High-dimensional estimation of quadratic variation based on penalized realized variance

Kim Christensen[1] · Mikkel Slot Nielsen[2] · Mark Podolskij[3]

Abstract

In this paper, we develop a penalized realized variance (PRV) estimator of the quadratic variation (QV) of a high-dimensional continuous Itô semimartingale. We adapt the principle idea of regularization from linear regression to covariance estimation in a continuous-time high-frequency setting. We show that under a nuclear norm penalization, the PRV is computed by soft-thresholding the eigenvalues of realized variance (RV). It therefore encourages sparsity of singular values or, equivalently, low rank of the solution. We prove our estimator is mini-max optimal up to a logarithmic factor. We derive a concentration inequality, which reveals that the rank of PRV is—with a high probability—the number of non-negligible eigenvalues of the QV. Moreover, we also provide the associated non-asymptotic analysis for the spot variance. We suggest an intuitive data-driven subsampling procedure to select the shrinkage parameter. Our theory is supplemented by a simulation study and an empirical application. The PRV detects about three–five factors in the equity market, with a notable rank decrease during times of distress in financial markets. This is consistent with most standard asset pricing models, where a limited amount of systematic factors driving the cross-section of stock returns are perturbed by idiosyncratic errors, rendering the QV—and also RV—of full rank.

Keywords Bernstein's inequality · LASSO estimation · Low rank estimation · Quadratic variation · Rank recovery · Realized variance · Shrinkage estimator

✉ Mark Podolskij
mark.podolskij@uni.lu

Kim Christensen
kim@econ.au.dk

Mikkel Slot Nielsen
mikkel.slot.nielsen@gmail.com

[1] Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark

[2] Department of Mathematics, Aarhus University, Aarhus, Denmark

[3] Department of Mathematics and Department of Finance, University of Luxembourg, Esch-sur-Alzette, Luxembourg

1 Introduction

The covariance matrix of asset returns is a central component, which is required in several contexts in financial economics, such as portfolio composition, pricing of financial instruments, or risk management (e.g., Andersen et al. (2003)). In the past few decades, estimation of quadratic variation (QV) from high-frequency data has been intensively studied in the field of econometrics. The standard estimator of QV, Σ, say over a window $[0, 1]$, is the realized variance (RV) (e.g., Andersen and Bollerslev (1998); Barndorff-Nielsen and Shephard (2002, 2004); Jacod (1994)). Given a decreasing sequence $(\Delta_n)_{n \geq 1}$ with $\Delta_n \to 0$ and equidistant observations $(Y_{i \Delta_n})_{i=0}^{\lfloor \Delta_n^{-1} \rfloor}$ of a d-dimensional semimartingale $(Y_t)_{t \in [0,1]}$, the RV is defined as

$$\widehat{\Sigma}_n = \sum_{k=1}^{\lfloor \Delta_n^{-1} \rfloor} (\Delta_k^n Y)(\Delta_k^n Y)^\top, \tag{1.1}$$

where $\Delta_k^n Y = Y_{k \Delta_n} - Y_{(k-1) \Delta_n}$ is the increment, and $^\top$ denotes the transpose operator. The asymptotic properties of $\widehat{\Sigma}_n$ are generally known, when the dimension d is fixed (e.g., Barndorff-Nielsen et al. (2006); Diop et al. (2013); Heiny and Podolskij (2020); Jacod (1994, 2008)). In particular, for any semimartingale $\widehat{\Sigma}_n$ is by definition a consistent estimator of Σ.

When modeling the dynamics of a large financial market over a relatively short time horizon, one is often faced with an ill-posed inference problem, where the number of assets d is comparable to (or even exceeding) the number of high-frequency data $\lfloor \Delta_n^{-1} \rfloor$ available to compute $\widehat{\Sigma}_n$. This renders RV inherently inaccurate, and it also implies that d cannot be treated as fixed in the asymptotic analysis. To recover an estimate of Σ in such a high-dimensional setting, the covariance matrix has to display a more parsimonious structure.

Currently, the literature on high-dimensional high-frequency data, mostly based on a continuous semimartingale model for the security price process, is rather scarce. Wang and Zou (2010) investigate optimal convergence rates for estimation of QV under sparsity constraints on its entries. Zheng and Li (2011) apply random matrix theory to estimate the spectral eigenvalue distribution of QV in a one-factor diffusion setting. Related work based on random matrix theory and eigenvalue cleaning, or on approximating the high-dimensional system by a low-dimensional factor model, can be found in, e.g., Cai et al. (2020); Hautsch et al. (2012); Lunde et al. (2016). Aït-Sahalia and Xiu (2017) study a high-dimensional factor model, where the eigenvalues of QV are assumed to be dominated by common components (spiked eigenvalue setting), and estimate the number of common factors under sparsity constraints on the idiosyncratic components.

In the context of joint modeling of the cross-section of asset returns, if a factor model is adopted, one should expect a small number of eigenvalues of Σ to be dominant (see also Sect. 4). In other settings, such as when the market is complete and some entries of Y_t reflect prices on derivative contracts (spanned by the existing assets and therefore redundant), one should even expect the rank of the volatility to be strictly smaller than the dimension. In such an instance, the rank of the covariance matrix is smaller than d, meaning that the intrinsic dimension of the system can be represented in terms of a driving Brownian motion of lower dimension. In general, identifying a low rank can, for instance, help with the economic interpretation of the model, or it may lighten the computational load related to its estimation or simulation. We refer to several recent studies on identification and estimation of the intrinsic dimension of the driving Brownian motion and eigenvalue analysis of QV in Aït-Sahalia and Xiu (2017, 2019); Fissler and Podolskij (2017); Jacod et al. (2008); Jacod

and Podolskij ([2013](), [2018]()), see also Kong ([2017](), [2020]()) in the high-dimensional setting. A related contribution of Pelger ([2019]()) incorporates a finite-activity jump process in the model.

In this paper, we develop a regularized estimator of Σ called the penalized realized variance (PRV). We draw from the literature of shrinkage estimation in linear regression (e.g., Hoerl and Kennard ([1970]()); Tibshirani ([1996]())). In particular, we propose the following LASSO-type estimator of Σ based on the RV in ([1.1]()):

$$\widehat{\Sigma}_n^\lambda = \arg\min_{A \in \mathbb{R}^{d \times d}} \left(\|\widehat{\Sigma}_n - A\|_2^2 + \lambda \|A\|_1 \right). \tag{1.2}$$

Here, $\lambda \geq 0$ is a tuning parameter that controls the degree of penalization, while $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the nuclear and Frobenius matrix norm. It has been shown in various contexts that estimators based on nuclear norm regularization possess good statistical properties, such as having optimal rates of convergence and being of low rank. For instance, Koltchinskii et al. ([2011]()) study such properties in the trace regression model and, in particular, in the matrix completion problem, where one attempts to fill out missing entries of a partially observed matrix. Negahban and Wainwright ([2011]()) adopt a rather general observation model and, among other things, cover estimation in near low rank multivariate regression models and vector autoregressive processes. Further references in this direction include Argyriou et al. ([2008]()); Bach ([2008]()); Candès and Recht ([2009]()); Recht et al. ([2010]()). While previous papers on nuclear norm penalization are solely dedicated to discrete-time models, the current work is—to the best of our knowledge—the first to study this problem in a continuous-time Itô setting. This implies a number of subtle technical challenges, since the associated high-frequency data are dependent and potentially non-stationary.

We provide a complete non-asymptotic theory for the PRV estimator in ([1.2]()) by deriving bounds for the Frobenius norm of its error. We show that it is minimax optimal up to a logarithmic factor. The derivation relies on advanced results from stochastic analysis combined with a Bernstein-type inequalities presented in Theorems [2.3]() and [2.4](), which constitute some of the major theoretical contributions of the paper. The latter builds upon recent literature on concentration inequalities for matrix-valued random variables, but it is more general as the variables are allowed to be both dependent and unbounded. We discuss the necessary conditions on d and Δ_n to ensure consistency results. We further show that in a "local-to-deficient" rank setting, where some eigenvalues are possibly decreasing as the dimension of the system increases, the penalized estimator in ([1.2]()) identifies the number of non-negligible eigenvalues of Σ with high probability.

The estimator depends on a tuning parameter λ. We exploit the subsampling approach of Christensen et al. ([2017]()) to choose a data-driven amount of shrinkage in the implementation. We also provide a related theoretical analysis for the local volatility, which is more likely to exhibit a lower rank than QV (the latter is only expected to have a near low rank in many settings).

In the empirical analysis, we look at the 30 members of the Dow Jones Industrial Average index to demonstrate that our results are consistent with a standard Fama–French three-factor structure, and one may expect a lower number of factors during crisis.

The paper proceeds as follows. In Sect. [2](), we establish concentration inequalities for the estimation error $\|\widehat{\Sigma}_n^\lambda - \Sigma\|_2$ and show minimax optimality up to a logarithmic factor. In Sect. [2.3](), we present sufficient conditions to ensure that the rank of $\widehat{\Sigma}_n^\lambda$ coincides with the number of non-negligible eigenvalues of Σ with high probability. In Sect. [2.4](), we show how the tuning parameter λ can be selected with a data-driven approach. In Sect. [3](), the associated non-asymptotic theory for estimation of the instantaneous variance is developed. In Sect. [4](), we design a simulation study to demonstrate the ability of our estimator to identify a low-dimensional factor model through rank($\widehat{\Sigma}_n^\lambda$). In Sect. [5](), we implement the estimator

on empirical high-frequency data from the 30 members of the Dow Jones Industrial Average index. The proofs are included in the appendix.

Notation

This paragraph introduces some notation used throughout the paper. For a vector or a matrix A the transpose of A is denoted by $A^\top$. We heavily employ the fact that any real matrix $A \in \mathbb{R}^{m \times q}$ admits a singular value decomposition (SVD) of the form:

$$A = \sum_{k=1}^{m \wedge q} s_k u_k v_k^\top \tag{1.3}$$

where $s_1 \geq \cdots \geq s_{m \wedge q}$ are the singular values of A, and $\{u_1, \ldots, u_{m \wedge q}\}$ and $\{v_1, \ldots, v_{m \wedge q}\}$ are orthonormal vectors. If $m = q$ and A is symmetric and positive semidefinite, its singular values coincide with its eigenvalues and the SVD is the orthogonal decomposition (meaning that one can take $v_k = u_k$). Sometimes we write $s_k(A)$, $v_k(A)$ and $u_k(A)$ to explicitly indicate the dependence on the matrix A. The rank of A is denoted by $\mathrm{rank}(A)$, whereas

$$\mathrm{rank}(A; \varepsilon) = \sum_{k=1}^{m \wedge q} \mathbb{1}_{[\varepsilon, \infty)}(s_k) \tag{1.4}$$

is the number of singular values of A exceeding a certain threshold $\varepsilon \in [0, \infty)$. Note that $\varepsilon \mapsto \mathrm{rank}(A; \varepsilon)$ is non-increasing on $[0, \infty)$, and càdlàg and piecewise constant on $(0, \infty)$. Furthermore, $\mathrm{rank}(A; 0) = m \wedge q$ and $\lim_{\varepsilon \downarrow 0} \mathrm{rank}(A; \varepsilon) = \mathrm{rank}(A)$. We denote by $\|A\|_p$ the *Schatten p-norm* of $A \in \mathbb{R}^{m \times q}$, i.e.,

$$\|A\|_p = \left(\sum_{k=1}^{m \wedge q} s_k^p \right)^{1/p} \tag{1.5}$$

for $p \in (0, \infty)$. Moreover, $\|A\|_\infty$ denotes the maximal singular value of A, i.e., $\|A\|_\infty = s_1$. In particular, in this paper we work with $p = 1$, $p = 2$, and $p = \infty$ corresponding to the nuclear, Frobenius, and spectral norm. The Frobenius norm is also induced by the inner product $\langle A, B \rangle = \mathrm{tr}(A^\top B)$ and we have the trace duality

$$\langle A, B \rangle \leq \|A\|_1 \|B\|_\infty. \tag{1.6}$$

For a linear subspace $S \subseteq \mathbb{R}^m$, $S^\perp$ denotes the linear space orthogonal to S and P_S stands for the projection onto S.

Given two sequences $(a_n)_{n \geq 1}$ and $(b_n)_{n \geq 1}$ in $(0, \infty)$, we write $a_n = o(b_n)$ if $\lim_{n \to \infty} a_n / b_n = 0$ and $a_n = O(b_n)$ if $\limsup_{n \to \infty} a_n / b_n$ is bounded. Furthermore, if both $a_n = O(b_n)$ and $b_n = O(a_n)$, we write $a_n \asymp b_n$. Finally, to avoid trivial cases we assume throughout the paper that the dimension d of the process $(Y_t)_{t \in [0,1]}$ is at least two.

2 Theoretical framework

We suppose a d-dimensional stochastic process $Y_t = (Y_{1,t}, \ldots, Y_{d,t})^\top$ is defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,1]}, \mathbb{P})$. In our context, Y_t is a time t random vector of log-prices of financial assets, which are traded in an arbitrage-free, frictionless market and therefore semimartingale (e.g., Delbaen and Schachermayer (1994)). In particular, we assume

$(Y_t)_{t \in [0,1]}$ is a continuous Itô semimartingale, which can be represented by the stochastic differential equation:

$$dY_t = \mu_t dt + \sigma_t dW_t, \qquad t \in [0, 1], \tag{2.1}$$

where $(W_t)_{t \in [0,1]}$ is an adapted d-dimensional standard Brownian motion, $(\mu_t)_{t \in [0,1]}$ is a progressively measurable and locally bounded drift with values in $\mathbb{R}^d$, and $(\sigma_t)_{t \in [0,1]}$ is a predictable and locally bounded stochastic volatility with values in $\mathbb{R}^{d \times d}$. In the above setting the QV (or integrated variance) is $\Sigma = \int_0^1 c_t dt$, where $c_t \equiv \sigma_t \sigma_t^\top$. We remark that the length $T = 1$ of the estimation window $[0, T]$ is fixed for expositional convenience and is completely without loss of generality, since a general T can always be reduced to the unit interval via a suitable time change.

Note that we exclude a jump component from (2.1). It should be possible to extend our analysis and allow for at least finite-activity jump processes. To do so, we can replace the RV with its truncated version, following the ideas of Mancini (2009). We leave the detailed analysis of more general jump processes for future work. Apart from these restrictions, our model is essentially nonparametric, as it allows for almost arbitrary forms of random drift, volatility, and leverage.

2.1 Upper bounds for the PRV

The goal of this section is to derive a sharp upper bound on the estimation error $\|\widehat{\Sigma}_n^\lambda - \Sigma\|_2$ that apply with high probability, where the PRV estimator $\widehat{\Sigma}_n^\lambda$ has been defined in (1.2). We first show that the PRV can alternatively be found by soft-thresholding of the eigenvalues in the orthogonal decomposition of $\widehat{\Sigma}_n$.

Proposition 2.1 *Let $\widehat{\Sigma}_n = \sum_{k=1}^d s_k u_k u_k^\top$ be the orthogonal decomposition of $\widehat{\Sigma}_n$. Then, the unique solution $\widehat{\Sigma}_n^\lambda$ to (1.2) is given by*

$$\widehat{\Sigma}_n^\lambda = \sum_{k=1}^d \max\left\{s_k - \frac{\lambda}{2}, 0\right\} u_k u_k^\top. \tag{2.2}$$

Proposition 2.1 is standard (see Koltchinskii et al. (2011)), but we state it explicitly for completeness. The interpretation of the PRV representation in (2.2) is that all "significant" eigenvalues of $\widehat{\Sigma}_n$ are shrunk by $\lambda/2$, while the smallest ones are set to 0. Hence, we only retain the principal eigenvalues in the orthogonal decomposition of $\widehat{\Sigma}_n$. The PRV can therefore account for a (near) low rank constraint. In the next proposition, we present a general non-probabilistic oracle inequality for the performance of $\widehat{\Sigma}_n^\lambda$.

Proposition 2.2 *Assume that $2\|\widehat{\Sigma}_n - \Sigma\|_\infty \leq \lambda$. Then, it holds that*

$$\|\widehat{\Sigma}_n^\lambda - \Sigma\|_2^2 \leq \inf_{A \in \mathbb{R}^{d \times d}} \left(\|A - \Sigma\|_2^2 + \min\left\{2\lambda\|A\|_1, 3\lambda^2 \operatorname{rank}(A)\right\}\right). \tag{2.3}$$

In particular, $\|\widehat{\Sigma}_n^\lambda - \Sigma\|_2^2 \leq \min\left\{2\lambda\|\Sigma\|_1, 3\lambda^2 \operatorname{rank}(\Sigma)\right\}$.

The statement of Proposition 2.2 is also a general algebraic result that is standard for optimization problems under nuclear penalization (e.g., Koltchinskii et al. (2011)). It says that an oracle inequality for $\widehat{\Sigma}_n^\lambda$ is available if we can control the stochastic error $\|\widehat{\Sigma}_n - \Sigma\|_\infty$. The latter is absolute key to our investigation.

In order to assess this error, we need to impose the following assumption on the norms of the drift and volatility processes.

Assumption (A): $\sup_{t\in[0,1]}\|\mu_t\|_2^2 \leq \nu_\mu$, $\sup_{t\in[0,1]}\mathrm{tr}(c_t) \leq \nu_{c,2}$, and $\sup_{t\in[0,1]}\|c_t\|_\infty \leq \nu_{c,\infty}$ for some constants ν_μ, $\nu_{c,2}$, and $\nu_{c,\infty}$ with $\nu_{c,\infty} \leq \nu_{c,2}$.

Mathematically speaking, Assumption (A) appears a bit strong, since it imposes an almost sure bound on the drift and volatility. We can weaken the requirement on the drift to a suitable moment condition without affecting the rate of $\|\widehat{\Sigma}_n - \Sigma\|_\infty$ in Theorem 2.4 below, but as usual the cost of this is more involved expressions. If the volatility does not meet the boundedness condition, which it does not for most stochastic volatility models, one can resort to the localization technique from Sect. 4.4.1 in Jacod and Protter (2012). In most financial applications, however, Assumption (A) is not too stringent if the drift and volatility do not vary strongly over short time intervals, such as a day.

The constants ν_μ, $\nu_{c,2}$, and $\nu_{c,\infty}$ may depend on d, but they should generally be as small as possible to get the best rate (see, e.g., (2.13)). For example, if the components of μ and c are uniformly bounded, we readily deduce that

$$\nu_\mu = O(d) \quad \text{and} \quad \nu_{c,2} = O(d). \tag{2.4}$$

To study the concentration of $\|\widehat{\Sigma}_n - \Sigma\|_\infty$, we present an exponential Bernstein-type inequality, which applies to matrix-valued martingale differences as long as the conditional moments are sufficiently well-behaved. While there are several related concentration inequalities (see, e.g., Minsker (2017); Tropp (2011, 2012, 2015) and references therein), existing results require that summands are either independent or bounded. Thus, to be applicable in our setting we needed to modify these.

Theorem 2.3 *Suppose that $(X_k)_{k=1}^n$ is a martingale difference sequence with respect to a filtration $(\mathcal{G}_k)_{k=0}^n$ and with values in the space of symmetric $d \times d$ matrices. Assume that, for some predictable sequence $(C_k)_{k=1}^n$ of random variables and constant $R \in (0, \infty)$,*

$$\left\|\mathbb{E}\left[X_k^p \mid \mathcal{G}_{k-1}\right]\right\|_\infty \leq \frac{p!}{2}R^p C_k, \quad \text{for } k = 1, \ldots, n \text{ and } p = 2, 3, \ldots \tag{2.5}$$

Then, for all $x, \nu \in (0, \infty)$, it holds that

$$\mathbb{P}\left(\left\|\sum_{k=1}^n X_k\right\|_\infty > x, \ \sum_{k=1}^n C_k \leq \nu\right) \leq 2d \exp\left(-\frac{x^2}{2(xR + \nu R^2)}\right). \tag{2.6}$$

Assume that Theorem 2.3 applies and that $C_1, \ldots, C_n$ in (2.5) can also be chosen to be deterministic. Then, with $\nu = \sum_{k=1}^n C_k$,

$$\mathbb{P}\left(\left\|\sum_{k=1}^n X_k\right\|_\infty > x\right) \leq 2d \exp\left(-\frac{x}{4\nu R^2}\min\{x, \nu R\}\right), \tag{2.7}$$

so $\|\sum_{k=1}^n X_k\|_\infty$ has a sub-exponential tail. The next theorem derives a concentration inequality for $\|\widehat{\Sigma}_n - \Sigma\|_\infty$. We remark that, although we have Theorem 2.3 at our disposal, the derivation of Theorem 2.4 requires a number of non-standard inequalities in order to prove that R in (2.5) can be chosen sufficiently small. For further details, see the discussion in relation to its proof in the supplementary file.

Theorem 2.4 *Suppose that Assumption (A) holds. Then, there exists an absolute constant γ such that*

$$\mathbb{P}\left(\|\widehat{\Sigma}_n - \Sigma\|_\infty > x\right) \leq 6d \exp\left(-\frac{x}{\gamma \nu_{c,2}\nu_{c,\infty}\Delta_n}\min\{x, \nu_{c,\infty}\}\right), \tag{2.8}$$

for all $x \geq \gamma \max\left\{\dfrac{\nu_{c,2}\nu_\mu\Delta_n}{\nu_{c,\infty}}, \sqrt{\nu_{c,2}\nu_\mu\Delta_n}, \nu_{c,\infty}\Delta_n\right\}$.

We now combine Proposition 2.2 and Theorem 2.4 to deliver a result on the concentration of $\|\widehat{\Sigma}_n - \Sigma\|_2$, which is the main statement of this section.

Theorem 2.5 *Suppose that Assumption (A) holds. For a given $\tau \in (0, \infty)$ consider a regularization parameter λ such that*

$$\lambda \geq \gamma \max \left\{ \sqrt{v_{c,2} v_{c,\infty} \Delta_n (\log(d) + \tau)}, \, v_{c,2} \Delta_n (\log(d) + \tau), \, \frac{v_{c,2} v_\mu \Delta_n}{v_{c,\infty}} \right\}, \tag{2.9}$$

where γ is a large absolute constant. Then, with probability at least $1 - e^{-\tau}$,

$$\|\widehat{\Sigma}_n^\lambda - \Sigma\|_2^2 \leq \|\Sigma - A\|_2^2 + \min\{2\lambda \|A\|_1, 3\lambda^2 \operatorname{rank}(A)\}, \tag{2.10}$$

for all $A \in \mathbb{R}^{d \times d}$.

If the drift term is non-dominant, in the sense that $v_\mu \leq v_{c,\infty}$, it follows that the regularization parameter λ should meet

$$\lambda \geq \gamma \max \left\{ \sqrt{v_{c,2} v_{c,\infty} \Delta_n (\log(d) + \tau)}, \, v_{c,2} \Delta_n (\log(d) + \tau) \right\}, \tag{2.11}$$

for an absolute constant γ. To get a concentration probability as large as possible without impairing the rate implied by (2.10), we should choose $\tau \asymp \log(d)$. Moreover, the first term of the maximum in (2.11) is largest for $1/\Delta_n \geq (\log(d) + \tau) v_{c,2}/v_{c,\infty}$. In light of these observations, the following corollary is an immediate consequence of Theorem 2.5, so we exclude the proof.

Corollary 2.6 *Suppose that Assumption (A) holds with $v_\mu \leq v_{c,\infty} \leq v_{c,2}$. Assume further that $\Delta_n^{-1} \geq 2 \frac{v_{c,2}}{v_{c,\infty}} \log(d)$ and choose the regularization parameter*

$$\lambda = \gamma \sqrt{v_{c,2} v_{c,\infty} \Delta_n \log(d)}, \tag{2.12}$$

where γ is a sufficiently large absolute constant. Then, it holds that

$$\|\widehat{\Sigma}_n^\lambda - \Sigma\|_2^2 \leq 3\gamma^2 v_{c,2} v_{c,\infty} \operatorname{rank}(\Sigma) \Delta_n \log(d) \tag{2.13}$$

with probability at least $1 - d^{-1}$.

It follows from (2.13) that the estimation error of $\widehat{\Sigma}_n^\lambda$ is closely related to the size of $v_{c,2}$ and $v_{c,\infty}$, which are both determined from the properties of the volatility process. If c is uniformly bounded, $v_{c,2} = O(d)$. As emphasized by Tropp (2015, Section 1.6.3), we can also often assume that $\|c_s\|_\infty$ and, hence, $v_{c,\infty}$ can be chosen independently of d. When this is the case, the rate implied by (2.13) is $\operatorname{rank}(\Sigma) \Delta_n d \log(d)$, and since $\operatorname{rank}(\Sigma) \leq d$, Corollary 2.6 implies consistency of $\widehat{\Sigma}_n^\lambda$ when both $\Delta_n \to 0$ and $d \to \infty$ so long as $d^2 \log(d) = o(\Delta_n^{-1})$. If $\operatorname{rank}(\Sigma)$ can be further bounded by a constant (that does not depend on d), the growth condition on d improves to $d \log(d) = o(\Delta_n^{-1})$. In contrast, for the estimation error $\|\widehat{\Sigma}_n - \Sigma\|_2^2 \to 0$, one cannot expect a better condition than $d^2 = o(\Delta_n^{-1})$, since this estimation error corresponds to a sum of d^2 squared error terms, each of the order Δ_n.

2.2 Minimax optimality

We denote by $\mathcal{C}_r$, for a given non-zero integer $r \leq d$, the subclass of $d \times d$ symmetric positive semidefinite matrices $\mathcal{S}_+^d$ whose *effective rank* is bounded by r:

$$\mathcal{C}_r = \left\{ A \in \mathcal{S}_+^d : r_e(A) \leq r \right\}, \tag{2.14}$$

where

$$r_e(A) \equiv \frac{\text{tr}(A)}{\|A\|_\infty}. \tag{2.15}$$

Compared to the rank, the effective rank is a more stable measure for the intrinsic dimension of a covariance matrix (see, e.g., (Vershynin 2010, Remark 5.53)). In the following we argue that $\widehat{\Sigma}_n^\lambda$, with λ of the form (2.12), is a minimax optimal estimator of Σ in the parameters $v_{c,2}$, $v_{c,\infty}$, and rank(Σ) over the parametric class of continuous Itô processes generated by (2.1) with no drift $\mu_s \equiv 0$ and a constant volatility $\sigma_s \equiv \sqrt{A}$ for $A \in \mathcal{C}_r$. To this end, denote by $\mathbb{P}_A$ a probability measure for which $(Y_t)_{t\in[0,1]}$ is defined as in (2.1) with no drift and constant volatility $c_s \equiv A$. In this setting, we can choose $v_{c,2} = \text{tr}(A)$ and $v_{c,\infty} = \|A\|_\infty$, which by Corollary 2.6 means that

$$\sup_{A\in\mathcal{C}_r} \mathbb{P}_A \left(\|\widehat{\Sigma}_n^\lambda - \Sigma\|_2^2 > \overline{\gamma} \|A\|_\infty^2 r_e(A) \text{rank}(A) \Delta_n \log(d) \right) \leq \frac{1}{d}, \tag{2.16}$$

for an absolute constant $\overline{\gamma}$ and $\Delta_n^{-1} \geq 2r \log(d)$.

Now, since the log-price increments $\Delta_1^n Y, \ldots, \Delta_{\lfloor \Delta_n^{-1} \rfloor}^n Y$ are i.i.d. Gaussian random vectors under $\mathbb{P}_A$, we can exploit existing results from the literature to assess the performance of $\widehat{\Sigma}_n^\lambda$. Hence, the following is effectively an immediate consequence of Theorem 2 in Lounici (2014). It shows that, up to a logarithmic factor, no estimator can do better than (2.16).

Theorem 2.7 *Let $\mathcal{C}_r$ be given as in* (2.14) *and suppose that $\lfloor \Delta_n^{-1} \rfloor \geq r^2$. Then, there exist absolute constants $\beta \in (0, 1)$ and $\underline{\gamma} \in (0, \infty)$ such that*

$$\inf_{\widehat{\Sigma}} \sup_{A\in\mathcal{C}_r} \mathbb{P}_A \left(\|\widehat{\Sigma} - \Sigma\|_2^2 > \underline{\gamma} \frac{\|A\|_\infty^2 r_e(A) \text{rank}(A)}{\lfloor \Delta_n^{-1} \rfloor} \right) \geq \beta, \tag{2.17}$$

where the infimum runs over all estimators $\widehat{\Sigma}$ of Σ based on $(\Delta_1^n Y, \ldots, \Delta_{\lfloor \Delta_n^{-1} \rfloor}^n Y)$.

2.3 Bound on the rank of PRV

In this section, we study the rank of $\widehat{\Sigma}_n^\lambda$ relative to the number of non-negligible eigenvalues and, in particular, the true rank of Σ. In line with Sect. 2.1, we begin by stating a general non-probabilistic inequality in Theorem 2.8. In the formulation of this result, we recall that rank$(A; \varepsilon)$ is the number of singular values of A exceeding $\varepsilon \in [0, \infty)$.

Theorem 2.8 *Suppose that $2\|\widehat{\Sigma}_n - \Sigma\|_\infty \leq \bar{\lambda}$ for some $\bar{\lambda} \in (0, \infty)$. Then*

$$\text{rank}(\Sigma; \lambda) \leq \text{rank}(\widehat{\Sigma}_n^\lambda) \leq \text{rank}\left(\Sigma; \frac{1}{2}(\lambda - \bar{\lambda})\right), \tag{2.18}$$

for any $\lambda \in (\bar{\lambda}, \infty)$. In particular, if $\lambda \in (\bar{\lambda}, s]$, where s is the smallest non-zero eigenvalue of Σ, then rank$(\widehat{\Sigma}_n^\lambda) = \text{rank}(\Sigma)$.

With this result in hand we can rely on the exponential inequality for the quantity $\|\widehat{\Sigma}_n - \Sigma\|_\infty$ established in Theorem 2.4 to show that, with high probability and in addition to converging to Σ at a fast rate, $\widehat{\Sigma}_n^\lambda$ automatically has the rank of Σ when we neglect eigenvalues of sufficiently small order. In particular, if Σ has full rank, but many of its eigenvalues are close to zero, $\widehat{\Sigma}_n^\lambda$ is of low rank and reflects the number of important components (or factors) in Σ.

Corollary 2.9 *Suppose that Assumption (A) holds with $v_\mu \leq v_{c,\infty} \leq v_{c,2}$ and fix $\delta \in (0, 1/2)$. Assume further that $\Delta_n^{-1} \geq 2\frac{v_{c,2}}{v_{c,\infty}}\log(d)$ and, for a sufficiently large constant γ depending only on δ, choose the regularization parameter λ as follows:*

$$\lambda = \gamma\sqrt{v_{c,2}v_{c,\infty}\Delta_n\log(d)}. \tag{2.19}$$

Then, with probability at least $1 - d^{-1}$, it holds that

$$\|\widehat{\Sigma}_n^\lambda - \Sigma\|_2^2 \leq 3\gamma^2 v_{c,2}v_{c,\infty}\,\text{rank}(\Sigma)\Delta_n\log(d), \tag{2.20}$$

and

$$\text{rank}(\Sigma; \lambda) \leq \text{rank}(\widehat{\Sigma}_n^\lambda) \leq \text{rank}(\Sigma; \delta\lambda). \tag{2.21}$$

If, in addition, $\lambda \leq s$, where s is the smallest non-zero eigenvalue of Σ, then both $\text{rank}(\widehat{\Sigma}_n^\lambda) = \text{rank}(\Sigma)$ and (2.20) hold with probability at least $1 - d^{-1}$.

In the setting of Corollary 2.9, it follows that with high probability an eigenvalue s of Σ affects $\text{rank}(\widehat{\Sigma}_n^\lambda)$ for large n if $\lambda = o(s)$, while it does not if $s = o(\lambda)$. The first condition says that, relative to the level of shrinkage, an eigenvalue is significant (or non-negligible), whereas the second says the opposite. Hence, the notion of negligibility depends on λ, which in turn depends on the model through the constants $v_{c,2}$ and $v_{c,\infty}$. However, we know that for $\widehat{\Sigma}_n^\lambda$ to be a consistent estimator of Σ, a necessary condition is that $\lambda \to 0$ as $n \to \infty$, which implies that for an eigenvalue of Σ to be negligible, it must tend to zero as n increases. In particular, if d is fixed, $\text{rank}(\widehat{\Sigma}_n^\lambda) = \text{rank}(\Sigma)$ with a probability tending to one as $n \to \infty$.

The following example illustrates a stylized setting, where many eigenvalues of Σ are negligible.

Example 2.10 Let $r \in \mathbb{N}$ be an absolute constant (i.e., independent of d and n). In line with the factor model studied in Aït-Sahalia and Xiu (2017), suppose that c_t is of the form

$$c_t = \beta e_t \beta^\top + g_t, \tag{2.22}$$

where $e_t \in \mathbb{R}^{r \times r}$ and $g_t \in \mathbb{R}^{d \times d}$ are predictable, symmetric, and positive definite processes, which are uniformly bounded such that $\|e_t\|_\infty \leq C_e$ and $\|g_t\|_\infty \leq C_g$ for some constants $C_e, C_g = O(1)$. The matrix $\beta \in \mathbb{R}^{d \times r}$ of factor loadings is deterministic and constant in time; a common assumption in the literature, which is also supported by Reiss et al. (2015). The form (2.22) of c_t can be motivated by a standard multi-factor model, where the total risk associated with any of the d assets can be decomposed into a systematic and an idiosyncratic component. The systematic component $\beta e_t \beta^\top$ is composed of loadings on the r underlying priced common factors in the economy, β, and the risk of those factors, e_t. The idiosyncratic component g_t is asset-specific and can therefore be reduced by diversification.

Suppose that, given an initial set of d individual securities, we start forming a new set of (orthogonalized) portfolios by taking linear combinations of the original assets, as prescribed by the APT model of Ross (1976), and a standard way to implement factor models in practice when constructing sorting portfolios. Then, these new assets are generally diversified enough to assume that $C_g = O(d^{-1})$. To identify the number of driving factors, r, we assume that $\|d^{-\gamma}\beta^\top\beta - I_r\|_\infty = o(1)$ (as $d \to \infty$) and $s_r(e_t) \geq \varepsilon$ for some absolute constants $\gamma \in [0, 1]$ and $\varepsilon \in (0, \infty)$. This corresponds to Assumption 4 in Aït-Sahalia and Xiu (2017), but it is slightly more general as the assets are assumed to be diversified portfolios.

Now, given a suitable drift process $(\mu_t)_{t\in[0,1]}$, Corollary 2.9 applies with regularization parameter $\lambda = O\left(d^\gamma\sqrt{\Delta_n\log(d)}\right)$. In particular, due to the fact $0 \leq s_k(\Sigma) - s_k(\beta E \beta^\top) \leq$

$\|\Gamma\|_\infty$ with $E = \int_0^1 e_t \, dt$ and $\Gamma = \int_0^1 g_t \, dt$, it follows from (2.21) that

$$\text{rank}(\beta E \beta^\top; \lambda) \leq \text{rank}(\widehat{\Sigma}_n^\lambda) \leq \text{rank}(\beta E \beta^\top; \delta\lambda - \|\Gamma\|_\infty) \qquad (2.23)$$

with probability at least $1 - d^{-1}$. By assumption $|s_r(\beta E \beta^\top) - d^\gamma s_r(E)| = o(d^\gamma)$ and $s_r(E) \geq \varepsilon$ (the former follows from arguments as in the proof of Theorem 1 in Aït-Sahalia and Xiu (2017)), and hence $d^\gamma = O(s_r(\beta E \beta^\top))$. Combining this with (2.23), we conclude that $\text{rank}(\widehat{\Sigma}_n^\lambda) = r$ with probability at least $1 - d^{-1}$ if both $\log(d) = o(\Delta_n^{-1})$ and $\Delta_n^{-1} = o(d^{2+2\gamma} \log(d))$, and n is large. Or to put it differently, $\text{rank}(\widehat{\Sigma}_n^\lambda)$ is, with a probability tending to one, exactly the number of underlying factors in the model.

We remark that the restrictions imposed above are too weak to ensure that the Frobenius estimation error $\|\widehat{\Sigma}_n^\lambda - \Sigma\|_2$ tends to zero, meaning that our estimator can be used to detect the underlying factor structure even in very high-dimensional settings. To ensure the estimation error is small, we need $d^{1+2\gamma} \log(d) = o(\Delta_n^{-1})$ rather than $\log(d) = o(\Delta_n^{-1})$.

2.4 Selection of tuning parameter

In view of Theorem 2.7 and Corollary 2.9, it follows that $\widehat{\Sigma}_n^\lambda$ can be of low rank and accurate, given optimal tuning of λ. However, as evident from (2.19), λ depends on the latent spot variance process $(c_t)_{t \in [0,1]}$ through $\nu_{c,2}$ and $\nu_{c,\infty}$ as well as the unknown absolute constant γ, whose value can be important in finite samples.

We remark that $\widehat{\Sigma}_n - \Sigma$ provides an estimate of the null matrix, but it is perturbed by randomness in the data. Hence, a good choice of shrinkage parameter exactly shrinks the eigenvalues of $\widehat{\Sigma}_n - \Sigma$ to zero (in view of problem (1.2) with $\widehat{\Sigma}_n$ replaced by $\widehat{\Sigma}_n - \Sigma$). By Proposition 2.1, this means $\lambda = 2\|\widehat{\Sigma}_n - \Sigma\|_\infty$.

The above is unobservable. To circumvent this problem and facilitate the calculation of our estimator in Sect. 4 and 5, we adopt a data-driven shrinkage selection by exploiting the subsampling technique of Christensen et al. (2017), see also Politis et al. (1999) and Kalnina (2011). To explain the procedure in short, suppose for notational convenience that $n = \Delta_n^{-1}$. We select an integer L—the number of subsamples—that divides n and assign log-returns successively to each subsample. Hence, the lth subsample consists of the increments $\left(\Delta_{(k-1)L+l}^n Y\right)_{k=1,\dots,n/L}$ for $l = 1, \dots, L$. We denote the associated subsampled RV by

$$\widehat{\Sigma}_{n,l} = \frac{1}{n/L} \sum_{k=1}^{n/L} (\sqrt{n}\Delta_{(k-1)L+l}^n Y)(\sqrt{n}\Delta_{(k-1)L+l}^n Y)^\top, \qquad (2.24)$$

Note that the random matrices in the sequence $(\widehat{\Sigma}_{n,l})_{l=1}^L$ are asymptotically conditionally independent, as $n \to \infty$. Moreover, it follows from Christensen et al. (2017) that as $n \to \infty$, $L \to \infty$ and $n/L \to \infty$, the half-vectorization of $\sqrt{n/L}(\widehat{\Sigma}_{n,l} - \widehat{\Sigma}_n)$ and $\sqrt{n}(\widehat{\Sigma}_n - \Sigma)$ converge in law to a common mixed normal distribution, derived in Barndorff-Nielsen and Shephard (2004, Theorem 1).

Hence, in each subsample $\lambda = 2\|\widehat{\Sigma}_n - \Sigma\|_\infty$ can be approximated by

$$\lambda_l = \frac{2}{\sqrt{L}} \|\widehat{\Sigma}_{n,l} - \widehat{\Sigma}_n\|_\infty \qquad (2.25)$$

As an estimator of λ, we therefore take the sample average:

$$\lambda^* = \frac{1}{L} \sum_{l=1}^L \lambda_l. \qquad (2.26)$$

3 Estimation of the local variance

As noted in Sect. 1, the rank of the local variance c_t is often much smaller than the rank of Σ. In fact, although Σ may be well-approximated by a matrix of low rank, we should expect that rank$(\Sigma) = d$. On the other hand, there are many situations where it is reasonable to expect that rank(c_t) is small, in which case it provides valuable insight about the complexity of the underlying model. This motivates developing a theory for the spot version of our penalized estimator with particular attention on its ability to identify rank(c_t).

To estimate c_t, we follow Jacod and Protter (2012) and apply a localized realized variance, which is defined over a block of $\lfloor h_n/\Delta_n \rfloor \geq 2$ log-price increments with $h_n \in (0, 1)$. The RV computed over the time window $[t, t+h_n]$, for $t \in (0, 1-h_n)$, is then defined as follows:

$$\widehat{\Sigma}_n(t; t + h_n) = \sum_{k=\lfloor t/\Delta_n \rfloor +1}^{\lfloor (t+h_n)/\Delta_n \rfloor} (\Delta_k^n Y)(\Delta_k^n Y)^\top. \tag{3.1}$$

The corresponding penalized version $\widehat{\Sigma}_n^\lambda(t; t + h_n)$ is computed as

$$\widehat{\Sigma}_n^\lambda(t; t + h_n) = \arg\min_{A \in \mathbb{R}^{d \times d}} \left(\|\widehat{\Sigma}_n(t; t + h_n) - A\|_2^2 + \lambda \|A\|_1 \right), \tag{3.2}$$

or by using Proposition 2.1 with $\widehat{\Sigma}_n$ replaced by $\widehat{\Sigma}_n(t; t+h_n)$. Then, the penalized estimator $\hat{c}_n^\lambda(t)$ of c_t is given by

$$\hat{c}_n^\lambda(t) = \frac{\widehat{\Sigma}_n^\lambda(t; t + h_n)}{h_n}. \tag{3.3}$$

Also, we write

$$\Sigma(t_1; t_2) = \int_{t_1}^{t_2} c_s \, ds, \tag{3.4}$$

for the QV over $(t_1, t_2]$ for arbitrary time points $t_1, t_2 \in [0, 1]$ with $t_1 < t_2$.

Recall that for a convex and increasing function $\psi : [0, \infty) \to [0, \infty)$ with $\psi(0) = 0$, the Orlicz norm of a random variable Z with values in the Hilbert space $(\mathbb{R}^{d \times d}, \| \cdot \|_2)$ is defined as:

$$\|Z\|_\psi \equiv \inf \{c > 0 : \mathbb{E}[\psi(\|Z\|_2/c)] \leq 1\}. \tag{3.5}$$

This setting includes, in particular, the $L^p(\mathbb{P})$ norms (with $\psi(s) = s^p$), but also the ψ_1 and ψ_2 norms for sub-exponential ($\psi(s) = e^s - 1$) and sub-Gaussian ($\psi(s) = e^{s^2} - 1$) random variables. For convenience, we further impose the mild restriction that the range of ψ is $[0, \infty)$, so it admits an inverse ψ^{-1} on $[0, \infty)$.

Theorem 3.1 *Suppose that Assumption (A) holds with $v_\mu \leq v_{c,\infty} \leq v_{c,2}$. Furthermore, let $t \in \left[\frac{h_n}{2}, 1 - \frac{3h_n}{2}\right]$ and assume $\sup_{0 \leq u < s \leq 1} \|c_s - c_u\|_\psi / \sqrt{s - u} \leq v_{c,\psi}$ and $h_n \geq 2\Delta_n \frac{v_{c,2}}{v_{c,\infty}} \log(d)$. Take the regularization parameter λ as*

$$\lambda = \gamma \sqrt{h_n \Delta_n v_{c,2} v_{c,\infty} \log(d)} \tag{3.6}$$

for a sufficiently large absolute constant γ. Then,

$$\|\hat{c}_n^\lambda(t) - c_t\|_2^2 \leq \kappa \gamma^2 \left(\frac{v_{c,2} v_{c,\infty} \Delta_n \, \mathrm{rank}\left(\Sigma\left(t - \frac{h_n}{2}; t + \frac{3h_n}{2}\right)\right)}{h_n} + h_n v_{c,\psi}^2 \right) \log(d), \tag{3.7}$$

with probability at least $1 - d^{-1} - \psi(\sqrt{\log(d)})^{-1}$ for some absolute constant κ.

The bound on the estimation error of $\hat{c}_n^\lambda(t)$ builds on Corollary 2.6, but Theorem 3.1 further enforces a smoothness condition on $(c_t)_{t\in[0,1]}$. It entails a trade-off between the smoothness of spot variance and the concentration probability associated with (3.7). For example, if $(c_t)_{t\in[0,1]}$ is 1/2-Hölder continuous in $L^2(\mathbb{P})$, which corresponds to setting $\psi(s) = s^2$, then we end up with a concentration probability converging to one at rate $\log(d)^{-1}$, which is slower than for the PRV. On the other hand, if $(c_t)_{t\in[0,1]}$ is 1/2-Hölder continuous in the sub-Gaussian norm $\psi(s) = e^{s^2} - 1$, the concentration probability converges to one at the rate d^{-1}, which is equivalent to the PRV.

Note that with d fixed, (3.7) reveals how to select the length of the estimation window h_n optimally. The upper bound depends on Δ_n/h_n and h_n, and for these terms to converge to zero equally fast, we should take $h_n \asymp \sqrt{\Delta_n}$. This is consistent with the literature on spot variance estimation; see, e.g., Jacod and Protter (2012).

The next result, which relies on Corollary 2.9, shows that the rank and performance of $\hat{c}_n^\lambda(t)$ can be controlled simultaneously.

Theorem 3.2 *Suppose that Assumption (A) holds with $v_\mu \le v_{c,\infty} \le v_{c,2}$ and fix $\delta \in (0, 1/4)$. Furthermore, let $t \in [\frac{h_n}{2}, 1 - \frac{3h_n}{2}]$ and assume that $\sup_{0\le u < s \le 1} \|c_s - c_u\|_\psi / \sqrt{s-u} \le v_{c,\psi}$. Suppose further that $h_n \ge 2\Delta_n \frac{v_{c,2}}{v_{c,\infty}} \log(d)$ and consider a regularization parameter λ such that*

$$\lambda = \gamma \sqrt{h_n \Delta_n v_{c,2} v_{c,\infty} \log(d)}, \qquad (3.8)$$

for a sufficiently large constant γ depending only on δ. Then, with probability at least $1 - d^{-1} - \psi(\sqrt{\log(d)})^{-1}$, it holds that

$$\|\hat{c}_n^\lambda(t) - c_t\|_2^2 \le \kappa\gamma^2 \left(\frac{v_{c,2} v_{c,\infty} \Delta_n \, \mathrm{rank}\left(\Sigma(t - \frac{h_n}{2}; t + \frac{3h_n}{2})\right)}{h_n} + h_n v_{c,\psi}^2 \right) \log(d), \quad (3.9)$$

and

$$\mathrm{rank}(c_t; \overline{\varepsilon}) \le \mathrm{rank}(\hat{c}_n^\lambda(t)) \le \mathrm{rank}(c_t; \underline{\varepsilon}), \qquad (3.10)$$

where

$$
\begin{aligned}
\underline{\varepsilon} &\equiv \delta\gamma \max\left\{ \sqrt{\frac{v_{c,2} v_{c,\infty} \Delta_n}{h_n}} - v_{c,\psi}\sqrt{h_n}, \, 0 \right\} \sqrt{\log(d)}, \\
\overline{\varepsilon} &\equiv \gamma \left(\sqrt{\frac{v_{c,2} v_{c,\infty} \Delta_n}{h_n}} + v_{c,\psi}\sqrt{h_n} \right) \sqrt{\log(d)},
\end{aligned}
\qquad (3.11)
$$

and κ is an absolute constant. If, in addition, $\underline{\varepsilon} > 0$ and $\overline{\varepsilon} \le s_{\mathrm{rank}(c_t)}(c_t)$, then both (3.9) and $\mathrm{rank}(\hat{c}_n^\lambda(t)) = \mathrm{rank}(c_t)$ hold with probability at least $1 - d^{-1} - \psi(\sqrt{\log(d)})^{-1}$.

Consider the setting of Theorem 3.2. For the upper bound in (3.10) to be useful, we must have that $\underline{\varepsilon} > 0$ or, equivalently,

$$v_{c,\psi}^2 h_n^2 < v_{c,2} v_{c,\infty} \Delta_n. \qquad (3.12)$$

Suppose that $v_{c,2}$, $v_{c,\infty}$, and $v_{c,\psi}$ do not depend on n. The inequality (3.12) can then be achieved in large samples by choosing $h_n = o(\sqrt{\Delta_n})$. However, as pointed out above one should take $h_n \asymp \sqrt{\Delta_n}$ in order to achieve an optimal rate for the estimation error $\|\hat{c}_n^\lambda(t) - c_t\|_2^2$. In that case, (3.12) translates directly into an upper bound on $v_{c,\psi}$, which concerns the smoothness of $(c_t)_{t\in[0,1]}$.

When can the term $\text{rank}(\Sigma(t - \frac{h_n}{2}; t + \frac{3h_n}{2}))$ in the upper bound of (3.9) be replaced by $\text{rank}(c_t)$? This question appears difficult to answer in general, but it is not too difficult to show that the replacement is valid if the volatility process is locally constant.

4 Simulation study

In this section, we do a Monte Carlo analysis to inspect the finite sample properties of the PRV, $\widehat{\Sigma}_n^\lambda$. The aim here is to demonstrate the ability of our estimator to detect the presence of near deficient rank in a large-dimensional setting. In view of Theorem 3.2, this can be achieved by a proper selection of λ. We simulate a $d = 30$-dimensional log-price process Y_t. The size of the vector corresponds to the number of assets in our empirical investigation. We assume Y_t follows a slightly modified version of the $r = 3$-factor model proposed in Aït-Sahalia and Xiu (2019):

$$\mathrm{d}Y_t = \beta \mathrm{d}F_t + \mathrm{d}Z_t, \quad t \in [0, 1], \tag{4.1}$$

where $F_t \in \mathbb{R}^r$ is a vector of systematic risk factors, which has the dynamic:

$$\mathrm{d}F_t = \alpha \mathrm{d}t + \sigma_t L \mathrm{d}W_t, \tag{4.2}$$

and $W_t \in \mathbb{R}^r$ is a standard Brownian motion. The drift term $\alpha \in \mathbb{R}^r$ is constant $\alpha = (0.05, 0.03, 0.02)^\top$. The random volatility matrix $\sigma_t = \text{diag}(\sigma_{1,t}, \sigma_{2,t}, \sigma_{3,t}) \in \mathbb{R}^{r \times r}$, where $\text{diag}(\cdot)$ is a diagonal matrix with coordinates $\cdot$, is simulated as a Heston (1993) square-root process:

$$\mathrm{d}\sigma_{j,t}^2 = \kappa_j(\theta_j - \sigma_{j,t}^2)\mathrm{d}t + \eta_j \sigma_{j,t}\left(\rho_j \mathrm{d}W_{j,t} + \sqrt{1 - \rho_j^2}\,\mathrm{d}\widetilde{W}_{j,t}\right), \tag{4.3}$$

for $j = 1, \ldots, r$, where $\widetilde{W}_t \in \mathbb{R}^r$ is an r-dimensional standard Brownian motion independent of W_t. As in Aït-Sahalia and Xiu (2019), the parameters are $\kappa = (3, 4, 5)^\top$, $\theta = (0.05, 0.04, 0.03)^\top$, $\eta = (0.3, 0.4, 0.3)^\top$, and $\rho = (-0.60, -0.40, -0.25)^\top$. The factor correlation is captured by the Cholesky component L:

$$\rho = LL^\top = \begin{bmatrix} 1.00 & 0.05 & 0.10 \\ 0.05 & 1.00 & 0.15 \\ 0.10 & 0.15 & 1.00 \end{bmatrix}. \tag{4.4}$$

The idiosyncratic component $Z_t \in \mathbb{R}^d$ is given by

$$\mathrm{d}Z_t = \sqrt{g_t}\mathrm{d}B_t, \tag{4.5}$$

where $g_t = \text{diag}(\gamma_{1,t}^2, \ldots, \gamma_{d,t}^2)$ with

$$\mathrm{d}\gamma_{j,t}^2 = \kappa_Z(\theta_Z - \gamma_{j,t}^2)\mathrm{d}t + \eta_Z \gamma_{j,t}\mathrm{d}\widetilde{B}_{j,t}, \tag{4.6}$$

for $j = 1, \ldots, d$, and $B_t \in \mathbb{R}^d$ and $\widetilde{B}_t \in \mathbb{R}^d$ are independent d-dimensional standard Brownian motions. Thus, g_t is the instantaneous variance of the unsystematic component. We set $\kappa_Z = 4, \theta_Z = 0.25$, and $\eta_Z = 0.06$. Hence, the idiosyncratic error varies independent in the cross-section of assets, but the parameters controlling the dynamics are common.

The above setup implies

$$c_t = \beta \sigma_t \rho \sigma_t \beta^\top + g_t, \tag{4.7}$$

so the spot covariance matrix has full rank at every time point t. However, it has only $r = 3$ large eigenvalues associated with the systematic factors, while the remaining $d - r$ associated
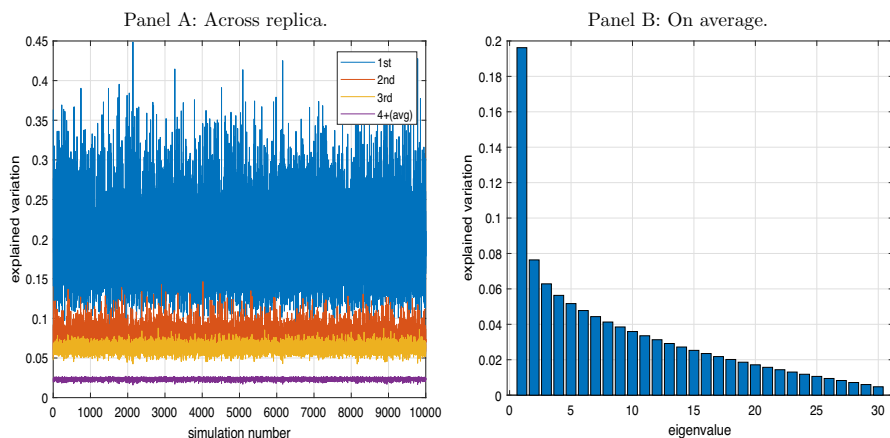
Fig. 1 Relative importance of eigenvalues in simulated model. *Note*. In Panel A, we plot for daily RV the relative size of the three largest eigenvalues, and also the average of the remaining 27 eigenvalues. In Panel B, we plot a histogram of the sample average of the relative size of each of the eigenvalues across simulations

with the idiosyncratic variance are relatively small. Note that in contrast to Aït-Sahalia and Xiu (2019), we allow for time-varying idiosyncratic variance.

We set $\Delta_n = 1/n$ with $n = 78$. This corresponds to 5-minute sampling frequency within a 6.5 hour window. Hence, d is relatively large compared to n. We construct 10,000 independent replications of this model. At the beginning of each simulation, we draw the associated factor loadings $\beta \in \mathbb{R}^{d \times r}$ at random such that the first column (interpreted as the loading on the market factor) are from a uniform distribution on $(0.25, 1.75)$, i.e. $\beta_{i1} \sim U(0.25, 1.75)$, $i = 1, \ldots, d$. The range of the support is consistent with the realized beta values reported in Table 1 in Sect. 5. The remaining columns are generated as $\beta_{ij} \sim N(0, 0.5^2)$, $i = 1, \ldots, d$ and $j = 2$ and 3.

In Panel A of Fig. 1, we plot the relative size of the three largest eigenvalues, extracted from the corresponding RV, together with the average of the remaining 27 eigenvalues. In Panel B, we include a histogram of the relative size of the eigenvalues across simulations. We observe that it is generally challenging to distinguish important factors from idiosyncratic variation, since the relative size of the eigenvalues decays smoothly with the exception of the largest (and perhaps the second largest) eigenvalue.

In each simulation, we employ the subsampling procedure described in Sect. 2.4 with $L = 6$ to choose λ^*.[1]

In Panel A of Fig. 2, we report the distribution of the λ^* parameter across simulations, which is here expressed in percent of the maximal eigenvalue of RV. Overall, the penalization is relatively stable with a tendency to shrink around one-fourth to one-third of the largest eigenvalue most of the times.

In Panel B, we plot a histogram of the relative frequency of the rank of the PRV. In general, the PRV does a good job at identifying the number of driving factors, given the challenging environment. Although there are variations in the rank across simulations, caused by the various sources of randomness incorporated in our model, the picture is close to the truth. This means we tend to identify about one-to-three driving factors. The average rank estimate

---

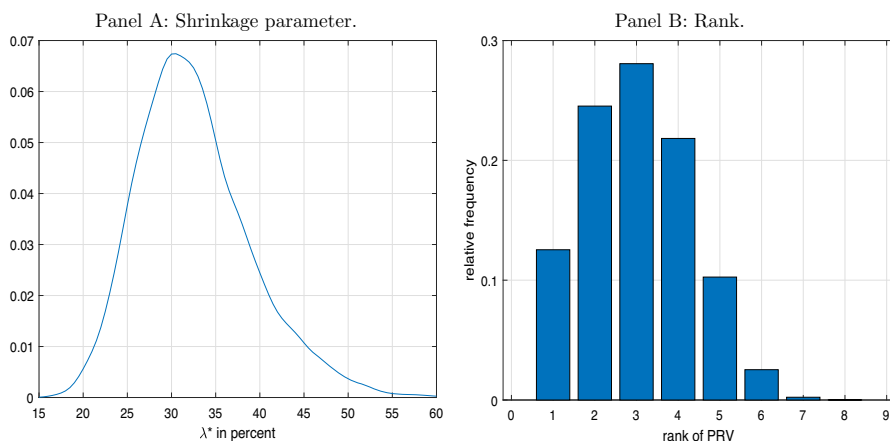[1] We also employed $L = 13$ subsamples, but there were no discernible change in the results.

Fig. 2 Properties of the PRV. *Note.* In Panel A, we report a kernel density estimate of the shrinkage parameter λ, expressed in percent of the maximal eigenvalue of the RV matrix. In Panel B, we show the relative frequency histogram of the associated rank of the PRV

is 1.79, so as expected there is a slight tendency to overshrink leading to a small downward bias. Based on these findings, we can confidently apply the PRV in the empirical application.

5 Empirical application

In this section, we apply the PRV to an empirical dataset. The sample period is from January 3, 2007 through May 29, 2020 (3,375 trading days in total) and includes both the financial crisis around 2007 – 2009 and partly the recent events related to Covid-19.

We look at high-frequency data from the 30 members of the Dow Jones Industrial Average (DJIA) index as of April 6, 2020.[2] The ticker codes of the various firms are listed in Table 1, along with selected descriptive statistics. As readily seen from Table 1, most of these companies are very liquid. In order to compute a daily RV matrix estimate we collect the individual 5-minute log-return series spanning the 6.5 hours of trading on U.S. stock exchanges from 9:30am to 4:00pm EST.[3] Hence, our analysis is based on a relatively large cross-section (i.e., $d = 30$) compared to the sampling frequency (i.e., $n = 78$). The realized beta in Table 1 is calculated with SPY acting as market portfolio. The latter is an exchange-traded fund that tracks the S&P500 and its evolution is representative of the overall performance of the U.S. stock market. The dispersion of realized beta is broadly consistent with the simulated values generated in Sect. 4.

In Fig. 3, we depict the factor structure in the time series of RV. In Panel A, we compute on a daily basis the proportion of the total variance (i.e., trace) explained by each eigenvalue, whereas Panel B reports the sample average of this statistic. As consistent with Aït-Sahalia

---

[2] On April 6, 2020, Raytheon Technologies (RTX) replaced United Technologies (UTX) in the DJIA index following a merger of United Technologies and Raytheon Company. Moreover, on April 2, 2019, Dow replaced DowDuPont following a spin-off from the parent company, which itself was a fusion between Dow Chemical Company and DuPont in 2017. We employ high-frequency data for the preceding member prior to each index update.

[3] We truncate 5-minute log-returns that exceed three local standard deviations, as gauged by the daily bipower variation estimator, in order to get shelter from potential jumps.

Table 1 Descriptive statistics of high-frequency data

| Code | N | σ_{RV} | σ_ϵ | Realized beta | |
				Median	$[q_{0.05}, q_{0.95}]$
AAPL	171,438	0.227	0.165	0.990	[0.539, 1.784]
AXP	42,379	0.279	0.207	1.016	[0.532, 1.730]
BA	41,579	0.249	0.192	0.918	[0.427, 1.489]
CAT	41,943	0.256	0.187	1.154	[0.570, 1.747]
CSCO	111,876	0.223	0.163	0.997	[0.498, 1.470]
CVX	54,037	0.223	0.163	0.887	[0.370, 1.407]
DIS	56,563	0.213	0.157	0.863	[0.411, 1.303]
DOW	35,108	0.240	0.179	1.002	[0.467, 1.567]
GS	42,098	0.287	0.216	1.091	[0.572, 1.812]
HD	50,777	0.222	0.163	0.878	[0.422, 1.343]
IBM	35,955	0.182	0.131	0.769	[0.402, 1.136]
INTC	126,855	0.233	0.174	1.026	[0.531, 1.586]
JNJ	54,527	0.151	0.120	0.534	[0.168, 0.982]
JPM	127,497	0.296	0.216	1.135	[0.680, 1.911]
KO	55,932	0.156	0.126	0.496	[0.116, 0.891]
MCD	36,777	0.168	0.134	0.523	[0.137, 0.902]
MMM	24,253	0.183	0.131	0.825	[0.426, 1.201]
MRK	60,637	0.197	0.161	0.695	[0.241, 1.166]
MSFT	162,004	0.211	0.154	0.957	[0.501, 1.543]
NKE	32,390	0.216	0.171	0.806	[0.360, 1.260]
PFE	97,351	0.196	0.159	0.726	[0.281, 1.184]
PG	53,787	0.157	0.128	0.490	[0.117, 0.918]
RTX	31,372	0.203	0.149	0.861	[0.445, 1.248]
TRV	18,855	0.219	0.175	0.688	[0.220, 1.190]
UNH	37,540	0.251	0.212	0.771	[0.247, 1.307]
V	44,085	0.228	0.188	0.835	[0.322, 1.336]
VZ	69,710	0.185	0.146	0.560	[0.139, 1.033]
WBA	37,302	0.218	0.181	0.786	[0.261, 1.308]
WMT	59,091	0.165	0.135	0.525	[0.154, 0.906]
XOM	90,346	0.203	0.146	0.809	[0.332, 1.261]

Note. "Code" is the ticker symbol, "N" is the number of transactions, "σ_{RV}" is the annualized square-root realized variance, "σ_ϵ" is the standard deviation of the idiosyncratic component. The latter is computed as $\sigma_\epsilon^2 = \sigma_{\mathrm{RV}}^2 - \beta^2 \sigma_{\mathrm{SPY}}^2$, where β is the realized beta between the asset and SPY (acting as market portfolio)

and Xiu (2019), we observe a pronounced dynamic evolution in the contribution of each eigenvalue to RV with notably changes corresponding to times of severe distress in financial markets. The first factor captures the vast majority of aggregate return variation (about 35% on average). It is followed by a few smaller—but still important—factors accountable for an incremental 25% – 30% of the total variance, whereas the last 25 or so eigenvalues are relatively small.

Next, we turn our attention to the PRV estimator. To select the shrinkage parameter λ, we follow the subsampling implementation from the simulation section. The relative frequency
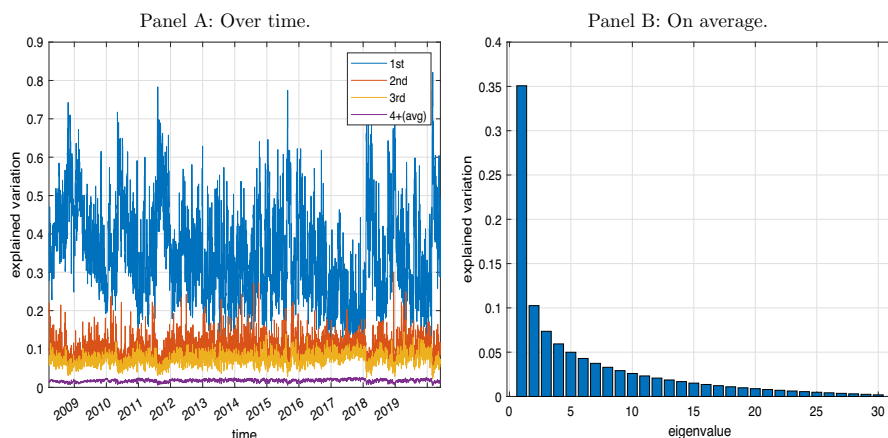
Fig. 3 Proportion of variance explained by each eigenvalue. *Note.* In Panel A, we plot for each day in the sample the relative size of the three largest eigenvalues of RV, as well as the average of the remaining 27. In Panel B, we plot a histogram of the time series average of the relative size of each eigenvalues
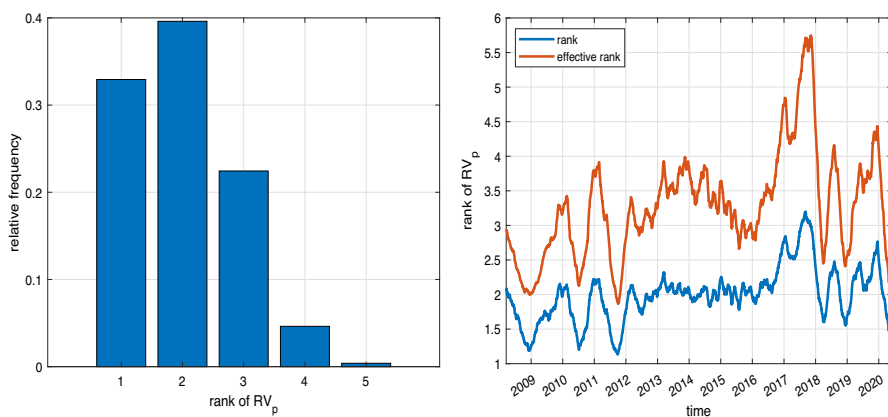

Fig. 4 Rank of PRV. *Note.* In Panel A, we report the relative frequency of the rank of the PRV. In Panel B, we compare the estimated rank to the effective rank, where both are smoothed over a three-month moving average window

histogram of the rank of the PRV is reported in Panel A of Fig. 4, whereas Panel B reports a three-month (90-day) moving average of the rank. The vast majority (around 95%) of the daily rank estimates are between one and three, which is consistent with a standard Fama–French three-factor interpretation of the data. There are relatively few rank estimates at four, and it never exceeds five (with about a dozen of the latter). In Panel B, we observe the rank varies over time in an intuitive fashion, often dropping close to a single-factor representation during times of crisis, where correlations tend to rise. As a comparison, we compute the effective rank of the RV (cf. (2.15)), which does not depend on a shrinkage parameter. There is a high association between the series, which is consistent with a soft-thresholding that eliminates smaller eigenvalues of the RV.

6 Conclusion

In this paper, we develop a novel and powerful penalized realized variance estimator, which is applicable to estimate the quadratic variation of high-dimensional continuous-time semi-martingales under a low rank constraint. Our estimator relies on regularization and adapts the principle ideas of the LASSO from regression analysis to the field of high-frequency volatility estimation in a high-dimensional setting. We derive a non-asymptotic analysis of our estimator, including bounds on its estimation error and rank. The estimator is found to be minimax optimal up to a logaritmic factor. We design a completely data-driven procedure for selecting the shrinkage parameter based on a subsampling approach. In a simulation study, the estimator is found to possess good properties. In our empirical application, we confirm a low rank environment that is consistent with a three-factor structure in the cross-section of log-returns from the large-cap segment of the U.S. stock market.

A Appendix of proofs

This appendix presents the proofs of the results from the main text.

Proof of Proposition 2.1 Since the map $A \mapsto L_n(A) \equiv \|\widehat{\Sigma}_n - A\|_2^2 + \lambda \|A\|_1$ is strictly convex, the minimizer A of L_n is uniquely determined by the property that $0 \in \mathbb{R}^{d \times d}$ belongs to the set

$$\partial L_n(A) = \left\{ 2(A - \widehat{\Sigma}_n) + \lambda \left(\sum_{k=1}^d u_k(A) v_k(A)^\top + P_{S_1(A)^\perp} W P_{S_2(A)^\perp} \right) : \|W\|_\infty \le 1 \right\}. \tag{A.1}$$

Here, $\partial L_n(A)$ denotes the subdifferential of L_n at A, and we employ Watson (1992) to get the explicit expression of this set in (A.1). Moreover, $\{u_1(A), \ldots, u_d(A)\}$ and $\{v_1(A), \ldots, v_d(A)\}$ are the vectors introduced in the notation paragraph associated with the SVD of A, whereas $S_1(A)$ and $S_2(A)$ are the corresponding linear spans, and P_S is the projection matrix onto the subspace S. With $\widehat{\Sigma}_n^\lambda$ given as in (2.2), we find that $\partial L_n(\widehat{\Sigma}_n^\lambda)$ coincides with the set of matrices V of the form:

$$V = -2 \sum_{k: s_k \le \lambda/2} s_k u_k u_k^\top + \lambda P_U W P_U, \tag{A.2}$$

where $W \in \mathbb{R}^{d \times d}$ is such that $\|W\|_\infty \le 1$, and U is the linear span of $\{u_k : s_k \le \lambda/2\}$. With $W = \frac{2}{\lambda} \sum_{k: s_k \le \lambda/2} s_k u_k u_k^\top$, it follows that $\|W\|_\infty \le 1$ by submultiplicativity of the norm and $P_U W P_U = W$. Hence, (A.2) shows that $0 \in \partial L_n(\widehat{\Sigma}_n^\lambda)$, so $\widehat{\Sigma}_n^\lambda$ is the unique minimizer of L_n. □

Proof of Proposition 2.2 The structure of the proof is based similar to the one of Theorem 1 in Koltchinskii et al. (2011), but is modified to fit our setting. Starting from the definition of $\widehat{\Sigma}_n^\lambda$, and since $L_n(A) = \|\widehat{\Sigma}_n - A\|_2^2 + \lambda \|A\|_1$, it follows that

$$L_n(\widehat{\Sigma}_n^\lambda) \le L_n(A) \tag{A.3}$$

for any $A \in \mathbb{R}^{d \times d}$. Note also that

$$\|\widehat{\Sigma}_n - A\|_2^2 - \|A - \Sigma\|_2^2 - \|\widehat{\Sigma}_n - \widehat{\Sigma}_n^\lambda\|_2^2 + \|\widehat{\Sigma}_n^\lambda - \Sigma\|_2^2 = 2\langle \widehat{\Sigma}_n^\lambda - A, \widehat{\Sigma}_n - \Sigma \rangle.$$

By using the above identity and (A.3),

$$\|\widehat{\Sigma}_n^\lambda - \Sigma\|_2^2 \le \|A - \Sigma\|_2^2 + 2\langle \widehat{\Sigma}_n^\lambda - A, \widehat{\Sigma}_n - \Sigma \rangle + \lambda \left(\|A\|_1 - \|\widehat{\Sigma}_n^\lambda\|_1 \right).$$

The first term in the minimum follows by observing that

$$2\langle \widehat{\Sigma}_n^\lambda - A, \widehat{\Sigma}_n - \Sigma \rangle + \lambda(\|A\|_1 - \|\widehat{\Sigma}_n^\lambda\|_1) \le 2\|\widehat{\Sigma}_n^\lambda - A\|_1 \|\widehat{\Sigma}_n - \Sigma\|_\infty + \lambda(\|A\|_1 - \|\widehat{\Sigma}_n^\lambda\|_1)$$
$$\le 2\lambda\|A\|_1,$$

where we employ the trace duality $\langle A_1, A_2 \rangle \le \|A_1\|_1 \|A_2\|_\infty$, the triangle inequality, and the assumption that $2\|\widehat{\Sigma}_n - \Sigma\|_\infty \le \lambda$. To show the second part of the proposition, we observe that if B is a subgradient of L_n at $\widehat{\Sigma}_n^\lambda$, then by definition

$$L_n(A) - L_n(\widehat{\Sigma}_n^\lambda) \ge \langle B, A - \widehat{\Sigma}_n^\lambda \rangle$$

or, equivalently,

$$\|A\|_2^2 + \lambda\|A\|_1 - \|\widehat{\Sigma}_n^\lambda\|_2^2 - \lambda\|\widehat{\Sigma}_n^\lambda\|_1 \ge \langle B + 2\widehat{\Sigma}_n, A - \widehat{\Sigma}_n^\lambda \rangle$$

for all matrices A. This shows that $B + 2\widehat{\Sigma}_n$ is a subgradient of the function $A \mapsto \|A\|_2^2 + \lambda\|A\|_1$ at $\widehat{\Sigma}_n^\lambda$ and, thus, $B = 2\widehat{\Sigma}_n^\lambda - 2\widehat{\Sigma}_n + \lambda\widehat{V}$ for an appropriate $\widehat{V} \in \partial\|\widehat{\Sigma}_n^\lambda\|_1$. Moreover, because $\widehat{\Sigma}_n^\lambda$ is a minimizer of L_n, there must exist $B \in \partial L_n(\widehat{\Sigma}_n^\lambda)$ such that $\langle B, \widehat{\Sigma}_n^\lambda - A \rangle \le 0$ for all $A \in \mathbb{R}^{d \times d}$ (see, e.g., (Clarke 1990), Section 2.4). Combining these facts, we establish that

$$2\langle \widehat{\Sigma}_n^\lambda, \widehat{\Sigma}_n^\lambda - A \rangle + \lambda\langle \widehat{V}, \widehat{\Sigma}_n^\lambda - A \rangle \le 2\langle \widehat{\Sigma}_n, \widehat{\Sigma}_n^\lambda - A \rangle \tag{A.4}$$

for some $\widehat{V} \in \partial\|\widehat{\Sigma}_n^\lambda\|_1$ and any $A \in \mathbb{R}^{d \times d}$. Note the identity

$$2\langle \widehat{\Sigma}_n^\lambda - \Sigma, \widehat{\Sigma}_n^\lambda - A \rangle = \|\widehat{\Sigma}_n^\lambda - \Sigma\|_2^2 + \|\widehat{\Sigma}_n^\lambda - A\|_2^2 - \|A - \Sigma\|_2^2. \tag{A.5}$$

Now, fix $A \in \mathbb{R}^{d \times d}$ and consider a generic matrix $V \in \partial\|A\|_1$. Then, if we subtract $\langle 2\Sigma + \lambda V, \widehat{\Sigma}_n^\lambda - A \rangle$ on both sides of (A.4), exploit (A.5), and note that $\langle \widehat{V} - V, \widehat{\Sigma}_n^\lambda - A \rangle \ge 0$ by the monotonicity of subdifferentials (see (Clarke 1990, Proposition 2.2.9)), we deduce that

$$\|\widehat{\Sigma}_n^\lambda - \Sigma\|_2^2 + \|\widehat{\Sigma}_n^\lambda - A\|_2^2 \le \|A - \Sigma\|_2^2 + \lambda\langle V, A - \widehat{\Sigma}_n^\lambda \rangle + 2\langle \widehat{\Sigma}_n - \Sigma, \widehat{\Sigma}_n^\lambda - A \rangle. \tag{A.6}$$

We shall bound each of the last two terms on the right-hand side of (A.6), starting with $\langle V, A - \widehat{\Sigma}_n^\lambda \rangle$. According to Watson (1992), with $r = \text{rank}(A)$ and $A = \sum_{k=1}^r s_k u_k v_k^\top$ being the SVD of A, the subdifferential $\partial\|A\|_1$ has the characterization:

$$\partial\|A\|_1 = \left\{ \sum_{k=1}^r u_k v_k^\top + P_{S_1^\perp} W P_{S_2^\perp} : \|W\|_\infty \le 1 \right\}.$$

Here, S_1 and S_2 denote the span of $\{u_1, \ldots, u_r\}$ and $\{v_1, \ldots, v_r\}$, respectively. By the polar decomposition, one finds that $W \in \mathbb{R}^{d \times d}$ with $\|W\|_\infty = 1$ and $\langle W, P_{S_1^\perp} \widehat{\Sigma}_n^\lambda P_{S_2^\perp} \rangle = \|P_{S_1^\perp} \widehat{\Sigma}_n^\lambda P_{S_2^\perp}\|_1$. Fixing V to be the subgradient associated with this choice of W,

$$\langle V, A - \widehat{\Sigma}_n^\lambda \rangle = \left\langle \sum_{k=1}^r u_k v_k^\top, A - \widehat{\Sigma}_n^\lambda \right\rangle - \langle W, P_{S_1^\perp} \widehat{\Sigma}_n^\lambda P_{S_2^\perp} \rangle$$

$$= \left\langle \sum_{k=1}^r u_k v_k^\top, P_{S_1}(A - \widehat{\Sigma}_n^\lambda) P_{S_2} \right\rangle - \|P_{S_1^\perp} \widehat{\Sigma}_n^\lambda P_{S_2^\perp}\|_1 \tag{A.7}$$

$$\le \|P_{S_1}(\widehat{\Sigma}_n^\lambda - A) P_{S_2}\|_1 - \|P_{S_1^\perp} \widehat{\Sigma}_n^\lambda P_{S_2^\perp}\|_1$$

$$\le \sqrt{r}\|\widehat{\Sigma}_n^\lambda - A\|_2 - \|P_{S_1^\perp} \widehat{\Sigma}_n^\lambda P_{S_2^\perp}\|_1.$$

The last term on the right-hand side of (A.6) can be handled by setting $B = \widehat{\Sigma}_n - \Sigma - P_{S_1^\perp}(\widehat{\Sigma}_n - \Sigma)P_{S_2^\perp}$. The Cauchy–Schwarz inequality and trace duality then delivers the estimate:

$$
\begin{aligned}
\langle \widehat{\Sigma}_n - \Sigma, \widehat{\Sigma}_n^\lambda - A \rangle &\leq |\langle B, \widehat{\Sigma}_n^\lambda - A \rangle| + |\langle \widehat{\Sigma}_n - \Sigma, P_{S_1^\perp} \widehat{\Sigma}_n^\lambda P_{S_2^\perp} \rangle| \\
&\leq \|B\|_2 \|\widehat{\Sigma}_n^\lambda - A\|_2 + \frac{\lambda}{2} \|P_{S_1^\perp} \widehat{\Sigma}_n^\lambda P_{S_2^\perp}\|_1.
\end{aligned}
\tag{A.8}
$$

For any given matrix $M \in \mathbb{R}^{d \times d}$:

$$
M - P_{S_1^\perp} M P_{S_2^\perp} = P_{S_1} M + P_{S_1^\perp} M \left(I_d - P_{S_2^\perp} \right) = P_{S_1} M + P_{S_1^\perp} M P_{S_2}.
$$

Thus,

$$
\|M - P_{S_1^\perp} M P_{S_2^\perp}\|_2 \leq \sqrt{r} \left(\|P_{S_1} M\|_\infty + \|P_{S_1^\perp} M P_{S_2^\perp}\|_\infty \right) \leq 2\sqrt{r} \|M\|_\infty.
$$

This shows that $\|B\|_2 \leq \lambda \sqrt{r}$. Hence, by combining (A.6) – (A.8) we get

$$
\|\widehat{\Sigma}_n^\lambda - \Sigma\|_2^2 + \|\widehat{\Sigma}_n^\lambda - A\|_2^2 \leq \|A - \Sigma\|_2^2 + 3\lambda \sqrt{r} \|\widehat{\Sigma}_n^\lambda - A\|_2.
$$

By subtracting $\|\widehat{\Sigma}_n^\lambda - A\|_2^2$ on both sides and using the inequality $\alpha^2/4 \geq \alpha\beta - \beta^2$ with $\alpha = 3\lambda\sqrt{r}$ and $\beta = \|\widehat{\Sigma}_n^\lambda - A\|_2$, we arrive at the second term in the minimum (with $A = \Sigma$):

$$
\|\widehat{\Sigma}_n^\lambda - \Sigma\|_2^2 \leq \|A - \Sigma\|_2^2 + 3\lambda^2 r.
$$

$\square$

Proof of Theorem 2.3 We set $S_n = \sum_{k=1}^n X_k$ and $v_n = \sum_{k=1}^n C_k$. The subadditivity of the maximal eigenvalue function $\lambda_{\max}(\cdot)$ implies that

$$
\lambda_{\max}\left(\theta S_n - \frac{\theta^2}{2(1-\theta)} v_n I_d \right) \geq \theta x - \frac{\theta^2}{2(1-\theta)} v,
\tag{A.9}
$$

whenever $\lambda_{\max}(S_n) \geq x$, $v_n \leq v$ and $\theta \in (0,1)$. By exponentiating both sides of (A.9), applying the spectral mapping theorem and that $\lambda_{\max}(A) \leq \operatorname{tr}(A)$ for any positive semidefinite matrix A,

$$
\begin{aligned}
\mathbb{P}\left(\lambda_{\max}(S_n) \geq x, \, v_n \leq v \right) &\leq \mathbb{P}\left(Y_n \geq \exp\left(\theta x - \frac{\theta^2}{2(1-\theta)} v \right) \right) \\
&\leq \mathbb{E}[Y_n] \exp\left(-\theta x + \frac{\theta^2}{2(1-\theta)} v \right),
\end{aligned}
\tag{A.10}
$$

where $Y_n = \operatorname{tr}\left[\exp\left(\theta S_n - \frac{\theta^2}{2(1-\theta)} v_n I_d \right) \right]$. Suppose for the moment that $R = 1$, so that $\|\mathbb{E}[X_k^p \mid \mathcal{G}_{k-1}]\|_\infty \leq \frac{p!}{2} C_k$ for all k and integers $p \geq 2$. Since $\mathbb{E}[X_k \mid \mathcal{G}_{k-1}] = 0$, it then follows that

$$
\begin{aligned}
\left\| \mathbb{E}\big[\exp(\theta X_k) \mid \mathcal{G}_{k-1} \big] \right\|_\infty &\leq 1 + \sum_{p=2}^\infty \frac{\theta^p \|\mathbb{E}[X_k^p \mid \mathcal{G}_{k-1}]\|_\infty}{p!} \\
&\leq 1 + \frac{\theta^2}{2(1-\theta)} C_k \leq \exp\left(\frac{\theta^2}{2(1-\theta)} C_k \right).
\end{aligned}
$$

Thus, the matrix

$$\exp\left(\frac{\theta^2}{2(1-\theta)}C_k\right)I_d - \mathbb{E}\big[\exp(\theta X_k) \mid \mathcal{G}_{k-1}\big]$$

is positive semidefinite, and hence it follows by Tropp (2011, Lemma 2.1) that $(Y_k)_{k=1}^n$ is a positive supermartingale with $Y_0 = d$. By combining this observation with (A.10) and choosing $\theta = x/(x+\nu)$, we conclude that

$$\mathbb{P}\left(\lambda_{\max}(S_n) \geq x,\ \nu_n \leq \nu\right) \leq d \exp\left(-\frac{x^2}{2(x+\nu)}\right).$$

The general result can now be deduced by taking $\widetilde{X}_k = X_k/R$. $\qquad\square$

A key ingredient to verify the conditions of Theorem 2.3 and prove Theorem 2.4 below is a suitable version of the Burkholder–Davis–Gundy inequality. We shall employ the one given in (Seidler and Sobukawa 2003, Lemma 2.2), which is restated here for convenience. Note that, although their result applies to martingales with values in a general Hilbert space, we specialize the formulation here to the finite-dimensional setting.

Lemma A.1 (Seidler and Sobukawa (2003)) *Let $m \in \mathbb{N}$ and $T \in (0, \infty)$. For any constant $p \in [2, \infty)$ there exists $\gamma_p \in (0, \infty)$ such that*

$$\mathbb{E}\left[\sup_{t\in[0,T]} \left\|\int_0^t \varphi_s \mathrm{d}W_s\right\|_2^p\right] \leq \gamma_p^p \mathbb{E}\left[\left(\int_0^T \|\varphi_t\|_2^2 \mathrm{d}t\right)^{p/2}\right]$$

for all predictable processes $\varphi\colon \Omega \times [0, T] \to \mathbb{R}^{m\times d}$ with $\int_0^T \mathbb{E}\big[\|\varphi_t\|_2^p\big]\mathrm{d}t < \infty$. Moreover, one can always take

$$\gamma_p = \frac{4p}{p-1}\sqrt{p + \frac{1}{2}}. \tag{A.11}$$

With the choice in (A.11), γ_p is of smaller asymptotic order (as $p \to \infty$) than the constant associated with the usual Burkholder–Davis–Gundy inequalities available by application of Itô's formula (cf. the proof of Proposition 2.1 in Marinelli and Röckner (2016)).

Proof of Theorem 2.4 We decompose the log-price into the drift and volatility component $Y_t = Y_t^\mu + Y_t^\sigma$, where $Y_t^\mu = \int_0^t \mu_s \mathrm{d}s$ and $Y_t^\sigma = \int_0^t \sigma_s \mathrm{d}W_s$. By the triangle inequality,

$$
\begin{aligned}
\mathbb{P}\left(\|\widehat{\Sigma}_n - \Sigma\|_\infty > x\right) \leq &\mathbb{P}\left(\left\|\sum_{k=1}^{\lfloor\Delta_n^{-1}\rfloor}(\Delta_k^n Y^\mu)(\Delta_k^n Y^\mu)^\top\right\|_\infty > \frac{x}{5}\right) \\
&+ 2\mathbb{P}\left(\left\|\sum_{k=1}^{\lfloor\Delta_n^{-1}\rfloor}(\Delta_k^n Y^\mu)(\Delta_k^n Y^\sigma)^\top\right\|_\infty > \frac{x}{5}\right) \\
&+ \mathbb{P}\left(\left\|\sum_{k=1}^{\lfloor\Delta_n^{-1}\rfloor}(\Delta_k^n Y^\sigma)(\Delta_k^n Y^\sigma)^\top - \int_0^{\lfloor\Delta_n^{-1}\rfloor\Delta_n} c_t \mathrm{d}t\right\|_\infty > \frac{x}{5}\right) \\
&+ \mathbb{P}\left(\left\|\int_{\lfloor\Delta_n^{-1}\rfloor\Delta_n}^1 c_t \mathrm{d}t\right\|_\infty > \frac{x}{5}\right) \\
\equiv &a_1 + 2a_2 + a_3 + a_4.
\end{aligned}
$$

$$\tag{A.12}$$

Jensen's inequality implies that for the first term in (A.12),

$$a_1 \leq \mathbb{P}\left(\sum_{k=1}^{\lfloor \Delta_n^{-1} \rfloor} \|\Delta_k^n Y^\mu\|_2^2 > \frac{x}{5} \right) \leq \mathbb{1}_{\{x < 5\nu_\mu \Delta_n\}}. \tag{A.13}$$

The fourth term can be bounded as:

$$a_4 \leq \mathbb{1}_{\{x < 5\nu_{c,\infty} \Delta_n\}}. \tag{A.14}$$

To handle the third term in (A.12), we note it has the form $a_3 = \sum_{k=1}^{\lfloor \Delta_n^{-1} \rfloor} (A_k - B_k)$, where

$$A_k = (\Delta_k^n Y^\sigma)(\Delta_k^n Y^\sigma)^\top, \quad B_k = \int_{(k-1)\Delta_n}^{k\Delta_n} c_t \mathrm{d}t,$$

and that $(A_k - B_k)_{k=1}^{\lfloor \Delta_n^{-1} \rfloor}$ is a martingale difference sequence with respect to the filtration $(\mathcal{F}_{k\Delta_n})_{k=0}^{\lfloor \Delta_n^{-1} \rfloor}$. Thus, we can apply Theorem 2.3 if we can control

$$M_{p,k} \equiv \left\| \mathbb{E}[(A_k - B_k)^p \mid \mathcal{F}_{(k-1)\Delta_n}] \right\|_\infty,$$

for integer $p \geq 2$. First note that, since $\|B_k\|_\infty \leq \nu_{c,\infty} \Delta_n$, the sub-multiplicativity of $\|\cdot\|_\infty$ and the binomial theorem imply that

$$M_{p,k} \leq \left\| \mathbb{E}[A_k^p \mid \mathcal{F}_{(k-1)\Delta_n}] \right\|_\infty + \sum_{i=1}^{p} \binom{p}{i} \mathbb{E}\left[\|A_k\|_\infty^{p-i} \mid \mathcal{F}_{(k-1)\Delta_n} \right] (\nu_{c,\infty} \Delta_n)^i. \tag{A.15}$$

We start by analyzing the second term on the right-hand side of (A.15). For an arbitrary integer $m \geq 1$ and any $\mathcal{F}_{(k-1)\Delta_n}$-measurable set A, it follows from Lemma A.1 that

$$\mathbb{E}\left[\|A_k\|_\infty^m \mathbb{1}_A \right] = \mathbb{E}\left[\left\| \int_{(k-1)\Delta_n}^{k\Delta_n} \sigma_t \mathbb{1}_A \mathrm{d}W_t \right\|_2^{2m} \right] \leq \gamma_{2m}^{2m} \mathbb{E}\left[\left(\int_{(k-1)\Delta_n}^{k\Delta_n} \mathrm{tr}(c_t)\mathrm{d}t \right)^m \mathbb{1}_A \right]$$

and, thus,

$$\mathbb{E}[\|A_k\|_\infty^m \mid \mathcal{F}_{(k-1)\Delta_n}]^{1/m} \leq \alpha m \nu_{c,2} \Delta_n \tag{A.16}$$

for a suitable absolute constant $\alpha \geq 1$. In particular, (A.16) shows that $\mathbb{E}\left[\|A_k\|_\infty^{p-i} \mid \mathcal{F}_{(k-1)\Delta_n} \right] \leq (\alpha p \nu_{c,2} \Delta_n)^{p-i}$ for $i = 1, \ldots, p$. Together with the mean value theorem and Stirling's formula, we can therefore establish that

$$\sum_{i=1}^{p} \binom{p}{i} \mathbb{E}\left[\|A_k\|_\infty^{p-i} \mid \mathcal{F}_{(k-1)\Delta_n} \right] (\nu_{c,\infty} \Delta_n)^i \leq (\alpha p \nu_{c,2} \Delta_n)^p \sum_{i=1}^{p} \binom{p}{i} \left(\frac{\nu_{c,\infty}}{\alpha p \nu_{c,2}} \right)^i$$

$$\leq \frac{p!}{2} (e\alpha \nu_{c,2} \Delta_n)^p \left[\left(1 + \frac{\nu_{c,\infty}}{\alpha p \nu_{c,2}} \right)^p - 1 \right]$$

$$\leq \frac{p!}{2} (2e\alpha \nu_{c,2} \Delta_n)^p \frac{\nu_{c,\infty}}{\nu_{c,2}}. \tag{A.17}$$

Now, look at the first term on the right-hand side of (A.15). For a fixed $u \in \mathbb{R}^d$ with $\|u\|_2 = 1$, it follows from the Cauchy–Schwarz inequality that

$$u^\top \mathbb{E}\left[A_k^p \mid \mathcal{F}_{(k-1)\Delta_n} \right] u = \mathbb{E}\left[\|A_k\|_\infty^{p-1} (u^\top (\Delta_k^n Y))^2 \mid \mathcal{F}_{(k-1)\Delta_n} \right]$$

$$\leq \mathbb{E}\left[\|A_k\|_\infty^{2(p-1)} \mid \mathcal{F}_{(k-1)\Delta_n} \right]^{1/2} \mathbb{E}\left[(u^\top \Delta_k^n Y)^4 \mid \mathcal{F}_{(k-1)\Delta_n} \right]^{1/2}. \tag{A.18}$$

The first term in (A.18) is handled by (A.16) with $m = 2(p-1)$:

$$\mathbb{E}\big[\|A_k\|_\infty^{2(p-1)} \mid \mathcal{F}_{(k-1)\Delta_n}\big]^{1/2} \le (2\alpha(p-1)v_{c,2}\Delta_n)^{p-1} \le \frac{p!}{2}\frac{(2e\alpha v_{c,2}\Delta_n)^p}{v_{c,2}\Delta_n}. \quad (A.19)$$

The second term is treated as:

$$\mathbb{E}\big[(u^\top \Delta_k^n Y)^4 \mid \mathcal{F}_{(k-1)\Delta_n}\big]^{1/2} \le \gamma_4^2 \mathbb{E}\bigg[\bigg(\int_{(k-1)\Delta_n}^{k\Delta_n} u^\top c_t u \, \mathrm{d}\bigg)^2 \mid \mathcal{F}_{(k-1)\Delta_n}\bigg]^{1/2} \quad (A.20)$$

$$\le \gamma_4^2 v_{c,\infty}\Delta_n,$$

where Lemma A.1 is used here for the conditional expectation. Consequently, by combining (A.18) – (A.20), we get the estimate

$$u^\top \mathbb{E}\big[A_k^p \mid \mathcal{F}_{(k-1)\Delta_n}\big]u \le \gamma_4^2 \frac{p!}{2}(2e\alpha v_{c,2}\Delta_n)^p \frac{v_{c,\infty}}{v_{c,2}}. \quad (A.21)$$

Since this analysis holds for an arbitrary unit vector u, the right-hand side of (A.21) is an upper bound for $\big\|\mathbb{E}[A_k^p \mid \mathcal{F}_{(k-1)\Delta_n}]\big\|_\infty$. This fact, together with (A.15) and (A.17), shows that

$$M_{p,k} \le \frac{p!}{2}(\tilde{\alpha}v_{c,2}\Delta_n)^p \frac{v_{c,\infty}}{v_{c,2}} \quad (A.22)$$

for a suitable absolute constant $\tilde{\alpha} \ge 2$. Now, it follows that Theorem 2.3 is applicable with $R = \tilde{\alpha}v_{c,2}\Delta_n$ and $C_k = C_1 = v_{c,\infty}/v_{c,2}$. Since

$$vR = \tilde{\alpha}v_{c,\infty}\lfloor \Delta_n^{-1}\rfloor\Delta_n \ge \tilde{\alpha}v_{c,\infty}(1-\Delta_n) \ge v_{c,\infty}$$

and $4vR^2 \le 4\tilde{\alpha}^2 v_{c,2}v_{c,\infty}\Delta_n$, the inequality (2.7) implies that

$$\mathbb{P}\left(\Big\|\sum_{k=1}^{\lfloor \Delta_n^{-1}\rfloor}(A_k - B_k)\Big\|_\infty > y\right) \le \tau(y) \equiv 2d\exp\left(-\frac{y}{4\tilde{\alpha}^2 v_{c,2}v_{c,\infty}\Delta_n}\min\{y, v_{c,\infty}\}\right). \quad (A.23)$$

In particular, (A.23) shows that

$$a_3 \le \tau\left(\frac{x}{5}\right). \quad (A.24)$$

Finally, for the second term in (A.12) we invoke the Cauchy–Schwarz inequality and consider an x for which $\frac{x^2}{25v_\mu\Delta_n} - v_{c,2} \ge \frac{v_{c,2}}{5v_{c,\infty}}x$ to deduce the following initial bound:

$$a_2 \le \mathbb{P}\left(\left(\sum_{k=1}^{\lfloor \Delta_n^{-1}\rfloor}\|\Delta_k^n Y^\mu\|_2^2\right)\left(\sum_{k=1}^{\lfloor \Delta_n^{-1}\rfloor}\|\Delta_k^n Y^\sigma\|_2^2\right) > \left(\frac{x}{5}\right)^2\right)$$

$$\le \mathbb{P}\left(\sum_{k=1}^{\lfloor \Delta_n^{-1}\rfloor}\|\Delta_k^n Y^\sigma\|_2^2 > \frac{x^2}{25v_\mu\Delta_n}\right) \quad (A.25)$$

$$\le \mathbb{P}\left(\sum_{k=1}^{\lfloor \Delta_n^{-1}\rfloor}\left(\|\Delta_k^n Y^\sigma\|_2^2 - \int_{(k-1)\Delta_n}^{k\Delta_n}\mathrm{tr}(c_t)\mathrm{d}t\right) > \frac{x^2}{25v_\mu\Delta_n} - v_{c,2}\right)$$

$$\le \mathbb{P}\left(\sum_{k=1}^{\lfloor \Delta_n^{-1}\rfloor}\left(\|\Delta_k^n Y^\sigma\|_2^2 - \int_{(k-1)\Delta_n}^{k\Delta_n}\mathrm{tr}(c_t)\mathrm{d}t\right) > \frac{v_{c,2}}{5v_{c,\infty}}x\right).$$

Next, observe that

$$X_k = \|\Delta_k^n Y^\sigma\|_2^2 - \int_{(k-1)\Delta_n}^{k\Delta_n} \mathrm{tr}(c_t)\mathrm{d}t, \qquad k = 1, \dots, \lfloor \Delta_n^{-1} \rfloor,$$

is a martingale difference sequence. By (A.16), we see that

$$\mathbb{E}\big[|X_k|^p \mid \mathcal{F}_{(k-1)\Delta_n}\big] \le \frac{p!}{2}(\alpha v_{c,2}\Delta_n)^p,$$

for a suitable absolute constant $\alpha \ge 2$, and thus (2.7) ensures that

$$\mathbb{P}\left(\Big|\sum_{k=1}^n X_k\Big| > y\right) \le \tau\left(\frac{v_{c,\infty}}{v_{c,2}}y\right),$$

where we assume $\tilde{\alpha}$ in τ is larger than α. This estimate together with (A.25) means that

$$a_2 \le \tau\left(\frac{x}{5}\right). \tag{A.26}$$

By combining (A.13), (A.14), (A.24), and (A.26) we conclude that

$$\mathbb{P}\big(\|\widehat{\Sigma}_n - \Sigma\|_\infty > x\big) \le 3\tau\left(\frac{x}{5}\right) \le 6d \exp\left\{-\frac{x}{100\tilde{\alpha}^2 v_{c,2}v_{c,\infty}\Delta_n}\min\{x, v_{c,\infty}\}\right\} \tag{A.27}$$

if $x \ge 5\Delta_n \max\{v_\mu, v_{c,\infty}\}$ and $\frac{x^2}{25v_\mu\Delta_n} - v_{c,2} \ge \frac{v_{c,2}}{5v_{c,\infty}}x$. In particular, (A.27) holds whenever $x \ge \gamma \max\left\{\frac{v_{c,2}v_\mu\Delta_n}{v_{c,\infty}}, \sqrt{v_{c,2}v_\mu\Delta_n}, v_{c,\infty}\Delta_n\right\}$, where γ is a sufficiently large absolute constant.   □

Note that, in the above calculations, one has to be careful in order to achieve optimal rates. Indeed, instead of the estimate (A.22), it might have been more natural to rely on the following (seemingly harmless) sequence of inequalities using (A.16) and Stirling's formula:

$$M_{p,k} \le 2^p \big(\mathbb{E}\big[\|A_k\|_\infty^p \mid \mathcal{F}_{(k-1)\Delta_n}\big] + \mathbb{E}\big[\|B_k\|_\infty^p \mid \mathcal{F}_{(k-1)\Delta_n}\big]\big) \le p!(2e\alpha v_{c,2}\Delta_n)^p. \tag{6.28}$$

However, with notation as in Theorem 2.3, this would result in $vR \asymp v_{c,2}$, while (A.22) satisfies $vR \asymp v_{c,\infty}$. Since vR determines the rate and $v_{c,\infty}$ can be of strictly smaller order than $v_{c,2}$ (see the discussion in the end of Sect. 2.1), this implies that a concentration inequality based on (6.28) cannot be optimal.

Proof of Theorem 2.5 Consider a fixed $\tau \in (0, \infty)$. Theorem 2.4 implies the existence of an absolute constant α such that

$$\mathbb{P}(2\|\widehat{\Sigma}_n - \Sigma\|_\infty > \lambda) \le 6d \exp\left(-\frac{\lambda}{\alpha v_{c,2}v_{c,\infty}\Delta_n}\min\{\lambda, v_{c,\infty}\}\right)$$

as long as $\lambda \ge \alpha \max\left\{\frac{v_{c,2}v_\mu\Delta_n}{v_{c,\infty}}, \sqrt{v_{c,2}v_\mu\Delta_n}, v_{c,\infty}\Delta_n\right\}$. Thus, under this restriction on λ it follows that $2\|\widehat{\Sigma}_n - \Sigma\|_\infty \le \lambda$ with probability at least $1 - e^{-\tau}$ if

$$6d \exp\left(-\frac{\lambda}{\alpha v_{c,2}v_{c,\infty}\Delta_n}\min\{\lambda, v_{c,\infty}\}\right) \le \exp(-\tau)$$

or, in particular, if

$$\lambda \ge \gamma \max\left\{\sqrt{v_{c,2}v_{c,\infty}\Delta_n(\log(d) + \tau)}, v_{c,2}\Delta_n(\log(d) + \tau), \frac{v_{c,2}v_\mu\Delta_n}{v_{c,\infty}}\right\} \tag{A.29}$$

for a suitably chosen absolute constant γ. In view of Proposition 2.2, the proof is complete.
                                                                                                    □

Proof of Theorem 2.7 Under $\mathbb{P}_A$, we have that $Z_k = \Delta_k^n Y / \sqrt{\Delta_n}$, $k = 1, \ldots, \lfloor \Delta_n^{-1} \rfloor$, are i.i.d. Gaussian random vectors with covariance matrix A. Consequently, since $\lfloor \Delta_n^{-1} \rfloor \geq r^2$, (Lounici 2014, Theorem 2) implies that

$$\inf_{\widetilde{\Sigma}} \sup_{A \in \mathcal{C}_r} \mathbb{P}_A \left(\| \widetilde{\Sigma} - A \|_2^2 > \underline{\gamma} \frac{\|A\|_\infty^2 r_e(A) \operatorname{rank}(A)}{\lfloor \Delta_n^{-1} \rfloor} \right) \geq \beta \tag{A.30}$$

for suitable absolute constants $\beta \in (0, 1)$ and $\underline{\gamma} \in (0, \infty)$, and this proves the result. $\qquad\square$

Proof of Theorem 2.8 Set $\hat{r} = \operatorname{rank}(\widehat{\Sigma}_n^\lambda)$. To show the upper bound in (2.18) it suffices to argue that Σ has at least $\hat{r}$ singular values which are larger than $(\lambda - \bar{\lambda})/2$ or, equivalently, $s_{\hat{r}}(\Sigma) \geq (\lambda - \bar{\lambda})/2$. Observe first that the representation (2.2) implies $s_{\hat{r}}(\widehat{\Sigma}_n) > \lambda/2$ (recall that $s_k(A)$ refers to the kth largest singular value of A). By using this inequality together with the fact that $A \mapsto s_k(A)$ is Lipschitz continuous with constant 1 with respect to the spectral norm, we establish that

$$s_{\hat{r}}(\Sigma) \geq s_{\hat{r}}(\widehat{\Sigma}_n) - |s_{\hat{r}}(\widehat{\Sigma}_n) - s_{\hat{r}}(\Sigma)| \geq \frac{\lambda - \bar{\lambda}}{2}.$$

In order to prove the lower bound of (2.18), we need to show that $s_{r_\lambda}(\widehat{\Sigma}_n^\lambda) > 0$ with $r_\lambda = \operatorname{rank}(\Sigma; \lambda)$. However, this follows immediately from the following sequence of inequalities:

$$s_{r_\lambda}(\widehat{\Sigma}_n^\lambda) \geq s_{r_\lambda}(\widehat{\Sigma}_n) - \frac{\lambda}{2} \geq s_{r_\lambda}(\Sigma) - \| \widehat{\Sigma}_n - \Sigma \|_\infty - \frac{\lambda}{2} \geq \frac{\lambda - \bar{\lambda}}{2} > 0.$$

The last part of the result is a direct consequence of the inequality (2.18), since $\operatorname{rank}(\Sigma; x) = \operatorname{rank}(\Sigma)$ for $x \in (0, s]$. This finishes the proof. $\qquad\square$

Proof of Corollary 2.9 For an arbitrary number $\bar{\lambda} \in (0, \infty)$, Proposition 2.2 and Theorem 2.8 imply that both

$$\| \widehat{\Sigma}_n^\lambda - \Sigma \|_2^2 \leq 3\lambda^2 \operatorname{rank}(\Sigma) \tag{A.31}$$

and

$$\operatorname{rank}(\Sigma; \lambda) \leq \operatorname{rank}(\widehat{\Sigma}_n^\lambda) \leq \operatorname{rank}(\Sigma; \tfrac{1}{2}(\lambda - \bar{\lambda})) \tag{A.32}$$

on the set $A(\bar{\lambda}) \equiv \{2\| \widehat{\Sigma}_n - \Sigma \|_\infty \leq \bar{\lambda}\}$ whenever $\lambda > \bar{\lambda}$. By the proof of Theorem 2.5, in particular (A.29) together with the fact that $v_\mu \leq v_{c,\infty}$, it follows that for any $\tau \in (0, \infty)$, the set $A(\bar{\lambda})$ has probability at least $1 - \exp(-\tau)$ if

$$\bar{\lambda} = \bar{\gamma} \max \left\{ \sqrt{v_{c,2} v_{c,\infty} \Delta_n (\log(d) + \tau)}, v_{c,2} \Delta_n (\log(d) + \tau) \right\} \tag{A.33}$$

for a sufficiently large absolute constant $\bar{\gamma}$. By considering $\tau = \log(d)$ and using that $\Delta_n^{-1} \geq 2\frac{v_{c,2}}{v_{c,\infty}} \log(d)$, the maximum in (A.33) equals its first term. With this choice of τ, the set $A(\bar{\lambda})$ has probability at least $1 - d^{-1}$, and $(\lambda - \bar{\lambda})/2 \geq \delta\lambda$ when λ is given by (2.19) as long as we choose $\gamma \geq \bar{\gamma}\sqrt{2}/(1 - 2\delta)$. Consequently, by plugging the specific values of λ and $\bar{\lambda}$ into (A.31) and (A.32), we have established the first part of the result. The last part of the result follows immediately from Theorem 2.8. $\qquad\square$

Proof of Theorem 3.1 First, let $\delta_n \in \{0, 1\}$ be defined such that

$$\lfloor (t + h_n) \Delta_n^{-1} \rfloor - \lfloor t \Delta_n^{-1} \rfloor = \lfloor (h_n + \delta_n \Delta_n) \Delta_n^{-1} \rfloor, \tag{A.34}$$

and set $\tau(s) = \varepsilon s + \lfloor t \Delta_n^{-1} \rfloor \Delta_n$ with $\varepsilon = h_n + \delta_n \Delta_n$. Clearly, $\bar{Y}_s = Y_{\tau(s)}$, $s \in [0, 1]$, is a diffusion with drift $\bar{\mu}_s = \varepsilon \mu_{\tau(s)}$, volatility $\bar{\sigma}_s = \sqrt{\varepsilon} \sigma_{\tau(s)}$, and driving (standard) Brownian

motion $\bar{W}_s = (W_{\tau(s)} - W_{\tau(0)})/\sqrt{\varepsilon}$. Moreover, its RV at time 1 based on a sampling frequency of $\bar{\Delta}_n = \Delta_n/\varepsilon$ coincides with $\widehat{\Sigma}_n(t; t + h_n)$:

$$\widehat{\Sigma}_n(t; t + h_n) = \sum_{k=1}^{\lfloor \bar{\Delta}_n^{-1} \rfloor} (\bar{\Delta}_k^n \bar{Y})(\bar{\Delta}_k^n \bar{Y})^\top, \qquad \bar{\Delta}_k^n \bar{Y} = \bar{Y}_{k\bar{\Delta}_n} - \bar{Y}_{(k-1)\bar{\Delta}_n}.$$

Hence, the first step is to apply Corollary 2.6 to the process $(\bar{Y}_s)_{s \in [0,1]}$ with sampling frequency $\bar{\Delta}_n$. To this end, note that

$$\sup_{s \leq 1} \|\bar{\mu}_s\|_2^2 \leq \varepsilon \nu_\mu, \quad \sup_{s \leq 1} \mathrm{tr}(\bar{\sigma}_s \bar{\sigma}_s^\top) \leq \varepsilon \nu_{c,2}, \quad \text{and} \quad \sup_{s \leq 1} \|\bar{\sigma}_s \bar{\sigma}_s^\top\|_\infty \leq \varepsilon \nu_{c,\infty}.$$

Since we also have $\varepsilon \in [h_n, 2h_n]$ and $h_n/\Delta \geq 2 \frac{\nu_{c,2}}{\nu_{c,\infty}} \log(d)$, it follows that the assumptions of Corollary 2.6 are satisfied and we deduce that

$$\left\| \widehat{\Sigma}_n^\lambda(t; t + h_n) - \bar{\Sigma} \right\|_2^2 \leq 3\gamma^2 \nu_{c,2} \nu_{c,\infty} h_n \Delta_n \, \mathrm{rank}(\bar{\Sigma}) \log(d) \tag{A.35}$$

with probability at least $1 - d^{-1}$ when λ meets (3.6) and γ is a sufficiently large absolute constant. Here $\bar{\Sigma}$ denotes the QV of $(\bar{Y}_s)_{s \in [0,1]}$ at time 1. By dividing both sides of the inequality (A.35) by h_n^2,

$$\left\| \hat{c}_n^\lambda(t) - \frac{\bar{\Sigma}}{h_n} \right\|_2^2 \leq 3\gamma^2 \frac{\nu_{c,2} \nu_{c,\infty} \Delta_n \, \mathrm{rank}(\bar{\Sigma}) \log(d)}{h_n}. \tag{A.36}$$

The error $\|h_n^{-1}\bar{\Sigma} - \varepsilon^{-1}\bar{\Sigma}\|_2$ can be bounded in the following way using that $h_n/\Delta_n \geq \frac{\nu_{c,2}}{2\nu_{c,\infty}}$:

$$\left\| \frac{\bar{\Sigma}}{h_n} - \frac{\bar{\Sigma}}{\varepsilon} \right\|_2^2 \leq \left(\frac{\nu_{c,2} \Delta_n}{h_n} \right)^2 \leq \frac{2\nu_{c,2} \nu_{c,\infty} \Delta_n}{h_n}. \tag{A.37}$$

Now, for any given $\beta \in (1, \infty)$, we want to argue that $\|\varepsilon^{-1}\bar{\Sigma} - c_t\|_2$ is small with probability $1 - \beta^{-1}$. To do so, note initially that $t \in A_n \equiv [\lfloor t\Delta_n^{-1} \rfloor \Delta_n, \lfloor t\Delta_n^{-1} \rfloor \Delta_n + \varepsilon]$ and that, for any $s, u \in A_n$, we have $\nu_{c,\psi} \geq \|c_s - c_u\|_\psi/\sqrt{2h_n}$. Using these two facts, and by relying on Jensen's and Markov's inequality, we do the following computations for an arbitrary $x > 0$:

$$\mathbb{P}\left(\left\| \frac{\bar{\Sigma}}{\varepsilon} - c_t \right\|_2 > x \right) \leq \mathbb{P}\left(\frac{1}{\varepsilon} \int_{A_n} \|c_s - c_t\|_2 ds > x \right)$$

$$\leq \mathbb{P}\left(\frac{1}{\varepsilon} \int_{A_n} \psi\left(\frac{\|c_s - c_t\|_2}{\|c_s - c_t\|_\psi} \right) ds > \psi\left(\frac{x}{\nu_{c,\psi}\sqrt{2h_n}} \right) \right).$$

$$\leq \psi\left(\frac{x}{\nu_{c,\psi}\sqrt{2h_n}} \right)^{-1}.$$

It follows that, with probability at least $1 - \beta^{-1}$,

$$\left\| \frac{\bar{\Sigma}}{\varepsilon} - c_t \right\|_2^2 \leq 2h_n \nu_{c,\psi}^2 \psi^{-1}(\beta)^2. \tag{A.38}$$

By choosing $\beta = \psi(\sqrt{\log(d)})$ and combining (A.36) – (A.38) we conclude that, with probability at least $1 - d^{-1} - \psi(\sqrt{\log(d)})^{-1}$,

$$\|\hat{c}_n^\lambda(t) - c_t\|_2^2 \leq \kappa\gamma^2 \left(\frac{\nu_{c,2} \nu_{c,\infty} \Delta_n \, \mathrm{rank}(\bar{\Sigma})}{h_n} + h_n \nu_{c,\psi}^2 \right) \log(d)$$

for a suitably chosen absolute constant κ. Since $A_n \subseteq [t - \Delta_n, t + h_n + \Delta_n]$ and $h_n \geq 2\Delta_n$, it follows that $\mathrm{rank}(\bar{\Sigma}) \leq \mathrm{rank}(\Sigma(t - \frac{h_n}{2}; t + \frac{3h_n}{2}))$, and the proof is complete. $\qquad \square$

Proof of Theorem 3.2 Similarly to the proof of Theorem 3.1, we consider the time-changed process $\bar{Y}_s = Y_{\tau(s)}$, $s \in [0, 1]$, and apply Corollary 2.9 to deduce that, with probability at least $1 - d^{-1}$,

$$\left\| \hat{c}_n^\lambda(t) - \frac{\bar{\Sigma}}{h_n} \right\|_2^2 \leq 3\gamma^2 \frac{v_{c,2} v_{c,\infty} \Delta_n \, \mathrm{rank}(\bar{\Sigma}) \log(d)}{h_n}$$

and

$$\mathrm{rank}(\bar{\Sigma}; \lambda) \leq \mathrm{rank}(\hat{c}_n^\lambda(t)) \leq \mathrm{rank}(\bar{\Sigma}; 2\delta\lambda). \tag{A.39}$$

Here we recall that $\tau(s) = \varepsilon s + \lfloor t \Delta_n^{-1} \rfloor \Delta_n$ and $\varepsilon = h_n + \delta_n \Delta_n$, where $\delta_n \in \{0, 1\}$ is chosen such that (A.34) holds. By following the exact same arguments as in the proof of Theorem 3.1 we obtain the estimate

$$\| \hat{c}_n^\lambda(t) - c_t \|_2^2 \leq \kappa \gamma^2 \left(\frac{v_{c,2} v_{c,\infty} \Delta_n \, \mathrm{rank}(\bar{\Sigma})}{h_n} + h_n v_{c,\psi}^2 \right) \log(d),$$

which applies with probability at least $1 - d^{-1} - \psi(\sqrt{\log(d)})^{-1}$ for a suitably chosen absolute constant κ. The inequality (3.9) follows immediately from the fact that $\mathrm{rank}(\bar{\Sigma}) \leq \mathrm{rank}(\Sigma(t - \frac{h_n}{2}; t + \frac{3h_n}{2}))$. To establish the bounds (3.10) on $\mathrm{rank}(\hat{c}_n^\lambda(t))$ note initially that, from the proof of Theorem 3.1 (particularly, (A.38)), we may in fact assume that

$$\left\| \frac{\bar{\Sigma}}{\varepsilon} - c_t \right\|_2^2 \leq 2 h_n v_{c,\psi}^2 \log(d)$$

on the event that we are considering. Consequently, since singular values are Lipschitz continuous with constant 1 with respect to $\| \cdot \|_2$,

$$\left| s_k \left(\frac{\bar{\Sigma}}{\varepsilon} \right) - s_k(c_t) \right| \leq v_{c,\psi} \sqrt{2 h_n \log(d)}$$

for $k = 1, \ldots, d$. By using this observation, the fact that $\varepsilon \in [h_n, 2h_n]$, and the explicit expression for λ the following two implications can be deduced:

$$s_k(\bar{\Sigma}) \geq 2\delta\lambda \quad \Longrightarrow \quad s_k(c_t) \geq \delta\gamma \left(\sqrt{\frac{v_{c,2} v_{c,\infty} \Delta_n}{h_n}} - v_{c,\psi} \sqrt{h_n} \right) \sqrt{\log(d)},$$

$$s_k(\bar{\Sigma}) < \lambda \quad \Longrightarrow \quad s_k(c_t) < \gamma \left(\sqrt{\frac{v_{c,2} v_{c,\infty} \Delta_n}{h_n}} + v_{c,\psi} \sqrt{h_n} \right) \sqrt{\log(d)}.$$

We have also imposed the innocent assumption that γ is chosen such that $\delta\gamma \geq \sqrt{2}$.

From these two implications we conclude that $\mathrm{rank}(\bar{\Sigma}; 2\delta\lambda) \leq \mathrm{rank}(c_t; \underline{\varepsilon})$ and $\mathrm{rank}(\bar{\Sigma}; \lambda) \geq \mathrm{rank}(c_t; \overline{\varepsilon})$, which (in view of (A.39)) establishes (3.10). The last statement in the result is a direct consequence of (3.10), and thus the proof is complete. $\qquad \square$

References

Aït-Sahalia Y, Xiu D (2017) Using principal component analysis to estimate a high dimensional factor model with high-frequency data. J Econ 201(2):384–399

Aït-Sahalia Y, Xiu D (2019) Principal component analysis of high-frequency data. J Am Stat Assoc 114(525):287–303

Andersen TG, Bollerslev T (1998) Answering the skeptics: yes, standard volatility models do provide accurate forecasts. Int Econ Rev 39(4):885–905

Andersen TG, Bollerslev T, Diebold FX, Labys P (2003) Modeling and forecasting realized volatility. Econometrica 71(2):579–625

Argyriou A, Evgeniou T, Pontil M (2008) Convex multi-task feature learning. Mach Learn 73(3):243–272

Bach FR (2008) Consistency of trace norm minimization. J Mach Learn Res 9(35):1019–1048

Barndorff-Nielsen OE, Graversen SE, Jacod J, Podolskij M, Shephard N (2006) A central limit theorem for realized power and bipower variations of continuous semimartingales. In: Kabanov Y, Lipster R, Stoyanov J (eds) From stochastic calculus to mathematical finance: the shiryaev festschrift. Springer, Heidelberg, pp 33–68

Barndorff-Nielsen OE, Shephard N (2002) Econometric analysis of realized volatility and its use in estimating stochastic volatility models. J Roy Stat Soc B 64(2):253–280

Barndorff-Nielsen OE, Shephard N (2004) Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. Econometrica 72(3):885–925

Cai TT, Hu J, Li Y, Zheng X (2020) High-dimensional minimum variance portfolio estimation based on high-frequency data. J Econ 214(2):482–494

Candès EJ, Recht B (2009) Exact matrix completion via convex optimization. Found Comput Math 9(6):717–772

Christensen K, Podolskij M, Thamrongrat N, Veliyev B (2017) Inference from high-frequency data: a subsampling approach. J Econ 197(2):245–272

Clarke FH (1990) Optimization and Nonsmooth Analysis. Society for Industrial and Applied Mathematics, Philadelphia, 1st edn

Delbaen F, Schachermayer W (1994) A general version of the fundamental theorem of asset pricing. Math Ann 300(1):463–520

Diop A, Jacod J, Todorov V (2013) Central limit theorems for approximate quadratic variations of pure jump Itô semimartingales. Stoch Process Appl 123(3):839–886

Fissler T, Podolskij M (2017) Testing the maximal rank of the volatility process for continuous diffusions observed with noise. Bernoulli 23(4B):3021–3066

Hautsch N, Kyj LM, Oomen RCA (2012) A blocking and regularization approach to high dimensional realized covariance estimation. J Appl Econom 27(4):625–645

Heiny J, Podolskij M (2020) "On estimation of quadratic variation for multivariate pure jump semimartingales," preprint arXiv:2009.02786

Heston SL (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev Financ Stud 6(2):327–343

Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12(1):55–67

Jacod J (1994) "Limit of random measures associated with the increments of a Brownian semimartingale," Preprint number 120, Laboratoire de Probabilitiés, Université Pierre et Marie Curie, Paris

Jacod J (2008) Asymptotic properties of realized power variations and related functionals of semimartingales. Stoch Process Appl 118(4):517–559

Jacod J, Lejay A, Talay D (2008) Estimation of the Brownian dimension of a continuous Itô process. Bernoulli 14(2):469–498

Jacod J, Podolskij M (2013) A test for the rank of the volatility process: the random perturbation approach. Ann Stat 41(5):2391–2427

Jacod J, Podolskij M (2018) On the minimal number of driving Lévy motions in a multivariate price model. J Appl Probab 55(3):823–833

Jacod J, Protter PE (2012) Discretization of processes, 2nd edn. Springer, Berlin

Kalnina I (2011) Subsampling high frequency data. J Econom 161(2):262–283

Koltchinskii V, Lounici K, Tsybakov AB (2011) Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. Ann Stat 39(5):2302–2329

Kong X-B (2017) On the number of common factors underlying large panel high-frequency data. Biometrika 104(2):397–410

Kong X-B (2020) A random-perturbation-based rank estimator of the number of factors. Biometrika 107(2):505–511

Lounici K (2014) High-dimensional covariance matrix estimation with missing observations. Bernoulli 20(3):1029–1058

Lunde A, Shephard N, Sheppard K (2016) Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. J Bus Econ Stat 34(4):504–518

Mancini C (2009) Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. Scand J Stat 36(2):270–296

Marinelli C, Röckner M (2016) On the maximal inequalities of Burkholder, Davis and Gundy. Expo Math 34(1):1–26

Minsker S (2017) On some extensions of Bernstein's inequality for self-adjoint operators. Stat Probab Lett 127(1):111–119

Negahban S, Wainwright MJ (2011) Estimation of (near) low-rank matrices with noise and high-dimensional scaling. Ann Stat 39(2):1069–1097

Pelger M (2019) Large-dimensional factor modeling based on high-frequency observations. J Econom 208(1):23–42

Politis DN, Romano JP, Wolf M (1999) Subsampling, 1st edn. Springer, Berlin

Recht B, Fazel M, Parrilo PA (2010) Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev 52(3):471–501

Reiss M, Todorov V, Tauchen G (2015) Nonparametric test for a constant beta between Itô semi-martingales based on high-frequency data. Stoch Process Appl 125(8):2955–2988

Ross SA (1976) The arbitrage theory of capital asset pricing. J Econ Theory 13(3):341–360

Seidler J, Sobukawa T (2003) Exponential integrability of stochastic convolutions. J Lond Math Soc 67(1):245–258

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J Roy Stat Soc B 58(1):267–288

Tropp J (2011) Freedman's inequality for matrix martingales. Electron Commun Probab 16(1):262–270

Tropp JA (2012) User-friendly tail bounds for sums of random matrices. Found Comput Math 12(4):389–434

Tropp JA (2015) An introduction to matrix concentration inequalities.' Foundations and Trends® in Machine Learning 8(1–2):1–230

Vershynin R (2010) Introduction to the non-asymptotic analysis of random matrices. In: Eldar YC, Kutyniok G (eds) Compressed sensing: theory and applications. Cambridge University Press, Cambridge, pp 210–268

Wang Y, Zou J (2010) Vast volatility matrix estimation for high-frequency financial data. Ann Stat 38(2):943–978

Watson GA (1992) Characterization of the subdifferential of some matrix norms. Linear Algebra Appl 170(1):33–45

Zheng X, Li Y (2011) On the estimation of integrated covariance matrices of high dimensional diffusion processes. Ann Stat 39(6):3121–3151