Estimation of time varying covariance matrices for large datasets

Y. Dendramis L. Giraitis G. Kapetanios

February 18, 2017

Abstract

Structural change is a fundamental issue in the statistical and econometric analysis of macroeconomic and financial data. Recently there has been considerable focus on developing econometric estimation and inference methods that allow for structural change in model parameters, using Bayesian or non-parametric kernel methods. In the context of the estimation of the covariance matrix of large dimensional panels, the development of such methods requires taking into account time variation, possible temporal dependence and the presence of heavy tailed distributions. In this paper we introduce a non-parametric version of regularisation techniques for estimating sparse large covariance matrices developed by Bickel and Levina (2008) and others. The paper includes a set of results on Bernstein type inequalities for dependent, possibly heavy-tailed, unbounded variables which are expected to be applicable widely in econometric and statistical analysis beyond the aim of estimating of large covariance matrices. We discuss the utility of robust thresholding methods and compare it with other estimators in simulations and and an empirical applications on portfolio management.

1 Introduction

The problem of structural change in econometric literature has been mainly addressed within either univariate or relatively small multivariate models and settings. A parallel, recent, development in statistics and econometrics relates to the analysis of increasingly large datasets where the number of variables is significantly larger, relative to the number of observations, than in standard econometric settings. There has been little work on the presence and analysis of structural change in large datasets. Existing research is primarily focused on factor models where large 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) and Stock and Watson (2002b).

This paper addresses this problem without recourse to factor modelling. In particular, a major consideration for large datasets relates to the estimation of their covariances. This obviously relevant problem is also particularly demanding since the dimension of the estimated object rises as a square of the dimension of the dataset under analysis. A number of different approaches 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).

In this paper we consider the problem of estimating the covariance matrix of a large dataset in the presence of structural change. Unlike the majority of the work on structural change we focus on smooth stochastic or deterministic change of the covariance matrix rather than structural breaks. We introduce new characterizations of smoothness. We follow the literature on local stationarity and, in particular, its recent extension to stochastic time-varying coefficient models by Giraitis, Kapetanios, and Yates (2014). There is a small recent literature that deals with this problem in the case of deterministic structural change, including Chen, Xu, and Wu (2013), Zhou, Lafferty, and Wasserman (2010) and Kolar and Xing (2011). However, all this work assumes that the volatility process is deterministic which is considered extremely restrictive for economic and financial data, both from a theoretical and an empirical point of view.

We develop a non-parametric version of existing regularisation techniques for sparse large covariance matrices introduced by Bickel and Levina (2008) and others. Our main objective is the investigation of the robustness of such procedures to time variation, dependence and heavy-tailedness of distributions. We provide a set of important new results on Bernstein type inequalities for dependent unbounded variables and compare the robust thresholding method to other estimators in simulations. The paper includes an empirical application on portfolio construction.

2 Thresholding Estimation

Given the p- variate sample $(\boldsymbol{y}_1, \cdots, \boldsymbol{y}_T)$, estimation of the population variance $p \times p$ matrix $\boldsymbol{\Sigma}$ by the sample variance is a well defined procedure when p is fixed. For large covariance matrices, where p increases with T the poor performance of the sample covariance matrix estimate can be improved by various regularization procedures which include the thresholding methods 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 time variation, dependence, heteroscedasticity and light/heavy tails of the data on consistency rates in the thresholding regularization estimation procedure of the covariance matrix. While in standard cases, heteroscedasticity is understood as the presence of a time varying deterministic covariance matrix $\boldsymbol{\Sigma} = \mathrm{var}(\boldsymbol{y}_t) = E[(\boldsymbol{y}_t - E\boldsymbol{y}_t)(\boldsymbol{y}_t' - E\boldsymbol{y}_t')]$, for data with random persistent (volatility) scaling components, the limit quantity estimated by $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}}$ is less obvious. As shown below, it combines a time varying deterministic matrix rescaled by time varying random volatility. In general, time variation requires implementation of local estimation of the covariance matrix, which, despite the possibility of different data generating processes, can be based on the same standard thresholding regularisation procedure. Our theoretical findings show that regularisation with a standard thresholding parameter $\lambda = \sqrt{\log p}/\sqrt{T}$ is robust to dependence, various types of distributions and heteroscedasticity.

To reflect the impact of dependence and heavy distributional tails on regularised estimation of a large dimensional covariance matrix we will consider two settings. In the first one,

$$\boldsymbol{y}_t = \boldsymbol{x}_t, \tag{1}$$

where $\boldsymbol{x}_t$ is an α-mixing p-variate process, and the covariance matrix is either constant over time, $\boldsymbol{\Sigma} \equiv \boldsymbol{\Sigma}^{(x)} = \mathrm{var}(\boldsymbol{x}_t)$, or time varying, $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t^{(x)} = \mathrm{var}(\boldsymbol{x}_t) = [\sigma_{ij,t}^{(x)}]$.

We shall write $(\boldsymbol{x}_t) \in \mathcal{M}$ to denote that $\boldsymbol{x}_t$ is an α-mixing (but not necessary stationary) sequence with mixing coefficients α_k such that $\alpha_k \leq c\phi^k$, $k \geq 1$, for some $0 < \phi < 1$ and $c > 0$.

The distribution of variables $\boldsymbol{x}_t$ may have light tails or heavy tails. We shall write $(\boldsymbol{x}_t) \in \mathcal{E}(s)$ to denote that the distribution of $x_{i,t}$ has light tails,

$$P\left[|x_{i,t}| \geq \zeta\right] \leq c_0 \exp(-c_1 \zeta^s) \text{ for all } i,\, t \text{ and } \zeta > 0 \tag{2}$$

for some $c_0, c_1 > 0$ and $s > 0$. In the case of heavy tails, we write $(\boldsymbol{x}_t) \in \mathcal{H}(\theta)$ to denote that for some $c_0 > 0$ and $\theta \geq 2$,

$$P\left[|x_{i,t}| \geq \zeta\right] \leq c_0 \zeta^{-\theta} \text{ for all } i,\, t \text{ and } \zeta > 0. \tag{3}$$

Thresholding estimation under stationarity. First we consider the case when then $\boldsymbol{x}_t$ is a stationary process. Then the mean $E\boldsymbol{x}_t = \boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma} = \mathrm{var}(\boldsymbol{x}_t)[\sigma_{ij}^{(x)}]$ does not change with t. Denote by $\hat{\boldsymbol{\Sigma}} = [\hat{\sigma}_{ij}]$ the sample variance estimate of $\boldsymbol{\Sigma}$,

$$\hat{\boldsymbol{\Sigma}} = T^{-1} \sum_{j=1}^{T} \boldsymbol{y}_j \boldsymbol{y}_j' - \bar{\boldsymbol{y}}\bar{\boldsymbol{y}}', \qquad \bar{\boldsymbol{y}}_t = T^{-1} \sum_{j=1}^{T} \boldsymbol{y}_j. \tag{4}$$

Hard and adaptive thresholding introduced by Bickel and Levina (2008) and Cai and Liu (2011) are two most commonly used approaches to regularize estimates of large dimensional covariance matrices, $\hat{\boldsymbol{\Sigma}}$, when dimension p increases with T. Hard thresholding is based on idea of setting the elements of the sample covariance matrix, which absolute values are smaller than some threshold λ, to zero. Regularising the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_t$, (4), of covariance matrix $\boldsymbol{\Sigma}$ by hard thresholding yields the estimate

$$T_\lambda\big(\hat{\Sigma}\big) = \big(\widehat{\sigma}_{ij} I(|\widehat{\sigma}_{ij}| > \lambda)\big). \tag{5}$$

Other thresholding operators can be defined, but the corresponding regularised estimates have similar properties to the hard thresholding, asymptotically, although they may have differ in finite samples.

Bickel and Levina (2008) showed that for i.i.d. Gaussian process $\boldsymbol{x}_t$ the regularized estimator $T_\lambda(\hat{\Sigma})$ of a covariance matrix Σ with sparsity parameter n_p satisfies the bound

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

$$\|T_\lambda(\hat{\Sigma})^{-1} - \Sigma^{-1}\| = O_p(n_p\lambda), \quad \text{if } n_p\lambda = o(1) \text{ and } ||\boldsymbol{\Sigma}|| \geq c > 0 \tag{7}$$

where $\|.\|$ denotes the spectral norm and κ can be fixed or chosen through cross validation.

As in Bickel and Levina (2008) and Fan, Liao, Mincheva (2013) we assume that $\boldsymbol{\Sigma} = \text{var}(\boldsymbol{x}_t)$ is approximately sparse, i.e., $n_p = n_p^{(x)}$, the maximum number of of non-zero elements in each row of $\boldsymbol{\Sigma}$ does not not grow too fast with p.

Theorem 1 *Let $\boldsymbol{y}_t = \boldsymbol{x}_t$ be as in (1), where $(\boldsymbol{x}_t) \in \mathcal{M}$ is a stationary α-mixing process.*

Then for sufficiently large $\kappa > 0$ and p, T such that

$$T \geq cp^\varepsilon \qquad (\exists c > 0;\ \varepsilon > 0) \tag{8}$$

the regularised estimate $T_\lambda(\hat{\Sigma})$ of $\boldsymbol{\Sigma} = \text{var}(\boldsymbol{x}_t)$ satisfies (6) and (7) in the following cases.

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

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

Thresholding estimation under heterescodesticity. To account for heteroscedasticity and time variation which arises in various economic applications due to structural change of the mean and variance (stochastic volatility) of observables $\boldsymbol{y}_t$, we shall use the local kernel type sample variance estimate $\hat{\boldsymbol{\Sigma}}_t = [\hat{\sigma}_{ij,t}]$, $1 \leq t \leq T$, given by

$$\hat{\boldsymbol{\Sigma}}_t = K_t^{-1} \sum_{k=1}^T b_{H,|t-k|} \boldsymbol{y}_k \boldsymbol{y}_k\prime - \bar{\boldsymbol{y}}_t \bar{\boldsymbol{y}}_t', \qquad \bar{\boldsymbol{y}}_t = K_t^{-1} \sum_{k=1}^T b_{H,|t-k|} \boldsymbol{y}_k \tag{9}$$

where $K_t = \sum_{k=1}^T b_{H,|t-k|}$ and $b_{H,|t-k|}$ are kernel weights with bandwidth parameter H specified below. Correspondingly, the regularized sample variance estimate is defined as

$$T_\lambda(\hat{\Sigma}_t) = \left(\widehat{\sigma}_{ij,t} I(|\widehat{\sigma}_{ij,t}| > \lambda)\right). \tag{10}$$

We shall focus on two types of heteroscedasticity.

(a) Under the first scenario, $\boldsymbol{y}_t$ is as in (1), and it is assumed that the covariance matrix $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t^{(x)} = \text{var}(\boldsymbol{x}_t) = [\sigma_{ij,t}^{(x)}]$ and the mean $E[\boldsymbol{x}_t] = \boldsymbol{\mu}_t$ are deterministic smooth bounded functions.

(b) Under the second scenario, we are interested in the asymptotic properties of the regularised sample covariance $T_\lambda(\hat{\mathbf{\Sigma}}_t)$ under more general forms of heteroscedasticity common in macroeconomic and financial modelling where data can be put in the form

$$\boldsymbol{y}_t = \boldsymbol{H}_t \boldsymbol{x}_t, \tag{11}$$

where $\boldsymbol{x}_t$ is an in (a) and $\boldsymbol{H}_t = (h_{ij,t})$ is a random persistent $p \times p$ (volatility) scaling matrix. Here, the object of the estimation (or the limit of $\hat{\mathbf{\Sigma}}_t$) is the matrix

$$\mathbf{\Sigma}_t = \boldsymbol{H}_t \mathrm{var}(\boldsymbol{x}_t) \boldsymbol{H}_t' = \boldsymbol{H}_t \mathbf{\Sigma}_t^{(x)} \boldsymbol{H}_t'.$$

As above, we assume that $\mathbf{\Sigma}_t^{(x)}$ is approximately sparse, i.e., $n_p = n_p^{(x)}$, the maximum number of non-zero elements in each row of $\mathbf{\Sigma}_t^{(x)}$ for all t does not not grow too fast with p. The sparsity parameter n_H of $\boldsymbol{H}_t$ is assumed to be bounded for all t and p.

We impose smoothness assumptions on $\mathbf{\Sigma}_t^{(x)}$. We shall write $(b_t) \in \Lambda_\beta$, $0 < \beta \leq 1$ to denote that the sequence of real numbers b_t has property

$$|b_t| \leq C, \qquad |b_t - b_s| \leq C\Big(\frac{|t-s|}{\min(t,s)}\Big)^{\beta} \quad \text{for all } t, s \geq 1 \qquad (\exists C > 0). \tag{12}$$

The notation $(\mathbf{\Sigma}_t^{(x)}) \in \Lambda_\beta$ means that all elements $\sigma_{ij,t}^{(x)}$, $i, j = 1, ..., p$ satisfy (12) with the same C. Establishing the consistency properties of the regularized estimate will require Λ_1 smoothness of $\mathbf{\Sigma}_t^{(x)}$ and the mean $\boldsymbol{\mu}_t^{(x)}$. Λ_β smoothness is an extension of the concept of Lipschitz β-smoothness of a continuous time function to the discrete time.

The following theorem establishes the consistency properties of the regularised estimate $T_\lambda\big(\hat{\mathbf{\Sigma}}_t\big)$ with hard threshold λ for a nonrandom covariance matrix $\mathbf{\Sigma}_t^{(x)}$. The sample covariance matrix $\hat{\mathbf{\Sigma}}_t$ is defined with the kernel weights

$$b_{H,|t-k|} = K(|t-k|/H), \tag{13}$$

where $H \to \infty$, $H = o(T)$ and $K(x)$, $x \in (0,\, a)$ is a continuous non-negative function with a finite or infinite support $[0,\, a)$, such that $\int K(x)dx = 1$ and for some $C > 0$ and $\nu > 2$

$$K(x) \leq C(1 + x^\nu)^{-1}, \quad |(d/dx)K(x)| \leq C(1 + x^\nu)^{-1}, \quad x \in (0,\, a). \tag{14}$$

In particular, 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 (14).

The hard thresholding, comparing to Theorem 1, is now based on higher λ's which depend the bandwidth H ("window size") and accounts for the change in the mean and the variance of the data.

Theorem 2 *Suppose that* $\boldsymbol{y}_t = \boldsymbol{x}_t$ *where* $(\boldsymbol{x}_t) \in \mathcal{M}$ *is an* α*-mixing process with time varying mean* $\boldsymbol{\mu}_t^{(x)} = E[\boldsymbol{x}_t]$ *and variance* $\boldsymbol{\Sigma}_t^{(x)} = \mathrm{var}(\boldsymbol{x}_t)$*. Suppose that* $(\boldsymbol{\mu}_t^{(x)}) \in \Lambda_1$*,* $(\boldsymbol{\Sigma}_t^{(x)}) \in \Lambda_1$*.*

Then, for any $1 \le t \le T$*, the regularised estimate* $T_\lambda(\hat{\Sigma}_t)$ *of* $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t^{(x)}$ *with threshold*

$$\lambda = \kappa\sqrt{\log p}\,\max(\frac{1}{\sqrt{H}}, \frac{H}{t}) \tag{15}$$

for sufficiently large $\kappa > 0$ *and* H*,* t *such that*

$$c_1 p^{\varepsilon_1} \le H \le c_1 t^{1-\varepsilon_2} \qquad (\exists c_1, c_2 > 0; \varepsilon_1, \varepsilon_2 > 0) \tag{16}$$

has the following properties.

(i) If $(\boldsymbol{x}_t) \in \mathcal{E}(s)$*,* $s > 0$ *then*

$$\left\|T_\lambda(\widehat{\boldsymbol{\Sigma}}_t) - \boldsymbol{\Sigma}_t\right\| = O_p(n_p\lambda), \tag{17}$$

$$\left\|T_\lambda(\widehat{\boldsymbol{\Sigma}}_t)^{-1} - \boldsymbol{\Sigma}_t^{-1}\right\| = O_p(n_p\lambda), \quad \text{if } n_p\lambda = o(1) \text{ and } ||\boldsymbol{\Sigma}_t|| \ge c > 0. \tag{18}$$

(ii) If $(\boldsymbol{x}_t) \in \mathcal{H}(\theta)$*,* $\theta > 4$*, then (17) and (18) hold if in addition* $\varepsilon_1 > 8/(\theta - 4)$ *in (16).*

(iii) Bandwidth $H_{opt} = t^{2/3}$ *yields the lowest threshold (15):* $\lambda_{opt} = \kappa\sqrt{\frac{\log p}{H_{opt}}} = \kappa\frac{\sqrt{\log p}}{t^{1/3}}$*.*

The second setup, $\boldsymbol{y}_t = \boldsymbol{H}_t\boldsymbol{x}_t$ allows for stochastic scaling (volatility) and is crucial for modelling economic and financial data. The scaling (volatility) process $\boldsymbol{H}_t = (h_{ij,t})$ is a random $p \times p$ matrix and appears in the limit $\boldsymbol{\Sigma}_t = \boldsymbol{H}_t\Sigma_t^{(x)}\boldsymbol{H}_t'$ of the regularized estimate $T_\lambda(\hat{\Sigma}_t)$. To ensure the existence of such a limit, the elements $(h_{ij,t})$ of $\boldsymbol{H}_t$ need to be non-stationary persistent (smooth) light-tailed processes. The latter is formalized as follows.

Notation $(\zeta_t) \in L_{1/2}(\alpha)$, $\alpha > 0$ means that (ζ_t) is light-tailed and "smooth":

$$|\zeta_t - \zeta_s| \le \Big(\frac{|t-s|}{\min(t,s)}\Big)^{1/2}\zeta_{ts} \quad \text{for all } t, s \ge 1, \text{ where } (\zeta_{ts}) \in \mathcal{E}(\alpha),\ (\zeta_t) \in \mathcal{E}(\alpha). \tag{19}$$

Notation $(\boldsymbol{H}_t^{(x)}) \in L_{1/2}(\alpha)$ below means that all elements $(h_{ij,t}^{(x)})$ of $\boldsymbol{H}_t$ satisfy (19). In Theorem 3 we assume that $\boldsymbol{\Sigma}_t^{(x)} = \mathrm{var}(\boldsymbol{x}_t)$ is sparse with sparsity parameter n_p which may increase with p, while the scaling matrix $\boldsymbol{H}_t$ for all t has a finite sparsity parameter n_H.

Theorem 3 *Suppose that* $\boldsymbol{y}_t = \boldsymbol{H}_t\boldsymbol{x}_t$ *where* $(\boldsymbol{x}_t) \in \mathcal{M}$ *is an* α*-mixing process. Let* $\boldsymbol{\mu}_t^{(x)} = E[\boldsymbol{x}_t]$ *and* $\boldsymbol{\Sigma}_t^{(x)} = \mathrm{var}(\boldsymbol{x}_t)$ *be time varying and smooth:* $(\boldsymbol{\mu}_t^{(x)}) \in L_{1/2}$*,* $(\boldsymbol{\Sigma}_t^{(x)}) \in \Lambda_{1/2}$*, and* $(\boldsymbol{H}_t) \in L_{1/2}(\alpha)$ *for some* $\alpha > 0$*.*

(a) For t *and* H *as in (16) the regularised estimate* $T_\lambda(\hat{\boldsymbol{\Sigma}}_t)$ *of* $\boldsymbol{\Sigma}_t = \boldsymbol{H}_t\Sigma_t^{(x)}\boldsymbol{H}_t'$ *with*

$$\lambda = \kappa(\log p)^\gamma \max(\frac{1}{\sqrt{H}}, (\frac{H}{t})^{1/2}), \qquad \gamma = (\alpha+4)/(2\alpha) \tag{20}$$

for sufficiently large $\kappa > 0$ is consistent and satisfies (17) and (18) in the following cases.

(i) $(\boldsymbol{x}_t) \in \mathcal{E}(s)$ *with* $s > 0$.

(ii) $(\boldsymbol{x}_t) \in \mathcal{H}(\theta)$ *with* $\theta > 4$ *and (16) holds with* $\varepsilon_1 > 8/(\theta - 4)$.

(b) Bandwidth $H_{opt} = t^{1/2}$ *yields the lowest threshold (20):* $\lambda_{opt} = \kappa \frac{(\log p)^\gamma}{H_{opt}^{1/2}} = \kappa \frac{(\log p)^\gamma}{t^{1/4}}$.

Presence of random scaling (volatility) $\boldsymbol{H}_t$ in the data leads to further increase of threshold. If the elements of $\boldsymbol{H}_t$ have Gaussian distribution and $(\boldsymbol{H}_t) \in L_{1/2}(2)$ then by (b), setting $H = t^{1/2}$, we arrive at the lowest threshold $\lambda = \kappa(\log p)^{3/2} H^{-1/2} = \kappa(\log p)^{3/2} t^{-1/4}$ applicable in this setting.

3 Exponential inequalities

We write $(\xi_t) \in \mathcal{M}$ to denote that $\xi_t - E\xi_t$ is an α-mixing (but not necessary stationary) sequence with the mixing coefficients α_k such that for some $c > 0$ and $0 < \phi < 1$,

$$\alpha_k \leq c_* \phi^k, \quad k \geq 1. \tag{21}$$

By $(\xi_t) \in \mathcal{E}(s)$, $s > 0$ we denote a sequence of random variables $\xi_1, \xi_2, ...$ which marginal distributions have light tails, i.e for some $b_0,\ b_1 > 0$,

$$P[|\xi_t| > \zeta] \leq b_0 \exp\big(-b_1 \zeta^s\big), \quad \zeta > 0, \quad t \geq 1. \tag{22}$$

By $(\xi_j) \in \mathcal{H}(\theta)$, $\theta > 1$ we denote sequence of random variables which marginal distributions have heavy tails, i.e. for some $b_0 > 0$,

$$P[|\xi_t| > \zeta] \leq b_0 |\zeta|^{-\theta}, \quad \zeta > 0, \quad t \geq 1. \tag{23}$$

Theorem 4 of Merlevéde, Peligrad, and Rio (2009) presents a Bernstein type inequality for bounded α- mixing random variables. The following lemma is a minor generalization of their result to a sequence of truncated random variables.

Lemma 1 *Let (ξ_t) be α-mixing sequence of random variables satisfying (21). Suppose that for some $p > 2$, $\max_k (E|\xi_k|^p)^{1/p} =: m^* < \infty$. Set $\xi_{D,k} := \xi_k I(|\xi_k| \leq D)$, $D > 0$.*

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

$$P\Big(\Big|\sum_{k=1}^{T} (\xi_{D,k} - E\xi_{D,k})\Big| \geq \zeta\Big) \leq \exp\Big(-\frac{c\zeta^2}{\bar{v}^2 T + D^2 + \zeta D \log^2 T}\Big), \tag{24}$$

where $\bar{v}^2 = m^*(1 + 24 \sum_{j=1}^{\infty} \alpha(j)^{1-2/p}) < \infty$ and c depends only on c_* appearing in (21).

Proof. By Theorem 14.1 of Davidson (1994) the truncated process $(\xi_{D,t})$ is also α-mixing with mixing coefficients $\tilde{\alpha}(k) \le \alpha(k)$. Thus, the bound (2.3) of Theorem 4 in Merlevéde *et al.* (2009) implies that (24) holds with $\bar{v}^2$ replaced by $v_D^2 = \sup_{i>0} \big(\mathrm{var}(\xi_{D,i}) + 2\sum_{j>i} |\mathrm{cov}(\xi_{D,i}, \xi_{D,j})|\big)$ where c depends only on c_*. To prove (24), it suffices to show that $v_D^2 \le \bar{v}^2$.

Conclusion (2.2) in Davydov (1968) applied with $p = q > 2$ implies that

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

Observe that $\mathrm{var}(\xi_{D,i}) \le E\xi_{D,i}^2 \le (E|\xi_{D,i}|^p)^{2/p} \le m^*$. Hence $v_D^2 \le m^*\big(1+24\sum_{j=1}^{\infty} \alpha(j)^{1-2/p}\big) = \bar{v}^2 < \infty$ which completes the proof. $\square$.

3.1 Exponential inequalities for unbounded variables

Next we establish Bernstein type inequalities for sums $S_T = T^{-1/2}\sum_{k=1}^T \xi_k$ of unbounded dependent α-mixing random variables (ξ_t) with light or heavy tails.

The upper bound in such inequalities will be described by functions

$$f_t(\gamma_1, \gamma_2, c, \zeta) = c_0\Big\{\exp\big(-c_1\zeta^{\gamma_1}\big) + \exp\big(-c_2(\frac{\zeta\sqrt{t}}{\log^2 t})^{\gamma_2}\big)\Big\}, \quad \zeta > 0, \ \ t \ \ge 2, \tag{25}$$

$$g_t(\gamma_1, \theta, c, \zeta) = c_0\Big\{\exp(-c_1\zeta_1^{\gamma}) + \zeta^{-\theta}t^{-(\theta/2-1)}\Big\},$$

where $\gamma_1 > 0$, $\gamma_2 > 0$, $\theta > 2$ and constant $c = (c_0, c_1, c_2)$ does not depends on ζ, t.

In particular, we will show that for $(\xi_j) \in \mathcal{E}(s)$, for all $\zeta \ge 0$, $T \ge 2$,

$$P(|S_T| > \zeta) \le f_T(2, \gamma, c, \zeta) = c_0\Big\{\exp\big(-c_1\zeta^2\big) + \exp\big(-c_2(\frac{\zeta\sqrt{T}}{\log^2 T})^{\gamma}\big)\Big\}, \tag{26}$$

with $\gamma = s/(s+1)$, while for $(\xi_j) \in \mathcal{H}(\theta)$, $\theta > 2$, for any fixed $2 < \theta' < \theta$,

$$P(|S_T| > \zeta) \le g_T(2, \theta', c, \zeta) = c_0\Big\{\exp(-c_1\zeta^2) + \zeta^{-\theta'}T^{-(\theta'/2-1)}\Big\}, \quad \zeta > 0, \ \ T \ge 2. \tag{27}$$

It suffices to establish the bounds (26) and (27) for $\zeta \ge 1$, since they can be extended to $0 < \zeta < 1$ by selecting large enough constant c_0.

The result (28) in Lemma 2 below significantly improves the exponential bound for $P(S_T \ge \zeta)$ obtained in Theorem 3.5 of Wooldridge and White (1991). Its proof combines exponential inequality for bounded random variables by Merlevéde *et al.* (2009) with truncation argument employed in White and Wooldridge (1991).

Lemma 2 *Let $(\xi_j) \in \mathcal{M}$ be a zero mean α-mixing sequence. Then for all $\zeta > 0$, $T \ge 2$,*

$$P(|S_T| \ge \zeta) \le \begin{cases} f_T(2, \gamma, c, \zeta) & \text{if } (\xi_j) \in \mathcal{E}(s), \ s > 0, \qquad (28) \\ g_T(2, \theta', c, \zeta) & \text{if } \ (\xi_j) \in \mathcal{H}(\theta), \ 2 < \theta' < \theta \qquad (29) \end{cases}$$

where $\gamma = s/(s+1)$ and c does not depend on ζ, T.

Proof. Withough restriction of generality assume that $\zeta \geq 1$.

We start with (28). We define the truncated process $w_k = \xi_k I(|\xi_k| \leq D)$, $v_t = \xi_t I(|\xi_t| > D)$ where the truncation constant $D = D_{T,\zeta}$ will be selected later. Set $v_k = \xi_k I(|\xi_k| > D)$. Then $\xi_k = w_k + v_k$, and

$$S_T = T^{-1/2} \sum_{k=1}^T (w_k - Ew_k) + T^{-1/2} \sum_{k=1}^T (v_k - Ev_k) =: s_{T,1} + s_{T,2}. \tag{30}$$

Then $P(|S_T| \geq \zeta) \leq P(|s_{T,1}| \geq \zeta/2) + P(|s_{T,2}| \geq \zeta/2)$. To prove (28), we need to show that $P(|s_{T,i}| \geq \zeta) \leq f_T(2, \gamma, c, \zeta)$, $i = 1, 2$, for some c.

By assumption, $(\xi_t) \in \mathcal{M}$ is α-mixing process which coefficients $\alpha(k)$ as in (21). Hence, by Theorem 14.1 in Davidson (1994), $(w_t) \in \mathcal{M}(\phi)$ and $(v_t) \in \mathcal{M}(\phi)$ are also α-mixing sequences with their respective mixing coefficients $\alpha_w(k)$ and $\alpha_v(k)$ having property

$$\alpha_w(k) \leq \alpha(k), \quad \alpha_v(k) \leq \alpha(k), \qquad k \geq 1. \tag{31}$$

Thus, by Lemma 1, for all $T \geq 2$ and $D > 0$,

$$P\big(|s_{T,1}| \geq \zeta\big) \leq \exp\big(- \frac{c_1 \zeta^2 T}{\bar{v}^2 T + D^2 + \zeta T^{1/2} D \log^2 T}\big) \tag{32}$$

where $c_1 > 0$ does not depend on T and D. Using on the r.h.s. of (32) the inequality

$$1/(|a| + |b| + |c|) \geq 1/3 \min(|a|, b|, |c|),$$

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

$$P\big(|s_{T,1}| \geq \zeta\big) \leq \exp\big(- c_1'^2\big) + \exp\big(- \frac{c_2'^2 T}{D^2}\big) + \exp\big(- \frac{c_2'^{1/2}}{D \log^2 T}\big), \quad \zeta > 0 \tag{33}$$

with $c_1' = c_1/(3v^2)$, $c_2' = c_1/3$. Setting $\Delta_T = T^{1/2}/\log^2 T$, (33) becomes

$$P\big(|s_{T,1}| \geq \zeta\big) \leq \exp\big(- c_1'^2\big) + \exp\big(- c_2'\big(\frac{\zeta \Delta_T}{D}\big)^2 \log^4 T\big) + \exp\big(- \frac{c_2' \zeta \Delta_T}{D}\big).$$

Select D such that $\zeta \Delta_T / D = D^s$. Then,

$$D = \big(\zeta \Delta_T\big)^{1/(s+1)}, \; D^s = \big(\zeta \Delta_T\big)^{s/(s+1)} \text{ and } \big(\zeta \Delta_T\big)/D = \big(\zeta \Delta_T\big)^{s/(s+1)}. \tag{34}$$

For $\zeta \geq 1$, $T \geq 2$ it holds $\zeta \Delta_T \geq \Delta_T \geq 1$. So, and $\big(\zeta \Delta_T\big)/D \geq 1$. Therefore,

$$\begin{aligned} P\big(|s_{T,1}| \geq \zeta\big) &\leq \exp\big(- c_1'^2\big) + 2 \exp\big(- \frac{c_2' \zeta \Delta_T}{D}\big) \\ &\leq \exp\big(- c_1'^2\big) + 2 \exp\big(- c_2' (\zeta \Delta_T)^{s/(s+1)}\big) \\ &\leq 2\Big(\exp\big(- c_1'^2\big) + \exp\big(- c_2'\big(\frac{\zeta \sqrt{T}}{\log^2 T}\big)^{s/(s+1)}\big)\Big). \end{aligned}$$

This proves (28) for $P(|s_{T,1}| \geq \zeta)$. Next, by Markov inequality,

$$P\big(|s_{T,2}| \geq \zeta\big) \leq \zeta^{-2} T^{-1} E\Big(\sum_{k=1}^{T}(v_k - Ev_k)\Big)^2 \leq \zeta^{-2} T^{-1} \sum_{j,k=1}^{T} \mathrm{cov}(v_j, v_k). \qquad (35)$$

Let $p, q > 1$, $1/p + 1/q < 1$ be such that $E|v_j|^p < \infty$, $E|v_k|^q < \infty$. Since $v_t - Ev_t$ 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),

$$|\mathrm{cov}(v_j, v_k)| \leq 12(E|v_j|^p)^{1/p}(E|v_j|^q)^{1/q}\alpha(|j-k|)^{1-1/p-1/q}, \ j \neq k, \qquad (36)$$

while for $j = k$, $\mathrm{var}(v_j) \leq Ev_j^2$. Therefore, setting $V_p := \max_{1\leq j\leq T}(E|v_j|^p)^{1/p}$, we obtain

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

where $e := 1 - 1/p - 1/q > 0$. Then (21) implies

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

Therefore,

$$P\big(|s_{T,2}| \geq \zeta\big) \leq C\zeta^{-2}(V_2^2 + V_p V_q) \qquad (37)$$

where C does not depend on T and D. Set $p = q = 2 + \delta$ where $\delta > 0$ is a small number. Then, by (37),

$$P\big(|s_{T,2}| \geq \zeta\big) \leq C\zeta^{-2}(V_2^2 + V_p^2) \leq C\zeta^{-2} V_p^2 \qquad (38)$$

because $V_2 = \max_j Ev_j^2 \leq \max_j (E|v_j|^p)^{2/p} = V_p^2$. For $D \geq 1$, by (86) it holds

$$E|v_j|^p = E[|\xi_j|^p I(|\xi_j| > D)] \leq c_0' \exp(-c_1'^s)$$

for some c_0', $c_1' > 0$ which do not depend on j and D. This implies $V_p \leq c_0'' \exp(-c_1'' D^s)$. Thus, there exists $c_0 > 0$, $c_2 > 0$ such that for all $\zeta \geq 1$, $T \geq 2$,

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

proving the bound (26) and completing the proof of (28).

Proof of (29). Let $(\xi_j) \in \mathcal{H}(\theta)$ and $s_{T,1}$, $s_{T,2}$ be as in (30). Let $2 < \theta' < \theta$ be fixed. It suffices to show that $P(|s_{T,i}| \geq \zeta) \leq g_T(2, \theta', c, \zeta)$, $i = 1, 2$ for some c. It suffices to consider the case $\zeta \geq 1$.

We start with the evaluation $P(|s_{T,1}| \ge \zeta)$. Set $D = C\zeta\sqrt{T}/\log^3(\zeta\sqrt{T}) \ge 1$ where $C > 0$. From (33) we obtain

$$P\big(|s_{T,1}| \ge \zeta\big) \le \exp\big(-c_1'^2\big) + \exp\big(-c_2'^3(\zeta\sqrt{T}))^2\big) + \exp\big(-c_2'(C/4)\log(\zeta\sqrt{T})\big).$$

Since $\log(\zeta\sqrt{T}) \ge \log(\sqrt{T}) \ge (1/2)\log T$, then for sufficiently large C it holds

$$P\big(|s_{T,1}| \ge \zeta\big) \le \exp\big(-c_1'^2\big) + 2(\zeta\sqrt{T})^{-\theta'} \le 2\big(\exp\big(-c_1'^2\big) + \zeta^{-\theta'}T^{-(\theta'/2-1)}\big).$$

This proves the bound (27) for $P\big(|s_{T,1}| \ge \zeta\big)$.

It remains to evaluate $P(|s_{T,2}| \ge \zeta)$. By (38), we have that $P(|s_{T,2}| \ge \zeta) \le C\zeta^{-2}V_p^2$ with $p = 2+\delta$. For $D \ge 1$, using (87) we obtain $E|v_j|^p = E[|\xi_j|^p I(|\xi_j| > D)] \le c_0'^{-(\theta-p)}$ for some $c_0' > 0$. This implies $V_p^2 \le c_0'^{-(\theta-p)(2/p)}$. Hence,

$$P(|s_{T,2}| \ge \zeta) \le C\zeta^{-2}D^{-(\theta-p)(2/p)} = C\zeta^{-2}(\zeta\sqrt{T})^{-(\theta'-2)}a_{T,\zeta}, \tag{39}$$

where

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

with $\gamma = (\theta-p)(2/p) - (\theta'-2) = \theta - \theta' - \theta(p-2)/p = \theta - \theta' - \theta\delta/p > 0$ when $\delta > 0$ is sufficiently small. Hence $a_{T,\zeta} \le C' < \infty$ where C' does not depend on $\zeta \ge 1$ and $T \ge 2$. Thus, (39) implies $P(|s_{T,2}| \ge \zeta) \le CC'^{-\theta'}T^{-(\theta'/2-1)}$ which proves the bound (27) for $P(|s_{T,2}| \ge \zeta)$.

This completes the proof of (29) and the lemma. □

3.2 Exponential inequalities for weighted variables

In this section we obtain Bernstein type inequalities for sums

$$S_{T,t} := H^{-1/2}\sum_{k=1}^{T} b_{H,|t-k|}(\xi_k - E\xi_k), \quad \tilde{S}_{T,t} := H^{-1/2}\sum_{k=1}^{T} b_{H,|t-k|}(\xi_k - E\xi_t),\ T \ge 2 \tag{40}$$

of α-mixing variables ξ_k with kernel weights $b_{H,k}$ defined as in (13) and (14). We assume that the bandwidth parameter $1 \le H \le T$ and $1 \le t \le T$ may vary with T. Under (14), such weights have property

$$b_{H,k} \le C(1+(k/H)^{\nu})^{-1}, \qquad |b_{H,k} - b_{H,k+1}| \le CH^{-1}(1+(k/H)^{\nu})^{-1}. \tag{41}$$

Hence, in (40) the variables ξ_k with k distant from t, i.e. $|t-k| > H$, are strongly down-weighted. In particular, equal weights $b_{H,k} = I(k = 0,1,...,H)$ generate rolling window sums (40).

The change of the mean $E\xi_k$ in $\tilde{S}_{T,t}$ will be restricted by smoothness assumption $(\xi_k) \in \Lambda_1$,

$$|E\xi_k| \le C, \qquad |E\xi_k - E\xi_t| \le C\Big(\frac{|t-k|}{\min(k,\ t)}\Big), \quad t,k \ge 1. \tag{42}$$

Lemma 3 *Let $S_{T,t}$, $\tilde{S}_{T,t}$ be as in (40), $(\xi_j) \in \mathcal{M}$ and $b_{H,k}$ satisfy (41) with $1 < H \le T$ and $\nu > 2$.*

(a) There exists $c > 0$ such that for all $\zeta > 0$, $0 \le t < T$ and $T \ge 2$,

$$P(|S_{T,t}| \ge \zeta) \le \begin{cases} f_H(2, \gamma, c, \zeta) & \text{if } (\xi_j) \in \mathcal{E}(s),\ s > 0, \quad \text{with } \gamma = s/(s+1), & (43) \\ g_H(2, \theta', c, \zeta) & \text{if } (\xi_j) \in \mathcal{H}(\theta), \quad \text{where } 2 < \theta' < \theta. & (44) \end{cases}$$

(b) In addition, if $(\xi_k) \in \Lambda_1$, then $P(|\tilde{S}_{T,t}| \ge \zeta)$ also satisfies (43)-(44) for $\zeta \ge aH^{3/2}/t$ for sufficiently large a.

We start the proof with the following technical lemma.

Lemma 4 *Let x_{tk}, $k, t \ge 1$ be random variables such that $\max_{t,k} E|x_{tk}| < \infty$ and a_{tk} and $v_{tk} > 0$ be real numbers such that*

$$\max_{t \ge 1} \textstyle\sum_{k=1}^t |a_{tk}| v_{tk} < \infty. \tag{45}$$

Then there exists $\varepsilon > 0$ such that for all $\zeta \ge 1$, $t \ge 2$,

$$P(|\sum_{k=1}^t a_{tk} x_{tk}| \ge \zeta) \le \varepsilon^{-1} \max_{1 \le k \le t} E\Big[\frac{|x_{tk}|}{\zeta v_{tk}} I(\frac{|x_{tk}|}{\zeta v_{tk}} \ge \varepsilon)\Big]. \tag{46}$$

Proof. By (45) there exists $\varepsilon > 0$ such that $\sum_{k=1}^t |a_{Tk} v_{tk}| < 1/(2\varepsilon)$, $t \ge 1$. Then $|x_{tk}| = |x_{tk}| I\big(|x_{_k}| < \varepsilon\zeta v_{tk}\big) + |x_{tk}| I\big(|x_{tk}| \ge \varepsilon\zeta v_{tk}\big) \le \varepsilon\zeta v_{tk} + x'_{tk,\zeta}$ where $x'_{tk,\zeta} := I\big(|x_{tk}| > \varepsilon\zeta v_{tk}\big)$. This implies

$$|\textstyle\sum_{k=1}^t a_{tk} x_{tk}| \le \zeta/2 + B_{t,\zeta}, \quad B_{t,\zeta} := \sum_{k=1}^t |a_{Tk}| x'_{tk,\zeta}.$$

Consequently,

$$\begin{aligned} P(|\textstyle\sum_{k=1}^t a_{tk} x_{tk}| \ge \zeta) &\le P(B_{T,\zeta} \ge \zeta/2) \le 2\zeta^{-1} \textstyle\sum_{k=1}^t |a_{tk}| E x'_{tk,\zeta} \\ &\le 2\zeta^{-1} (\textstyle\sum_{k=1}^t |a_{tk}| v_{tk}) \max_{1 \le k \le t} v_{tk}^{-1} E[x'_{tk,\zeta}] \le \varepsilon^{-1} \max_{1 \le k \le t} (\zeta v_{tk})^{-1} E[x'_{tk,\zeta}]. \quad \Box \end{aligned}$$

Proof of Lemma 3. Without restriction of generality assume that $\zeta \ge 1$.

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

$$S_{T,t} = H^{-1/2} \textstyle\sum_{k=1}^{t-1} b_{H,k} (\xi'_k - E\xi'_k) + H^{-1/2} \sum_{k=0}^{T-t} b_{H,k} (\xi''_k - E\xi''_k) =: s^{(1)}_{T,t} + s^{(2)}_{T,t}. \tag{47}$$

It suffices to show that $P(|s^{(\ell)}_{T,t}| \ge \zeta)$, $\ell = 1, 2$ satisfy the bounds (43)-(44). We provide the proof for $s^{(1)}_{T,t}$. (For $s^{(2)}_{T,t}$ the proof is similar). Set $x_k = \sum_{i=1}^k (\xi'_i - E\xi'_i)$, $k \le t-1$. Summation by parts gives

$$s^{(1)}_{T,t} = H^{-1/2} \textstyle\sum_{k=1}^{t-2} (b_{H,k} - b_{H,k+1}) x_k + H^{-1/2} b_{H,t-1} x_{t-1} = \sum_{k=1}^{t-1} a_{tk} x_k \tag{48}$$

where $a_{tk} = H^{-1/2}(b_{H,k} - b_{H,k+1})$, $k \le t-2$; $a_{t,t-1} = H^{-1/2} b_{H,t-1}$. Set $v_{tk} = (\max(k,H))^{1/2}$. By (41),

$$\begin{aligned}\textstyle\sum_{k=1}^{t-1} |a_{tk}| v_{tk} &\le \textstyle\sum_{k=1}^{t-2} |b_{H,k} - b_{H,k+1}| \big(\frac{\max(k,H)}{H}\big)^{1/2} + b_{H,t-1} \big(\frac{\max(t-1,H)}{H}\big)^{1/2} \\ &\le CH^{-1} \textstyle\sum_{k=1}^{t-2} (1 + (k/H)^{p-1/2})^{-1} + C \big(\frac{\max(t-1,H)}{H}\big)^{-p+1/2} \le C, \quad t \ge 2.\end{aligned}$$

Therefore, by Lemma 4, there exists $\varepsilon > 0$ such that

$$p_{T,\zeta} \le \varepsilon^{-1} \max\nolimits_{1 \le k < t} (\zeta v_{tk})^{-1} E[|x_k| I(|x_k| > \varepsilon \zeta v_{tk})],$$

and, setting $y_k = k^{-1/2} x_k$, we obtain

$$p_{T,\zeta} \le \varepsilon^{-1} \max_{1 \le k < t} (\zeta \nu_k)^{-1} E[|y_k| I(|y_k| \ge \varepsilon \zeta \nu_k], \quad \nu_k := \big(\frac{\max(k,H)}{k}\big)^{1/2}. \tag{49}$$

Notice that $\nu_k = 1$ for $k \ge H$; $\nu_k = (H/k)^{1/2} \ge 1$ for $k < H$.

Proof of (43). Let $(\xi_k) \in \mathcal{E}(s)$. Then, (28) implies $P(|y_k| \ge \zeta) \le f_k(2, \gamma, c, \zeta)$, $\zeta > 0, k \ge 2$. Thus, from Lemma 6(iii) we obtain $E[|y_k| I(|y_k| \ge \varepsilon \zeta \nu_k] \le f_k\big(2, \gamma, c', \varepsilon \zeta \nu_k\big)$ for some c'. Thus, by (49),

$$p_{T,\zeta} \le C \max_{1 \le k < t} f_k\big(2, c'; \varepsilon \zeta \nu_k\big). \tag{50}$$

Since $f_k \ge f_s$ if $k > s$, then for $k \ge H$, $f_k\big(2, \gamma, c, \varepsilon \zeta \nu_k\big) = f_k\big(2, \gamma, c, \varepsilon \zeta\big) \le f_H\big(2, \gamma, c, \varepsilon \zeta\big)$. In turn, for $1 \le k < H$, it holds $\nu_k = (H/k)^{1/2} \ge 1$, $\nu_k \sqrt{k} = \sqrt{H}$, and therefore,

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

Together with (50), this yields $p_{T,\zeta} \le f_H(2, \gamma, c, \zeta)$ which implies (43).

Proof of (44). Let $(\xi_j) \in \mathcal{H}(\theta)$. Then by (29), $P(|y_k| \ge \zeta) \le g_k(2, \theta', c, \zeta)$ for $k \ge 2$, and from Lemma 6(iv) we obtain $E[|y_k| I(|y_k| \ge \varepsilon \zeta \nu_k)] \le (\zeta \nu_k)^{-1} g_k\big(2, \theta', c, \varepsilon \zeta \nu_k\big)$ for some c. Since $\zeta \nu_k \ge 1$, by (49),

$$p_{T,\zeta} \le C \max_{1 \le k < t} g_k\big(2, \theta', c', \varepsilon \zeta \nu_k\big). \tag{51}$$

Let $k \ge H$. Then $\nu_k = 1$ and $g_k\big(2, \theta', c, \varepsilon \zeta \nu_k\big) = g_k\big(2, \theta', c, \varepsilon \zeta\big) \le g_H\big(2, \theta', c, \varepsilon \zeta\big)$ since $g_k \le g_s$ if $k > s$ (see definition (27)).

Let $k \le H$. Then $\nu_k = (H/k)^{1/2}$ and $(\zeta \nu_k)^{-\theta'} k^{-(\theta'/2-1)} \le \zeta^{-2} (\zeta \nu_k \sqrt{k})^{-\theta'+2} = \zeta^{-\theta'} H^{-(\theta'/2-1)}$ which implies $g_k\big(2, \theta', c', \varepsilon \zeta \nu_k\big) \le g_H\big(2, \theta', c', \varepsilon \zeta\big)$. Together with (51), this implies $p_{T,\zeta} \le g_H(2, \theta', c, \zeta)$ which proves (44).

(b) Write $\tilde{S}_{T,t} = S_{T,t} + r_{T,t}$ where $r_{T,t} := H^{-1/2}\sum_{k=1}^{T} b_{H,|t-k|}(E\xi_k - E\xi_t)$. If $|r_{T,t}| < \zeta/2$, then

$$P(|\tilde{s}_{T,t}| \ge \zeta) \le P\big(|s_{T,t}| + |r_{T,i}| \ge \zeta\big) \le P\big(|s_{T,t}| \ge \zeta/2\big).$$

Hence, it suffices to show that there exists $C > 0$ such that $|r_{T,t}| \le \zeta/2$ when $\zeta \ge C(H^{3/2}/t)$. By (42), $|E\xi_k - E\xi_t| \le C(|k-t|/\min(t,k))$. By assumption, $b_{H,k}$ satisfies (41) with parameter $\nu > 2$. Hence using Lemma 8 with $\gamma = 1$ we obtain that for $H/t \le 1/(\log t)^{1/(\nu-2)}$,

$$|r_{T,t}| \le CH^{-1/2}\sum_{k=1}^{T} b_{H,|t-k|}(\tfrac{|t-k|}{\min(t,k)}) \le C(H^{3/2}/t)\big(1 + (H/t)^{\nu-2}\log t\big) \le C_1(H^{3/2}/t).$$

Hence $|r_{T,t}| \le \zeta/2$ when $\zeta \ge 2C_1(H^{3/2}/t)$. This completes the proof of the part (b). □

3.3 Exponential inequalities with random scaling

In this section we establish exponential inequalities for sums

$$S^{(h)}_{T,t} := H^{-1/2}\sum_{k=1}^{T} b_{H,|t-k|}h_k(\xi_k - E\xi_k), \quad \tilde{S}^{(h)}_{T,t} := H^{-1/2}\sum_{k=1}^{T} b_{H,|t-k|}(h_k\xi_k - h_tE\xi_t), \quad (52)$$

of an α- mixing process ξ_k in presence of the random scaling ("volatility"). Differently from ARCH modelling, where h_k is a stationary volatility process, in our setting h_t is assumed to be a persistent non-stationary process. Its persistence is guarantied by smoothness assumption

$$h_t - h_k = \big(\frac{|t-k|}{\min(t,k)}\big)^{1/2}\xi_{tk}, \qquad t,k \ge 1, \quad (53)$$

and the magnitude of ξ_{tk} and h_k is controlled by requirement of light tailed distribution:

$$(h_k) \in \mathcal{E}(\alpha), \quad (\xi_{tk}) \in \mathcal{E}(\alpha), \qquad \exists \alpha > 0. \quad (54)$$

We suppose that kernel weights $b_{H,k}$ satisfy (41). In (52) we assume that that $0 \le t \le T$ and $1 < H \le T$ may vary with T.

Lemma 5 *Let $S^{(h)}_{T,t}$, $\tilde{S}^{(h)}_{T,t}$ be as in (52). Suppose that $b_{H,t}$ satisfies (41) with parameters $\nu > 2$ and H, and (h_t) satisfy (53) and (54) with parameter $\alpha > 0$. Let $0 < \theta < 1$.*

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

$$P(|S^{(h)}_{T,t}| \ge \zeta) \le \begin{cases} f_H\big(\gamma_1, \gamma_2, c, \zeta\min(1, t^{1/2}/H)\big) & \text{if } (\xi_j) \in \mathcal{E}(s),\ s > 0, \quad (55) \\ g_H\big(\gamma_1, \theta', c, \zeta\min(1, t^{1/2}/H)\big) & \text{if } (\xi_j) \in \mathcal{H}(\theta),\ 2 < \theta' < \theta, \quad (56) \end{cases}$$

where $\gamma_1 = 2\alpha/(2+\alpha)$, $\gamma_2 = \alpha s/(\alpha + s + 1)$.

In addition, $P(|\tilde{S}^{(h)}_{T,t}| \ge \zeta)$ satisfies corresponding bounds (55)-(56) if

$$|E\xi_k| \le C, \qquad |E\xi_k - E\xi_t| \le C(\frac{|t-k|}{\min(k,t)})^{1/2}, \quad t,k \ge 1. \quad (57)$$

Proof of Lemma 5. Without restriction of generality assume that $\zeta \geq 1$.

Denote $h'_k := h_{t-k}$, $\xi'_k := \xi_{t-k}$, $h''_k := h_{t+k}$, $\xi''_k := \xi_{t+k}$ for $k \geq 0$. As in (47) write

$$S^{(h)}_{T,t} = H^{-1/2} \sum_{k=1}^{t-1} b_{H,k} h'_k \xi'_k + H^{-1/2} \sum_{k=0}^{T-t} b_{H,k} h''_k \xi''_k =: s^{(h)}_{T,t;1} + s^{(h)}_{T,t;2}.$$

Proof of (55)-(56) for $S^{(h)}_{T,t}$ reduces to verification of these bounds for $s^{(h)}_{T,t;1}$ and $s^{(h)}_{T,t;2}$.

We start with $s^{(h)}_{T,t;1}$. As in (48), setting $x_k = \sum_{i=1}^{k} (\xi'_i - E\xi'_i)$, $k \geq 1$, summation by parts yields

$$\begin{aligned} s^{(h)}_{T,t;1} &= H^{-1/2} \sum_{k=1}^{t-2} (b_{H,k} h'_k - b_{H,k+1} h'_{k+1}) x_k + H^{-1/2} b_{H,t-1} h'_{t-1} \\ &= \Big\{ H^{-1/2} \sum_{k=1}^{t-2} (b_{H,k} - b_{H,k+1})(h'_k x_k) + H^{-1/2} b_{H,t-1} h'_{t-1} x_{t-1} \Big\} \\ &\quad + H^{-1/2} \sum_{k=1}^{t-2} b_{H,k} (h'_k - h'_{k+1}) x_k =: s^{(1)}_{T,t;1} + s^{(2)}_{T,t;1}. \end{aligned} \tag{58}$$

Hence, it suffices to verify (55)-(56) for $s^{(1)}_{T,t;1}$ and $s^{(2)}_{T,t;1}$.

1) The sum $s^{(1)}_{T,t;1}$ can be obtained from $s^{(1)}_{T,t}$ in (48) replacing x_k by $h_k x_k$. Therefore, by the same argument as in the proof of (49), there exists $\varepsilon > 0$ such that

$$p'_{T,\zeta} := P(|s^{(1)}_{T,t}| \geq \zeta) \leq C \max_{k=1,\dots,T} \zeta_k^{-1} E[|y'_k| I(|y'_k| \geq \varepsilon \zeta_k], \quad \zeta \geq 1, \;\; T \geq 2, \tag{59}$$

where $y'_k = h'_k y_k$, $y_k = k^{-1/2} x_k$, and $\nu_k := \big(\frac{\max(k,H)}{k}\big)^{1/2}$.

1a) Let $(\xi_k) \in \mathcal{E}(s)$. First we will show that

$$P(|y'_k| \geq \zeta) \leq f_k(\gamma_1, \gamma_2, c, \zeta), \qquad \zeta > 0, \;\; k \geq 2. \tag{60}$$

Assumption (54) implies $(h'_k) \in \mathcal{E}(\alpha)$, while (28) implies $P(|y_k| \geq \zeta) \leq f_k(2, \gamma, c, \zeta)$, $\zeta > 0$, $k \geq 2$. Therefore (60) follows from (80). As seen in the proof of (43), bounds (60) and (59) imply $p'_{T,\zeta} \leq f_H\big(\gamma_1, \gamma_2, c, \zeta\big)$. Thus, $p'_{T,\zeta}$ satisfies (55).

1b) Let $(\xi_j) \in \mathcal{H}(\theta)$, and $0 < \theta' < \theta$. Then,

$$P(|y'_k| \geq \zeta) \leq g_k(\gamma_1, \theta', c, \zeta), \qquad \zeta > 0, \;\; k \geq 2 \tag{61}$$

where $\gamma = 2\alpha/(2+\alpha)$. To establish (61), note that $(h'_k) \in \mathcal{E}(\alpha)$ and (29) implies $P(|y_k| \geq \zeta) \leq g_k(2, \theta', c, \zeta)$. Hence, (61) follows from (81). In turn, (59) and (61) imply $p'_{T,\zeta} \leq g_H\big(\gamma_1, \theta', c, \zeta\big)$ as seen in the proof of (44). Hence $p'_{T,\zeta}$ satisfies (56).

2) Next we establish corresponding bounds for $s^{(2)}_{T,t;1}$. To evaluate $p''_{T,\zeta} := P(|s^{(2)}_{T,t;1}| \geq \zeta)$, we use Lemma 4. Denote $\xi_{tk} = (t-k)^{1/2}(h'_k - h'_{k+1})$, $1 \leq k \leq t-2$. Then,

$$s^{(2)}_{T,t;1} = H^{-1/2} \sum_{k=1}^{t-2} b_{H,k} (h'_k - h'_{k+1}) x_k = \sum_{k=1}^{t-2} \frac{b_{H,k}}{H^{1/2}(t-k)^{1/2}} \xi_{tk} x_k,$$

Set $v_{Tk} = (\max(k,H)^{1/2}/(H/t^{1/2})$. By (93),

$$\sum_{k=1}^{t-2} \frac{b_{H,k} v_{Tk}}{H^{1/2}(t-k)^{1/2}} \le \frac{1}{(H/t)^{1/2}} H^{-1} \sum_{k=1}^{T} b_{H,|t-k|} (\frac{|t-k| \vee H}{k})^{1/2} \le C, \qquad T \ge 2.$$

Hence, by Lemma 4, there exists $\varepsilon > 0$ such that

$$p''_{T,\zeta} :\le \varepsilon^{-1} \max_{1 \le k \le t-2} E\Big[\frac{|\xi_{tk} x_k|}{\zeta v_{Tk}} I(\frac{|\xi_{tk} x_k|}{\zeta v_{Tk}} \ge \varepsilon)\Big].$$

Setting $y''_k = \xi_{tk} y_k$ where $y_k = k^{-1/2} x_k$ and writing $\zeta v_{Tk} k^{-1/2} = \zeta' \nu_k$ where $\nu_k = \big(\max(k,H)/k\big)^{1/2}$ and $\zeta'^{1/2}/H)$, we obtain

$$p''_{T,\zeta} \le C \max_{1 \le k \le t-2} E\Big[\frac{|y''_k|}{\zeta' \nu_k} I(\frac{|y''_k|}{\zeta' \nu_k} \ge \varepsilon)\Big]. \tag{62}$$

By (53)-(54), the variables $\xi_{tk} = (t-k)^{1/2}(h_{t-k} - h_{t-k+1})$ have property $(\xi_{tk}) \in \mathcal{E}(\alpha)$. Observe that the bound (62) for $p''_{T,\zeta}$ is of the the same type as (59) for $p'_{T,\zeta}$

If $(\xi_k) \in \mathcal{E}(s)$, then, arguing as in 1a) above we obtain that $p''_{T,\zeta} \le f_H\big(\gamma_1, \gamma_2, c, \zeta'\big)$, and therefore $p''_{T,\zeta}$ satisfies (55). If $(\xi_j) \in \mathcal{H}(\theta)$, and $0 < \theta' < \theta$ then same argument as 1b) above implies $p''_{T,\zeta} \le g_H\big(\gamma_1, \theta', c, \zeta'\big)$, and whence $p''_{T,\zeta}$ satisfies (56).

This completes the proof of (55)-(56) for $p''_{T,\zeta}$, and for $s^{(h)}_{T,t;1}$, too.

Similar arguments as above allow to establish the bounds (55)-(56) for $s^{(h)}_{T,t;2}$, which completes the proof of Lemma 5a).

To prove Lemma 5b), we need to verify (55)-(56) for $P(|\tilde{S}^{(h)}_{T,t}| \ge \zeta)$. Write

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

Since $S^{(h)}_{T,t}$ satisfies (55)-(56), it suffices to show that $r_{T,t}$ satisfies (55)-(56) too. By (57),

$$|h_k E\xi_k - h_t E\xi_t| \le |h_k(E\xi_k - E\xi_t)| + |E\xi_t|\,|h_k - h_t| \le C(\frac{|t-k|}{\min(t,k)})^{1/2}(|h_k| + |\tilde{\xi}_{tk}|)$$

where $\tilde{\xi}_{tk} := (\frac{\min(t,k)}{|t-k|})^{1/2}(h_k - h_t)$. Hence,

$$|r_{T,t}| \le C H^{-1/2} \sum_{k=1}^{T} b_{H,|t-k|}(\frac{|t-k|}{\min(t,k)})^{1/2} z_k, \quad z_k := |h_k| + |\tilde{\xi}_{tk}|.$$

Let $v_{Tt} = \max(1, t^{1/2}/H)$. By Lemma 8, $H^{-1/2} \sum_{k=1}^{T} b_{H,|t-k|}(\frac{|t-k|}{t \wedge k})^{1/2} v_t \le C$. Therefore, by Lemma 4, there exists $\varepsilon > 0$ such that

$$P(|r_{T,t}| \ge \zeta) \le \varepsilon^{-1} \max_{1 \le k \le T} E\Big[\frac{|z_k|}{\zeta v_{Tt}} I(\frac{|z_k|}{\zeta v_{Tt}} \ge \varepsilon)\Big]. \tag{63}$$

Assumption (54) implies that $(z_k) \in \mathcal{E}(\alpha)$. Therefore, $P(|z_k| \ge \zeta) \le c_0 \exp(-c_1|\zeta|^\alpha)$, $\zeta > 0$, $k \ge 1$, and by Lemma 7(i), $E[|z_k| I(|z_k| \ge \varepsilon \zeta v_{Tt})] \le c'_0 \exp(-c'_1|\zeta v_{Tt}|^\alpha)$. Noting that $(\zeta v_{Tt})^{-1} \le 1$ for $\zeta \ge 1$, this together with (63) imply $P(|r_{T,t}| \ge \zeta) \le c'_0 \exp(-c'_1|\zeta v_t|^\alpha)$.

Since $\alpha \ge \gamma_1 = 2\alpha(2+\alpha)$, clearly $P(|r_{T,t}| \ge \zeta)$ satisfies the bounds (55)-(56) which proves the claim of the part b) of the lemma. □

Proofs of Theorems 1-3

Proof of Theorem 2. The proof of Theorem 1, p. 2582-2584 in Bickel and Levina (2008) implies that for $\lambda > 0$ and $0 < \epsilon < 1$,

$$||T_\lambda(\hat{\Sigma}_t) - \hat{\Sigma}_t|| \le 2\lambda n_p + 3Mn_p + MN$$

where $M = \max_{i,j} |\hat{\sigma}_{ij,t} - \sigma_{ij,t}|$, $N = \max_i \sum_{j=1}^p I\big(|\hat{\sigma}_{ij,t} - \sigma_{ij,t}| \ge (1-\epsilon)\lambda\big)$. Set for simplicity $\epsilon = 1/2$. We will show that $M = O_P(\lambda)$ by verifying that

$$P\big(M \ge \lambda/2\big) \to 0. \tag{64}$$

The latter yields $N = O_P(\lambda)$ because $P\big(N > 0\big) \le P\big(M \ge \lambda/2\big) \to 0$. This proves the required claim (17): $||T_\lambda\big(\hat{\Sigma}_t\big) - \hat{\Sigma}_t|| = O_p(\lambda n_p)$. In turn, to prove (64), we will show that

$$\max_{ij} P\big(|\hat{\sigma}_{ij,t} - \sigma_{ij,t}| > \lambda/2\big) = o(p^{-2}) \tag{65}$$

uniformly in i, j which implies $P\big(M \ge \lambda/2\big) \le \sum_{ij} P\big(|\hat{\sigma}_{ij,t} - \sigma_{ij,t}| \ge \lambda/2\big) = o(1)$, $p \to \infty$.

Set $z_k = y_{ik}y_{jk}$. Notice that $\sigma_{ij,t} = \mathrm{cov}(y_{it}, y_{jt}) = Ez_t - Ey_{it}Ey_{jt}$. Then,

$$\begin{aligned}
\hat{\sigma}_{ij,t} - \sigma_{ij,t} &= K_t^{-1} \textstyle\sum_{k=1}^T b_{H,|t-k|} y_{ik}y_{jk} - \bar{y}_{it}\bar{y}_{jt} - \sigma_{ij,t} \\
&= s_{T,ij,t} - (\bar{y}_{it} - Ey_{it})(\bar{y}_{jt} - Ey_{jt}), \ \ s_{T,ij,t} := K_t^{-1} \textstyle\sum_{k=1}^T b_{H,|t-k|}(z_k - Ez_t), \text{ and} \\
P\big(|\hat{\sigma}_{ij,t} - \sigma_{ij,t}| > \lambda/2\big) &\le P\big(|s_{T,ij,t}| \ge \lambda/4\big) + P\big(|\bar{y}_{it} - Ey_{it}|\,|\bar{y}_{jt} - Ey_{jt}| \ge \lambda/4\big) \\
&\le P\big(|s_{T,ij,t}| \ge \lambda/4\big) + P\big(|\bar{y}_{it} - Ey_{it}| \ge \sqrt{\lambda/4}\big) + P\big(|\bar{y}_{jt} - Ey_{jt}| \ge \sqrt{\lambda/4}\big).
\end{aligned} \tag{66}$$

Hence, to prove (65), it suffices to show that uniformly in i, j,

$$\text{a)} \quad P\big(|s_{T,ij,t}| > \lambda/4\big) = o(p^{-2}), \qquad \text{b)} \quad P\big(|\bar{y}_{it} - Ey_{it}| > (\lambda/4)^{1/2}\big) = o(p^{-2}). \tag{67}$$

We now turn to the proof of (67). First we gather intermediate facts. Assumption $(\boldsymbol{y}_t) \in \mathcal{M}$ implies that the process $z_k - Ez_k$ is α-mixing: $(z_k) \in \mathcal{M}$. Observe that $K_t/H \to 1$ while definition $\lambda = \kappa(\log p)^{1/2} \max(H^{-1/2}, H/t)$ and (16) imply $\lambda \le \kappa c_*$ where $c_* < \infty$, and

$$\lambda H^{1/2} \ge \kappa(\log p)^{1/2}, \quad (\lambda H)^{1/2} = \lambda^{-1/2}(\lambda H^{1/2}) \ge (\kappa/l_*)^{1/2}(\log p)^{1/2}. \tag{68}$$

(i) Let $(y_{ik}) \in \mathcal{E}(s)$. Then, $(z_k) \in \mathcal{E}(s/2)$, while property $(Ez_k) \in \Lambda_1$ follows from equality $Ez_k = Ey_{ik}y_{jk} = \mathrm{cov}(y_{ik}, y_{jk}) + Ey_{ik}Ey_{jk}$ using assumptions $(\sigma_{ij,k}) \in \Lambda_1$, $(Ey_{ik}) \in \Lambda_1$.

Hence, from Lemma 3(b) and (43) we obtain that with $\gamma = (s/2)(1 + s/2)$,

$$\begin{aligned}
&P\big(|s_{T,ij,t}| > \lambda/4\big) = P\big(H^{1/2}|s_{T,ij,t}| > \lambda H^{1/2}/4\big) \le f_H(2, \gamma, c, \lambda H^{1/2}/4), \\
&P\big(|\bar{y}_{it} - Ey_{it}| > (\lambda/4)^{1/2}\big) = P\big(|H^{1/2}(\bar{y}_{it} - Ey_{it})| > (\lambda H/4)^{1/2}\big) \le f_H(2, \gamma, c, (\lambda H/4)^{1/2}).
\end{aligned} \tag{69}$$

The function $f_H(2,\gamma,c,\zeta)$ given by (25) is non-increasing in ζ. Hence, by (68) the r.h.s. of (69) can be bounded by $f_H(2,\gamma,c,\kappa'^{1/2})$ where $\kappa'=\min(\kappa/4,(\kappa/(4c_*))^{1/2})$. Since

$$f_H(2,\gamma,c,\kappa'^{1/2})\le c_0\Big\{\exp(-c_1\kappa'^2\log p)+\exp\big(-c_2\big(\kappa'^{1/2}\frac{H^{1/2}}{\log^2 H}\big)^{\gamma}\Big\}=o(p^{-2})$$

because $c_1\kappa'^2>2$ for large enough κ and because $\log p=o\big((H^{1/2}/\log^2 H)^{\gamma}\big)$ under assumption (16). This proves (67)(a,b) and (17).

(ii) Let $(y_{ik})\in\mathcal{H}(\theta)$. Then, $(z_k)\in\mathcal{H}(\theta/2)$, and as in (i), $(Ez_k)\in\Lambda_1$. Therefore, Lemma 3(b) and (44) yield that for any $\theta'\in(2,\,\theta/2)$,

$$\begin{aligned}&P\big(|s_{T,ij,t}|>\lambda/4\big)=P\big(H^{1/2}|s_{T,ij,t}|>\lambda H^{1/2}/4\big)\le g_H(2,\theta'^{1/2}/4),\\&P\big(|\bar y_{it}-Ey_{it}|>(\lambda/4)^{1/2}\big)=P\big(|H^{1/2}(\bar y_{it}-Ey_{it})|>(\lambda H/4)^{1/2}\big)\le g_H(2,\theta'^{1/2}).\end{aligned}\qquad(70)$$

Since $g_H(2,\gamma,c,\zeta)$ given by (25) is a non-increasing function in ζ, then by (68) the r.h.s. of (70) can be bounded by $g_H(2,\theta',c,\kappa'^{1/2})$ with κ' as in (i). In turn,

$$g_H(2,\theta',c,\kappa'^{1/2})\le c_0\Big\{\exp(-c_1\kappa'^2\log p)+\frac{1}{(\kappa'^{1/2})^{\theta'}}\frac{1}{H^{\theta'/2-1}}\Big\}=o(p^{-2})\qquad(71)$$

because $c_1\kappa'^2>2$ for large enough κ and because $p^2=o(H^{\theta'/2-1})$ under assumption $H\ge c_1p^{\varepsilon_1}$, $\varepsilon_1>8/(\theta-4)$ if θ' is selected close enough to $\theta/2$. This verifies (67)(a,b) and (17).

To prove (18), set $B:=T_\lambda\big(\hat\Sigma_t\big)$, $A:=\Sigma_t$. By assumption, $\|A\|\ge c>0$ and $n_p\lambda=o(1)$. By (17), $\|B-A\|=O_p(n_p\lambda)=o_p(1)$. Thus $\|B\|\ge\|A+(B-A)\|\ge\|A\|-\|B-A\|\ge c-o_p(1)\ge c(1+o_p(1))$. This implies $\|B^{-1}\|=O_P(1)$.

Hence, $\|B^{-1}-A^{-1}\|=\|A^{-1}(A-B)B^{-1}\|\le\|A^{-1}\|\,\|A-B\|\,\|B^{-1}\|\le c^{-1}O_p(n_p\lambda)O_P(1)$ $=O_p(n_p\lambda)$ which proves (18).

(iii) Bandwidth $H=t^{2/3}$ minimizes $\max(H^{-1/2},(H/T))$, so $\lambda=\kappa(\log p)^{1/2}\max(H^{-1/2},(H/T))\ge\lambda_{opt}=\kappa(\log p)^{1/2}t^{-1/3}$. □

Proof of Theorem 3. We start with the proof of the part (a).

As in Theorem 2, to prove (17) it suffices to verify (65). Observe that $\boldsymbol{y}_k=\boldsymbol{H}_k\boldsymbol{x}_t=(y_{1k},...,y_{pk})$, $y_{ik}=\sum_{u=1}^p h_{iu,k}x_{uk}$, $y_{ik}y_{jk}=\sum_{u,v=1}^p h_{iu,k}h_{jv,k}x_{uk}x_{vk}$, and $\boldsymbol{\Sigma}_t=\boldsymbol{H}_t\boldsymbol{\Sigma}_t^{(x)}\boldsymbol{H}_t'=(\sigma_{ij,t})$ where $\sigma_{ij,t}=\sum_{u,v=1}^p h_{iu,t}h_{ju,t}\sigma_{uv,t}^{(x)}$. Since $\sigma_{uv,t}^{(x)}=E[x_{ut}x_{vt}]+Ex_{ut}Ex_{vt}$, then $\sigma_{ij,t}=\sum_{u,v=1}^p h_{iu,t}h_{ju,t}E[x_{ut}x_{vt}]+(\sum_{u=1}^p h_{iu,t}E[x_{ut}])(\sum_{v=1}^p h_{iv,t}E[x_{vt}])$. Then,

$$\begin{aligned}\hat\sigma_{ij,t}-\sigma_{ij,t}&=K_t^{-1}\textstyle\sum_{k=1}^T b_{H,|t-k|}y_{ik}y_{jk}-\bar y_{it}\bar y_{jt}-\sigma_{ij,t}\\&=\textstyle\sum_{u,v=-1}^p\tilde s_{ij,uv,t}+\sum_{u=1}^p s_{iu,t}\sum_{v=1}^p s_{iv,t}\quad\text{where}\\\tilde s_{ij,uv,t}&:=K_t^{-1}\textstyle\sum_{k=1}^T b_{H,|t-k|}\big(h_{iu,k}h_{jv,k}x_{uk}x_{vk}-h_{iu,t}h_{jv,t}E[x_{ut}x_{vt}]\big),\\s_{iu,t}&:=K_t^{-1}\textstyle\sum_{k=1}^T b_{H,|t-k|}\big(h_{iu,k}x_{uk}-h_{iu,t}Ex_{ut}\big).\end{aligned}$$

Because sparsity parameter n_H of $\boldsymbol{H}_t$ is finite and does not depend on t, p, for any fixed (i,j) the sum $\sum_{u,v=1}^{p} \tilde{s}_{ij,uv,t}$ includes only a finite number ($\le n_H^2$) of non-zero terms $\tilde{s}_{ij,uv,t}$. Similarly, the sum $\sum_{u=1}^{p} s_{iu,t}$ includes only a finite number ($\le n_H$) of non-zero terms $s_{iu,t}$. Thus, to prove (65), similarly as in (67), it suffices to verify that for $\nu = (4n_H^2)^{-1}$ uniformly in i, j, it holds

$$\text{a)} \quad P\big(|\tilde{s}_{ij,uv,t}| > \lambda\nu\big) = o(p^{-2}), \quad \text{b)} \quad P\big(|s_{iu,t}| > \sqrt{\lambda\nu}\big) = o(p^{-2}). \tag{72}$$

Let i, j, u, v be fixed. Define $\tilde{x}_k := x_{uk}x_{vk}$, $\tilde{h}_k := h_{iu,k}h_{jv,k}$. Assumption $(x_{uk}) \in \mathcal{M}$ implies $(\tilde{x}_k) \in \mathcal{M}$ while smoothness assumption $(Ex_{uk}) \in \Lambda_{1/2}$ implies $(E\tilde{x}_k) \in \Lambda_{1/2}$. Moreover, it is easy to see that assumption $(h_{ij,k}) \in L_{1/2}(\alpha/2)$ implies $(\tilde{h}_k) \in L_{1/2}(\alpha/2)$. By definition, $\lambda = \kappa(\log p)^{1/\gamma_1} \max(H^{-1/2}, (H/t)^{1/2})$, $\gamma_1 = 2\alpha/(\alpha+4)$. This together with (16) implies $\lambda \le \kappa c_*$ where $c_* < \infty$. Therefore,

$$\lambda H^{1/2} \ge \kappa(\log p)^{1/\gamma_1}, \qquad (\lambda H)^{1/2} = \lambda^{-1/2}(\lambda H^{1/2}) \ge (\kappa/l_*)^{1/2}(\log p)^{1/\gamma_1}. \tag{73}$$

(i) Let $(x_{it}) \in \mathcal{E}(s)$. Then, $(\tilde{x}_k) \in \mathcal{E}(s/2)$. By (55) of Lemma 5,

$$\begin{aligned} P\big(|\tilde{s}_{ij,uv,t}| > \lambda\nu\big) &= P\big(H^{1/2}|\tilde{s}_{ij,uv,t}| > \lambda\nu H^{1/2}\big) \le f_H\big(\gamma_1, \gamma_2, c, \lambda\nu H^{1/2} \min(1, t^{1/2}H^{-1})\big) \quad (74)\\ P\big(|s_{iu,t}| > (\lambda\nu)^{1/2}\big) &= P\big(|H^{1/2}s_{iu,t}| > (\lambda\nu H)^{1/2}\big) \le f_H\big(\gamma_1, \gamma_2, c, (\lambda\nu H)^{1/2} \min(1, t^{1/2}H^{-1})\big) \end{aligned}$$

where $\gamma_1 = 2(\alpha/2)/(\alpha/2+2) = 2\alpha/(\alpha+4)$, $\gamma_2 = (\alpha/2)(s/2)/(\alpha/2+s/2+1) = \alpha s/(2\alpha+2s+4)$.

Function $f_H(\gamma_1, \gamma_2, c, \zeta)$ is non-increasing in ζ. So, in view of (73), the r.h.s. of (74) can be bounded by $f_H(\gamma_1, \gamma_2, c, \kappa'^{1/\gamma_1})$ where $\kappa' = \min(\kappa\nu, (\kappa\nu/c_*)^{1/2})$. Observe that

$$f_H(\gamma_1, \gamma_2, c, \kappa'^{1/\gamma_1}) \le c_0 \Big\{ \exp(-c_1\kappa'^{\gamma_1} \log p) + \exp\Big(-c_2\Big(\kappa'^{1/\gamma_1} \frac{H^{1/2}}{\log^2 H}\Big)^{\gamma_2}\Big\} = o(p^{-2})$$

because $c_1\kappa'^{\gamma_1} > 2$ when κ is sufficiently large and $\log p = o\big((H^{1/2}/\log^2 H)^{\gamma_2}\big)$ under assumption (16). This completes the proof of (72)(a,b) and (17).

(ii) Let $(x_{it}) \in \mathcal{H}(\theta)$. Then, $(\tilde{x}_k) \in \mathcal{H}(\theta/2)$ and by (56) of Lemma 5, for any $\theta' \in (2, \theta/2)$,

$$\begin{aligned} P\big(|\tilde{s}_{ij,uv,t}| > \lambda\nu\big) &\le g_H(\gamma_1, \theta', c, \lambda\epsilon'^{1/2} \min(1, t^{1/2}H^{-1})), \quad (75)\\ P\big(|s_{iu,t}| > (\lambda\nu)^{1/2}\big) &\le g_H(\gamma_1, \theta'^{1/2} \min(1, t^{1/2}H^{-1})). \end{aligned}$$

Since $g_H(\gamma_1, \gamma_2, c, \zeta)$ is a non-increasing function in ζ, by then by (73), the r.h.s. of (75) can be bounded by

$$g_H(\gamma_1, \theta', c, \kappa'^{1/\gamma_1}) \le c_0 \Big\{ \exp(-c_1\kappa'^{\gamma_1} \log p) + \frac{1}{(\kappa'^{1/\gamma_1})^{\theta'}} \frac{1}{H^{\theta'/2-1}} \Big\} = o(p^{-2})$$

by the same argument as in (71). This verifies (72)(a, b) and completes the proof of (17).

The claim (18) follows using the same argument as in the proof of Theorem 2.

(b) Bandwidth $H = t^{1/2}$ minimizes $\max(H^{-1/2}, (H/T)^{1/2})$, so $\lambda = \kappa(\log p)^\gamma \max(H^{-1/2}, (H/T)^{1/2}) \geq \lambda_{opt} = \kappa(\log p)^\gamma t^{-1/4}$. □

Proof of Theorem 1. To show (6), as in the proof of Theorem 2 it suffices to prove that for $\lambda = \kappa(T^{-1/2}\log p)^{1/2}$ the elements $\hat{\sigma}_{ij}$ of the sample variance matrix $\hat{\boldsymbol{\Sigma}} = (\hat{\sigma}_{ij})$ given by (4) uniformly in i, j satisfy

$$\max_{ij} P\big(|\hat{\sigma}_{ij} - \sigma_{ij}| > \lambda/2\big) = o(p^{-2}). \tag{76}$$

Set $z_k = y_{ik}y_{jk}$. Without restriction of generality assume that $Ey_{ik} = 0$, $Ey_{jk} = 0$. Then $\sigma_{ij} = Ez_t$ and

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

As in (67), to prove (76) it suffices to show that uniformly in i, j,

$$\text{a)} \quad P\big(|s_{T,ij}| > \lambda/4\big) = o(p^{-2}), \quad \text{b)} \quad P\big(|\bar{y}_i| > (\lambda/4)^{1/2}\big) = o(p^{-2}). \tag{77}$$

Assumption $(\boldsymbol{y}_t) \in \mathcal{M}$, implies $(z_k) \in \mathcal{M}$, and similarly as in (68),

$$\lambda T^{1/2} \geq \kappa(\log p)^{1/2}, \quad (\lambda T)^{1/2} = \lambda^{-1/2}(\lambda T^{1/2}) \geq (\kappa/l_*)^{1/2}(\log p)^{1/2}. \tag{78}$$

(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, from (28) of Lemma 2 we obtain that with $\gamma = (s/2)(1 + s/2)$,

$$\begin{aligned} P\big(|s_{T,ij}| > \lambda/4\big) &= P\big(T^{1/2}|s_{T,ij}| > \lambda T^{1/2}/4\big) \leq f_T(2, \gamma, c, \lambda T^{1/2}/4), \\ P\big(|\bar{y}_i| > (\lambda/4)^{1/2}\big) &= P\big(|T^{1/2}\bar{y}_i| > (\lambda T/4)^{1/2}\big) \leq f_T(2, \gamma, c, (\lambda T/4)^{1/2}) \end{aligned}$$

which together with (78) implies (77)(a,b) by the same argument as in the proof of (69).

(ii) Let $(y_{ik}) \in \mathcal{H}(\theta)$. Then, $(z_k) \in \mathcal{H}(\theta/2)$, and from (29) of Lemma 2 it follows that for any $\theta' \in (2, \theta/2)$,

$$\begin{aligned} P\big(|s_{T,ij}| > \lambda/4\big) &= P\big(T^{1/2}|s_{T,ij}| > \lambda T^{1/2}/4\big) \leq g_T(2, \theta'^{1/2}/4), \\ P\big(|\bar{y}_{it}| > (\lambda/4)^{1/2}\big) &= P\big(|T^{1/2}\bar{y}_i| > (\lambda T/4)^{1/2}\big) \leq g_T(2, \theta'^{1/2}). \end{aligned} \tag{79}$$

By the same argument as in the proof of (70) it follows that (79) and (78) imply (77)(a,b). This completes the proof of (6).

Property (7) follows using the same argument as in the proof of (17) in Theorem 2. □

4 Appendix. Auxiliary results

This section contains auxiliary results used in the proofs.

Lemma 6 *Let $x \in \mathcal{E}(\alpha)$, $(x_t) \in \mathcal{E}(\alpha)$, $\alpha > 0$.*

(i) If $y \in \mathcal{E}(\alpha')$, $\alpha' > 0$ then $xy \in \mathcal{E}(\tilde\alpha)$, $\tilde\alpha = \alpha\alpha'/(\alpha+\alpha')$.
Moreover, $x + y \in \mathcal{E}(\min(\alpha,\alpha'))$ and $|z| \le |x|$ implies $z \in \mathcal{E}(\alpha)$.

(ii) If $y \in \mathcal{H}(\theta)$, $\theta > 0$ then $xy \in \mathcal{H}\big(\theta'\big)$ for any $0 < \theta' < \theta$.
(iii) If $P(|y_t| \ge \zeta) \le f_t(2,\gamma,c,\zeta)$, $\zeta > 0$, $t \ge 2$ with $\gamma > 0$, then

$$P(|x_t y_t| \ge \zeta) \le f_t(\gamma_1,\gamma_2,c',\zeta), \quad \zeta > 0, \quad t \ge 2, \tag{80}$$

where $\gamma_1 = 2\alpha/(2+\alpha)$, $\gamma_2 = \alpha\gamma/(\alpha+\gamma)$ and c' does not depend on t, ζ.
(iv) If $P(|y_t| \ge \zeta) \le g_t(2,\theta,c,\zeta)$, $\zeta > 0$, $t \ge 2$ where $\theta > 2$, then for $2 < \theta' < \theta$,

$$P(|x_t y_t| \ge \zeta) \le g_t(\gamma_1,\theta',c',\zeta), \quad \zeta > 0, \quad t \ge 2, \tag{81}$$

where $\gamma_1 = 2\alpha/(2+\alpha)$ and c' does not depend on t, ζ.

Proof. The function $f(x) = x^\alpha + c(v/x)^{\alpha'}$, $x > 0$ where $v > 0$, $c > 0$ achieves its unique minimum at $x_0 = (c\alpha'^{1/(\alpha+\alpha')}v^{\alpha'/(\alpha+\alpha')}$ because x_0 is a unique solution of equation $f'^{\alpha-1} - c\alpha'^{\alpha'}x^{-1} = 0$ and $f''(x_0) = x_0^{\alpha-2}\alpha(\alpha+\alpha') > 0$. Thus,

$$f(x) \ge f(x_0) = c'^{\tilde\alpha}, \quad x \ge 0 \tag{82}$$

where $\tilde\alpha = \alpha\alpha'/(\alpha+\alpha')$ and $c' = (c\alpha'^{\alpha/(\alpha+\alpha')}(1+\alpha/\alpha')$.

Let $p, q > 1$, $1/p + 1/q = 1$. Then

$$\begin{aligned} P(|xy| \ge \zeta) \quad &= \textstyle\sum_{k=0}^\infty P(\{|x| \in [k,k+1)\} \cap \{|xy| \ge \zeta\}) \\ &\le \textstyle\sum_{k=0}^\infty P^{1/p}(|x| \in [k,k+1))P^{1/q}(|y| \ge \zeta/(k+1)). \end{aligned}$$

Since $x \in \mathcal{E}(\alpha)$, then for $k \ge 0$, $P(|x| \in [k,k+1)) \le P(|x| \ge k) \le c_0' \exp(-2c_1'^\alpha)$, $k \ge 0$ for some $c_0' > 0$, $c_1' > 0$. Denote $g_{k\zeta} := \exp(-c_1'^\alpha)P^{1/q}(|y| \ge \zeta/k)$. Then,

$$\begin{aligned} P(|xy| \ge \zeta) \quad &\le C \textstyle\sum_{k=0}^\infty \exp(-c_1'k^\alpha)g_{k+1,\zeta} \\ &\le C \max_{k\ge1} g_{k\zeta} \textstyle\sum_{k=0}^\infty \exp(-c_1'k^\alpha) \le C \max_{k\ge1} g_{k\zeta}. \end{aligned} \tag{83}$$

Next we evaluate $\max_{k\ge1} g_{k\zeta}$ in the cases (i)-(iv).

(i) Let $y \in \mathcal{E}(\alpha')$. Then $P(|y| \geq \zeta/k) \leq c_0 \exp(-c_1(\zeta/k)^{\alpha'})$ for $\zeta > 0$. Therefore, by (82),

$$\max_{k\geq 1} g_{k,\zeta} \leq C \max_{k\geq 1} \exp\big(- c_1'^{\alpha} + (\zeta/k)^{\alpha'})\big) \leq C \exp(-c_2''\zeta^{\tilde{\alpha}}), \quad k \geq 1,$$

Hence, $P(|xy| \geq \zeta) \leq C \exp(-c_2''\zeta^{\tilde{\alpha}})$, i.e. $xy \in \mathcal{E}(\tilde{\alpha})$.

Property $x+y \in \mathcal{E}(\min(\alpha, \alpha'))$ follows from $P(|x+y| \geq \zeta) \leq P(|x| \geq \zeta/2)+P(|y| \geq \zeta/2)$. If $z \leq x$ then $P(|z| \geq \zeta) \leq P(|x| \geq \zeta)$ implies $z \in \mathcal{E}(\alpha)$.

(ii) Let $y \in \mathcal{H}(\theta)$ and $0 < \theta' < \theta$. Set $q = \theta/\theta' > 1$. Then $\theta/q = \theta'$, and $P^{1/q}(|y| \geq \zeta/k) \leq C(\zeta/k)^{-\theta/q} = C(\zeta/k)^{-\theta'}$. Hence,

$$\max_{k\geq 1} g_{k\zeta} \leq C\zeta^{-\theta'} \max_{k\geq 1} \exp(-c_1'^{\alpha})k^{\theta'} \leq C\zeta^{-\theta'}.$$

By (83), this implies $P(|xy| \geq \zeta) \leq C\zeta^{-\theta'}$, i.e. $xy \in \mathcal{H}(\theta')$.

(iii) Without restriction of generality, we assume that $\zeta \geq 1$. By (83),

$$P(|x_t y_t| \geq \zeta) \leq C \max_{k\geq 1} g_{k\zeta}. \tag{84}$$

To evaluate $g_{k\zeta} = \exp(-c_1'^{\alpha}) f_t^{1/q}(2, \gamma, c, \zeta/k)$, denote $\zeta_t = \zeta\sqrt{t}/\log^2 t$. Using inequality

$$(a + b)^{1/q} \leq a^{1/q} + b^{1/q}, \qquad a,\, b > 0 \tag{85}$$

we obtain

$$\begin{aligned} f_t^{1/q}(2, \gamma, c, \zeta/k) &\leq C\Big(\exp(-c_1(\zeta/k)^2) + \exp(-c_2(\zeta_t/k)^{\gamma})\Big)^{1/q} \\ &\leq C\Big(\exp(-(c_1/q)(\zeta/k)^2) + \exp(-(c_2/q)(\zeta_t/k)^{\gamma}\Big). \end{aligned}$$

Hence, $g_{k,\zeta} \leq C\{\exp(-c_1''(k^{\alpha} + (\zeta/k)^2)) + \exp(-c_2''(k^{\alpha} + (\zeta_t/k)^{\gamma})\big)$, and by (82),

$$g_{k,\zeta} \leq c_0^*\big(\exp(-c_1^*\zeta^{\gamma_1}) + \exp(-c_2^*\zeta_t^{\gamma_2})\big) = f_t(\gamma_1, \gamma_2, c^*, \zeta), \quad k \geq 1,$$

with $\gamma_1 = 2\alpha/(2 + \alpha)$, $\gamma_2 = \alpha\gamma/(\gamma + \alpha)$. Thus, (84) implies $P(|x_t y_t| \geq \zeta) \leq f_t(\gamma_1, \gamma_2, c'', \zeta)$ which proves (iii).

(iv) Let $\zeta \geq 1$. In (iv), (84) holds with $g_{k\zeta} = \exp(-c_1'^{\alpha}) g_t^{1/q}(2, \theta, c, \zeta/k)$. Let $2 < \theta' < \theta$. Select $q > 1$ such that $\theta/\theta' > q$ and $(\theta - 2)/(\theta' - 2) > q$. By (85),

$$\begin{aligned} g_t^{1/q}(2, \theta', c, \zeta/k) &\leq C\Big(\exp\{-c_1(\zeta/k)^2\} + (\zeta/k)^{-\theta} t^{-(\theta/2-1)}\Big)^{1/q} \\ &\leq C\{\exp\{-(c_1/q)(\zeta/k)^2\} + \zeta^{-\theta/q} t^{-(\theta/2-1)/q} k^{\theta/q}\}. \end{aligned}$$

Selection of $q > 1$ implies $\theta/q > \theta'$ and $(\theta/2 - 1)/q > \theta'/2 - 1$. Since $\zeta \geq 1$ and $t \geq 1$, then

$$g_t^{1/q}(2, \theta', c, \zeta/k) \leq C\{\exp\{-(c_1/q)(\zeta/k)^2\} + \zeta^{-\theta'} t^{-(\theta'/2-1)} k^{\theta/q}\}.$$

Hence,

$$\max_{k\ge1} g_{k\zeta} \le C \max_{k\ge1} \exp\{-c_1''(k^\alpha + (\zeta/k)^2)\} + C\zeta^{-\theta'} t^{-(\theta'/2-1)} \max_{k\ge1}\{\exp(-c_2' k^\alpha) k^{\theta/q}\}$$
$$\le c_0^*\Big(\exp(-c_1^*\zeta^{\gamma_1}) + \zeta^{-\theta'} t^{-(\theta'/2-1)}\Big) = g_t(\gamma_1, \theta'^*, \zeta)$$

with $\gamma_1 = 2\alpha/(2+\alpha)$. Then (84) implies $P(|x_t y_t| \ge \zeta) \le g_t(\gamma_1, \theta'^*, \zeta)$ which proves (iv). □

Lemma 7 *Let* $0 < \gamma \le 1$.
(i) Let ξ *be a zero random variable. Then for all* $\zeta > 0$,

$$E\big(|\xi|^\gamma I(|\xi| > \zeta)\big) \le \begin{cases} c_0 \exp(-c_1\zeta^s) & \text{if } \xi \in \mathcal{E}(s),\ s > 0 \quad (86) \\ c_0 \zeta^{\gamma-\theta} & \text{if } \xi \in \mathcal{H}(\theta),\ \gamma < \theta \quad (87) \end{cases}$$

for some $c_0 > 0$, $c_1 > 0$ *which do not depend on* ζ.
(ii) Let s_t *be a sequence of zero mean random variables such that* $P(|s_t| \ge \zeta) \le f_t(\gamma_1, \gamma_2, c', \zeta)$ *for all* $\zeta > 0$, $t \ge 2$ *for some* $\gamma_1 > 0$, $\gamma_2 > 0$ *and* c. *Then,*

$$E\big[|s_t|^\gamma I(|s_t| > \zeta)\big] \le f_t(\gamma_1, \gamma_2, c', \zeta), \quad \zeta > 0, \quad t \ge 2, \tag{88}$$

where c' *does not depend on* ζ, t.
(iii) Let $Es_t = 0$ *and* $P(|s_t| \ge \zeta) \le g_t(\gamma_1, \theta, c, \zeta)$ *for all* $\zeta > 0$, $t \ge 2$ *for some* $\theta > \gamma$, $\gamma_1 > 0$ *and* c. *Then,*

$$E\big[|s_t|^\gamma I(|s_t| > \zeta)\big] \le (\max(\zeta, 1))^\gamma g_t(\gamma_1, \theta, c', \zeta), \quad \zeta > 0, \quad t \ge 2, \tag{89}$$

where c' *does not depend on* ζ, t.

Proof. Without restriction generality let $\zeta \ge 1$. Denote $F(x) = P(|\xi| \ge x)$. Then

$$E[|\xi|^\gamma I(|\xi| > \zeta)] = -\int_\zeta^\infty x^\gamma dF(x) = -\zeta^\gamma F(\zeta) + \int_\zeta^\infty x^{\gamma-1} F(x)dx. \tag{90}$$

(i) If $(\xi_k) \in \mathcal{E}(s)$, then $F(x) \le c_0' \exp\big(-2c_1'^s\big)$ for some $c_0', c_1' > 0$, and (90) implies (86):

$$E[|\xi|^\gamma I(|\xi| > \zeta)] \le F^{1/2}(\zeta)\big(\zeta^\gamma F^{1/2}(\zeta) + \int_1^\infty x^{\gamma-1} F^{1/2}(x)dx\big) \le C\exp(-c_1'\zeta^s).$$

If $(\xi) \in \mathcal{H}(\theta)$, then $F(x) \le c_0'^{-\theta}$ and (90) implies (87):

$$E[|\xi|^\gamma I(|\xi| > \zeta)] \le C\big(\zeta^\gamma |\zeta|^{-\theta} + \int_\zeta^\infty x^{\gamma-1} x^{-\theta} dx\big) \le C\zeta^{\gamma-\theta}.$$

(ii) Let again $\zeta \ge 1$. Definition (25) of f_t implies that $f_t(\gamma_1, \gamma_2, c, \zeta) \le c_0 \exp(-2c_1\zeta^{\min(\gamma_1,\gamma_2)})$, $\zeta > 0$, $t \ge 2$ for some $c_0, c_1 > 0$. In addition, $f_t(\gamma_1, \gamma_2, c, x) \le f_t(\gamma_1, \gamma_2, c, \zeta)$ for $x \ge \zeta$. Thus, by (90), for $\zeta \ge 1$,

$$E[|s_t|^\gamma I(|s_t| > \zeta)] \le f_t^{1/2}(\gamma_1, \gamma_2, c, \zeta)\big(\zeta^\gamma f_t^{1/2}(\gamma_1, \gamma_2, c, \zeta) + \int_1^\infty x^{\gamma-1} f_t^{1/2}(\gamma_1, \gamma_2, c, x)dx\big)$$
$$\le C f_t^{1/2}(\gamma_1, \gamma_2, c, \zeta) \le C f_t(\gamma_1, \gamma_2, c', \zeta)$$

for some c' in view of (85). This proves (88).

(iii) Let $\zeta \ge 1$. Since $P(|s_t| \ge \zeta) \le g_t(\gamma_1, \theta, c, \zeta)$, (90) implies

$$E[|s_t|^\gamma I(|s_t| > \zeta)] \le \zeta^\gamma g_t(\gamma_1, \theta, c, \zeta) + \int_\zeta^\infty x^{\gamma-1} g_t(\gamma_1, \theta, c, x) dx. \tag{91}$$

By (25), $g_t(\gamma_1, \theta, c, \zeta) \le c_0\{\exp(-2c_1\zeta^{\gamma_1}) + \zeta^{-\theta} t^{-(\theta/2-1)}\}$ for some $c_0, c_1 > 0$. Thus,

$$\int_\zeta^\infty x^{\gamma-1} g_t(\gamma_1, \theta, c, x) dx \le C\Big(\exp(-c_1\zeta^{\gamma_1}) \int_\zeta^\infty x^{\gamma-1} \exp(-c_1 x^{\gamma_1}) dx + \int_\zeta^\infty x^{\gamma-1} x^{-\theta} t^{-(\theta/2-1)} dx\Big)$$
$$\le C\zeta^\gamma \Big(\exp(-c_1\zeta^{\gamma_1}) + \zeta^{-\theta} t^{-(\theta/2-1)}\Big) = \zeta^\gamma g_t(\gamma_1, \theta, c', \zeta)$$

for some c'. This together with (91) proves (89). □

Lemma 8 *Let $b_{H,k}$ satisfy (41) with $\nu > 2$ and let $0 < \theta < 1$. Then for all $1 \le H \le \theta t \le T$, $T \ge 2$,*

$$H^{-1} \sum_{k=1}^T b_{H,|t-k|} \Big(\frac{|t-k|}{\min(t,k)}\Big)^\gamma \le \begin{cases} C(H/t)^\gamma & \text{if } 0 < \gamma < 1 \\ C\big[(H/t) + (H/t)^{\nu-1} \log t\big] & \text{if } \gamma = 1, \end{cases} \tag{92}$$

$$H^{-1} \sum_{k=1}^T b_{H,|t-k|} \Big(\frac{\max(|t-k|, H)}{k}\Big)^{1/2} \le C(H/t)^{1/2}, \tag{93}$$

where $C > 0$ does not depend on t, T, H.

Proof. Denote by $q_{\gamma,H}$ the l.h.s. of (92). By (41), $b_{H,|t-k|} \le C(1 + (|t-k|/H)^\nu)^{-1}$ where $\nu > 2$. Write $q_{\gamma,H} = H^{-1} \sum_{k=t/2+1}^T [...] + H^{-1} \sum_{k=1}^{t/2} [...] = q_{\gamma,H;1} + q_{\gamma,H;1}$. Then, for $0 < \gamma \le 1$,

$$q_{\gamma,H;1} \le C(H/t)^\gamma H^{-1} \sum_{k=t/2+1}^T b_{H,|t-k|} \, |(t-k)/H|^\gamma$$
$$\le C(H/t)^\gamma \big[H^{-1} \sum_{k=1}^T b_{H,|t-k|} \, |(t-k)/H|^\gamma\big] \le C(H/t)^\gamma. \tag{94}$$

On the other hand, for $k < t/2$ we have $b_{T,|t-k|} \le C(H/t)^\nu$. Thus,

$$q_{\gamma,H;2} \le C(H/t)^\nu H^{-1} t^\gamma \sum_{k=1}^{t/2} k^{-\gamma}.$$

Let $0 < \gamma < 1$. Then $\sum_{k=1}^{t/2} k^{-\gamma} \le Ct^{1-\gamma}$ which implies $q_{\gamma,H;2} \le C(H/t)^{\nu-1} \le C(H/t)$ since $\nu > 2$ and $H/t \le 1$. Together with (94) this proves (92).

Let $\gamma = 1$. Then $\sum_{k=1}^{t/2} k^{-1} \le C \log t$ and $q_{\gamma,H;2} \le C(H/t)^{\nu-1} \log t$. This and (94) implies (92).

To prove (93), denote by $\tilde{q}_{\gamma,H}$ the l.h.s. of (93). Write $\tilde{q}_{\gamma,H} = H^{-1} \sum_{k=1:\,|t-k| \le H}^T [...] + H^{-1} \sum_{k=1:\,|t-k|>H}^T [...] =: \tilde{q}_{\gamma,H;1} + \tilde{q}_{\gamma,H;2}$. First we bound $\tilde{q}_{\gamma,H;1}$. By assumption, $H \le \theta t$. Therefore, $|k-t| \le H$ implies $k \ge t - H \ge (1-\alpha)t$, and $(H/k)^{1/2} \le C(H/t)^{1/2}$. Thus,

$$\tilde{q}_{\gamma,H;1} \le C(H/t)^{1/2} [H^{-1} \sum_{k=1:\,|t-k| \le H}^T b_{T,|t-k|}] \le C(H/t)^{1/2}.$$

On the other hand,

$$\tilde{q}_{\gamma,H;2} \le H^{-1} \sum_{k=1}^T b_{T,|t-k|} \big(\tfrac{|t-k|}{\min(t,k)}\big)^{1/2} \le C(H/t)^{1/2}$$

by (92). These bounds imply (93). □

References

Abadir, K. M., W. Distaso, and F. Zikes (2014): "Design-free estimation of variance matrices," *Journal of Econometrics*, 181, 165–180.

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.

Davidson, J. (1994): *Stochastic Limit Theory*. Oxford University Press.

Davydov, Y. A. (1968): "Convergence of Distributions Generated by Stationary Stochastic Processes," *Theory of Probability & Its Applications*, 13, 691–696.

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.

Giraitis, L., G. Kapetanios, and T. Yates (2014): "Inference on multivariate heteroscedastic time varying random coefficient models," *Mimeo, Queen Mary University of London*.

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 (2004): "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, 88, 365–411.

Merlevéde, F., M. Peligrad, and E. Rio (2009): "Bernstein inequality and moderate deviations under strong mixing conditions," *Institute of Mathematical Statistics Collections*, 5, 273–292.

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.

Wooldridge, J. M., and H. White (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.