-1} \exp(-c_1 x^{\gamma_1}) dx + \int_\zeta^\infty x^{\gamma-\theta-1} t^{-(\theta/2-1)} dx \right) \\ &\leq C \zeta^\gamma \left(\exp(-c_1 \zeta^{\gamma_1}) + \zeta^{-\theta} t^{-(\theta/2-1)} \right) = \zeta^\gamma g_t(\gamma_1, \theta, c', \zeta) \end{aligned}$$

for some c' . This together with (C.15) proves (C.12). \square

Lemma C3 Let $b_{H,k}$ satisfy (43) with $\nu > 3$ and $0 \leq \gamma < 2$. Then for $1 \leq t, H \leq T$, $T \geq 1$,

$$H^{-1} \sum_{k=1}^T b_{H,|t-k|} \left(\frac{|t-k|}{t \vee k} \right)^\gamma \leq C \left(\frac{H}{H \vee t} \right)^\gamma, \quad (\text{C.16})$$

$$H^{-1} \sum_{k=1}^T b_{H,|t-k|} \left(\frac{|t-k| \vee H}{k} \right)^{1/2} \leq C \left(\frac{H}{H \vee t} \right)^{1/2}, \quad (\text{C.17})$$

where $C > 0$ does not depend on t, T, H .

Proof. Notice that

$$\left(\frac{H}{H \vee t} \right)^\gamma = \left(\left(\frac{H}{t} \right)^\gamma \wedge 1 \right) = \left(\frac{H \wedge t}{t} \right)^\gamma. \quad (\text{C.18})$$

By (43), $b_{H,k} \leq C(1 + (k/H)^\nu)^{-1}$ for $k \geq 0$ where $\nu \geq 3$. Therefore, for $0 \leq \gamma < 2$,

$$H^{-1} \sum_{k=1}^T b_{H,k} (k/H)^\gamma \leq C, \quad \max_{k \geq 1} b_{H,k} (k/H)^\gamma \leq C, \quad (\text{C.19})$$

where C does not depend on H, T .

Denote by $I_{\gamma,H}$ the l.h.s. of (C.16). Then, by (C.19), noting that $t \vee k \geq t$,

$$I_{\gamma,H} = H^{-1} \sum_{k=1}^T b_{H,|t-k|} \left(\frac{|t-k|}{t \vee k} \right)^\gamma = H^{-1} \sum_{k=1}^T b_{H,|t-k|} \left(\frac{|t-k|}{H} \right)^\gamma \left(\frac{H}{t \vee k} \right)^\gamma \leq C \left(\frac{H}{t} \right)^\gamma.$$

On the other hand, since $|t-k|/(t \vee k) \leq 1$, using (C.19) we obtain

$$I_{\gamma,H} \leq H^{-1} \sum_{k=1}^T b_{H,|t-k|} \leq C$$

which together with (C.18) proves (C.16).

To prove (C.17), denote by I_H the l.h.s. of (C.17). Write

$$I_H = H^{-1} \sum_{k=t/2+1}^T [\dots] + H^{-1} \sum_{k=1}^{t/2} [\dots] =: I_{H;1} + I_{H;2}.$$

Then,

$$\begin{aligned} I_{H;1} &\leq \left[H^{-1} \sum_{k=t/2+1}^T b_{T,|t-k|} \left(\frac{|t-k| \vee H}{H} \right)^{1/2} \left(\frac{H}{k} \right)^{1/2} \right] \\ &\leq C \left(\frac{H}{t} \right)^{1/2} \left[H^{-1} \sum_{k=1}^T b_{T,|t-k|} \left(\frac{|t-k| \vee H}{H} \right)^{1/2} \right] \leq C \left(\frac{H}{t} \right)^{1/2} \end{aligned}$$

by (C.19). On the other hand, for $1 \leq k \leq t/2$, it holds $|t-k| \geq t/2$. Then,

$$1 = (|t-k|/H)(H/|t-k|) \leq 2(|t-k|/H)(H/t),$$

and

$$I_{H;2} \leq 2H^{-1} \sum_{k=1}^{t/2} \{b_{T,|t-k|} (\frac{|t-k| \vee H}{H})^{1/2} \frac{|t-k|}{H}\} (\frac{H}{t}) (\frac{H}{k})^{1/2}.$$

Then, using the second claim of (C.19), we obtain

$$I_{H;2} \leq C (\frac{H^{1/2}}{t}) \sum_{k=1}^{t/2} k^{-1/2} \leq C (\frac{H}{t})^{1/2}.$$

The bounds for $I_{H;1}$ and $I_{H;2}$ imply $I_H \leq C(H/t)^{1/2}$.

In view of (C.18), to prove (C.17), it remains to show that $I_H \leq C$. By (C.19),

$$\begin{aligned} I_H &\leq H^{-1} \sum_{k=1}^T b_{T,|t-k|} (\frac{|t-k| \vee H}{H})^{1/2} (\frac{H}{k})^{1/2} \\ &\leq H^{-1} \sum_{k=1}^{2H} b_{T,|t-k|} (\frac{|t-k| \vee H}{H})^{1/2} (\frac{H}{k})^{1/2} + 2H^{-1} \sum_{k=2H}^T b_{T,|t-k|} (\frac{|t-k| \vee H}{H})^{1/2} \\ &\leq CH^{-1/2} \sum_{k=1}^{2H} k^{-1/2} + CH^{-1} \sum_{k=1}^T b_{T,|t-k|} (\frac{|t-k| \vee H}{H})^{1/2} \leq C, \end{aligned}$$

where $C < \infty$ does not depend on t , H and T . This proves $I_H \leq C$ and (C.17), and completes the proof of the lemma. \square

References

- Bickel, P. and E. Levina (2008) Covariance regularization by thresholding. *The Annals of Statistics* 36, 2577–2604.
- Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press.
- Davydov, Y. (1968) Convergence of distributions generated by stationary stochastic processes. *Theory of Probability and Its Applications* 13, 691–696.
- Merlevede, F., M. Peligrad and E. Rio (2009) Bernstein inequality and moderate deviations under strong mixing conditions. *IMS Collections High Dimensional Probability V: The Luminy Volume* 5, 273–292.
- White, H. and J. M. Wooldridge (1991) Some results on sieve estimation with dependent observations. In: *Semiparametric and Nonparametric Methods in Econometrics and Statistics*, ed. by W. Barnett, J. Powell, and G. Tauchen, pp. 459–493. Cambridge University Press.