A restricted eigenvalue condition for unit-root non-stationary data

Etienne Wijler
Vrije Universiteit Amsterdam
Department of Econometrics and Data Science
August 30, 2022

Abstract

In this paper, we develop a restricted eigenvalue condition for unit-root non-stationary data and derive its validity under the assumption of independent Gaussian innovations that may be contemporaneously correlated. The method of proof relies on matrix concentration inequalities and offers sufficient flexibility to enable extensions of our results to alternative time series settings. As an application of this result, we show the consistency of the lasso estimator on ultra high-dimensional cointegrated data in which the number of integrated regressors may grow exponentially in relation to the sample size.

Keywords: REC, random matrix theory, lasso, cointegration.

JEL-Codes: C32, C55

1 Introduction

Modern time series applications frequently involve the analysis of high-dimensional datasets in which the number of time series $N$ is large in relation to, and may exceed, the number of temporal observations $T$. A rather successful approach to modelling such datasets has been to impose some form of sparsity on the underlying data generating process (DGP) and to incorporate $\ell_1$-based regularization such as the lasso to reduce model complexity and/or recover the true sparsity pattern (e.g Basu and Michailidis 2015, Medeiros and Mendes 2016, Masini et al. 2019). However, these approaches heavily rely on stationarity of the considered time series, thereby either excluding the presence of stochastic trends or requiring correct identification and corrections of the order of integration. As noted by Smeekes and Wijler (2020), a key difficulty that hinders extensions of the theory for lasso-type estimators to high-dimensional non-stationary settings is the absence of a verifiable restricted eigenvalue condition (REC). Accordingly, we aim to fill this gap in the literature by proposing an appropriately scaled REC for unit root non-stationary time series and verifying its validity directly on the gram matrix with the use of matrix concentration inequalities. This result is then used to derive novel finite sample error-bounds for the lasso and prove its asymptotic consistency on (ultra) high-dimensional (co)integrated datasets.

The behaviour of $\ell_1$-based regularization on stationary time series data is well-researched. Basu and Michailidis (2015) and Kock and Callot (2015) consider lasso based estimation of high-dimensional VAR processes driven by Gaussian innovations and derive tight error bounds for the prediction and estimation error. These bounds show that the estimation error deteriorates in the dimension at a logarithmic rate, thereby validating the lasso in (ultra) high-dimensional time series settings. Medeiros and Mendes (2016) consider the adaptive lasso and derive both estimation and selection consistency on weakly dependent and possibly heteroskedastic time series data without imposing Gaussianity on the predictors or innovations. By replacing the assumption of Gaussianity by a set of moment restrictions, they provide a considerable generalization to previous results, albeit at the cost of error bounds that deteriorate at faster rates...
in the dimension. Masini et al. (2019) generalize the aforementioned results on sparse VARs estimated with the lasso by deriving error bounds under the assumption of weak sparsity, while only assuming the innovation vector to be geometrically strong ($\alpha$-)mixing and a martingale difference process. Finally, Wong et al. (2020) provide further generalizations by avoiding the assumption of a specific parametric form of the DGP and relying only on (strict) stationarity and geometrically decaying $\beta$-mixing coefficients to establish error bounds for the lasso for subweibull random vectors.

Owing to the lack of suitable restricted eigenvalue conditions, the literature on the lasso in the non-stationary setting is substantially more scarce, with current results being limited to fixed or moderate dimensional settings only. For example, in the fixed dimensional setting, Liao and Phillips (2015) develop a penalized estimator for a vector error-correction model (VECM) and derive its consistency and pointwise limit distribution. Lee et al. (2021) consider a two-stage adaptively weighted lasso procedure applied to (co)integrated data for which the oracle property is shown to hold. Koo et al. (2020) consider the lasso applied to predictive regressions with covariates that are integrated of order one and derive error bounds under the assumption that the gram matrix satisfies an REC. The authors correctly note that the resulting lower bound that is assumed on the restricted eigenvalue may decrease to zero as the number of integrated time series grows, thereby inflating the estimation error at an unknown rate. Accordingly, consistency is again only derived for the case of a fixed number of integrated time series. Moving on to the moderate dimensional setting, Liang and Schienle (2019) develop a sparse VECM estimator that is argued to be consistent in a setting where the number of stochastic trends in the data increases at a rate slightly slower than $T^{1/4}$. Smeekes and Wijler (2020) consider lasso-type estimation of a single-equation error-correction model and derive consistency in an asymptotic framework in which the number of integrated time series is allowed to grow at rate $T^{1/4}$. In addition, the authors show that the minimum eigenvalue of the usual non-stationary gram matrix converges to zero as the dimension increases, but can be bounded away from zero with probability converging to one when scaled up by its dimension if the latter grows no faster than rate $T^{1/2}$.

To the best of the author’s knowledge, this paper is the first to develop and verify a restricted eigenvalue condition that enables theoretical analysis of $\ell_1$-regularized estimators in true high-dimensional (co)integrated datasets. Since the gram matrix based on integrated time series does not converge to a deterministic matrix, the commonly used indirect approach of imposing a restricted eigenvalue condition on the limit of the sample covariance matrix is not well-suited to high-dimensional settings. Therefore, we instead verify the condition directly on an appropriately scaled version of the gram matrix. Outside the realm of time series, a broad line of research exists that is concerned with direct verification of RE type conditions. In particular, letting $X$ be a $(T \times N)$-matrix with independent rows, numerous authors have shown that $\frac{1}{T}X'X$ satisfies the RE condition under Gaussian or sub-Gaussian tail conditions (e.g. Zhou, 2009; Raskuti et al. 2010; Rudelson and Zhou, 2012; Tropp, 2013; Kasiviswanathan and Rudelson, 2018). To the best of my knowledge, however, no contributions that deal with dependent rows exist in the current literature, let alone the case in which the rows are strongly persistent as is the case when dealing with integrated time series.

The RE condition developed in this paper enables the theoretical analysis of methods based on $\ell_1$-regularization, which we illustrate by deriving tight error bounds for the lasso when applied to cointegrated regression. We then use these error bounds to show that consistency can be maintained even when the number of potentially relevant integrated regressors grows exponentially in the sample size. Furthermore, we conjecture that the reliance on matrix concentration inequalities in the proofs offers sufficient flexibility for prospective future theoretical developments, which we motivate by providing several suggestions for valuable extensions.

The paper proceeds as follows. Section 2 presents the assumptions on the data, the restricted eigenvalue and the theoretical validity of the latter. Next, in Section 3 we apply the results of the
In this section, our main theoretical result is derived. In particular, we derive a finite-sample bound on the probability that the restricted eigenvalue of the gram matrix lies above a positive, universal constant. Further simplify notation, we let $Y_A = Y_{A,A}$. The main theoretical result brought forward in this paper concerns the behaviour of the restricted eigenvalues via matrix concentration inequalities.

Remark 1. Assumption [1] imposes the innovations to be independently and identically distributed according to a multivariate Gaussian distribution. Somewhat surprisingly, the choice for a Gaussian distribution is not motivated for its exponential tail decay, although this property is exploited in the results to follow. Instead, the Gaussian distribution’s invariance to orthonormal transformations is the key ingredient that enables the derivation of probabilistic bounds on restricted eigenvalues via matrix concentration inequalities.

The main theoretical result brought forward in this paper concerns the behaviour of the minimum restricted eigenvalue of $SS'$. 

**Theorem 1.** Assume that $\left(\frac{\log s}{c_\sigma}\right)^2 < N$. Then, under Assumption [1] there exists $\kappa_0 > 0$ such that

$$\mathbb{P}\left( \kappa \left( S, \frac{s \log^{3/4} N}{T}, s, c_0 \right) \geq \kappa_0 \right) \geq 1 - (4T + C_2s)e^{-C_1s \log N},$$

where $\kappa_0 := \kappa_0(c_0, c_\sigma, C_\sigma, C_1)$, whose expression is given in (34), is a constant that is independent of $s, N, T$, polynomially decreases in $c_0, C_\sigma, C_1$, and polynomially increases in $c_\sigma$.

A direct implication of Theorem [1] is that $\mathbb{P}\left( \kappa \left( S, \frac{s \log^{3/4} N}{T}, s, c_0 \right) \geq \kappa_0 \right) \rightarrow 1$ whenever $Te^{-s \log N} \rightarrow 0$ and $se^{-s \log N} \rightarrow 0$. Hence, perhaps somewhat counter-intuitively, the restricted eigenvalue condition is more likely to hold when $s$ and $N$ are large. The reason underlying this behaviour is the scaling factor of $s \log^{3/4} N$ included in the restricted minimum eigenvalue. We stress, however, that this is no free lunch, as the scaling factor will inversely affect the convergence rates of $\ell_1$-regularized estimators, as will become apparent in the following section.
3 Lasso estimation of cointegrated data

Suppose a researcher is interested in modelling a scalar-valued time series $y_t$ based on an $N$-dimensional vector time series $x_t$ by considering the model

$$y_t = \beta' x_t + \epsilon_{y,t},$$
$$x_t = x_{t-1} + \epsilon_{x,t}.$$  \hspace{1cm} (3)

The vector $\beta$ is assumed to be sparse, with $S_\beta = \{j : \beta_j \neq 0\}$ and $|S_\beta| = s$. Hence, $y_t$ is a unit root non-stationary time series that is cointegrated with $s$ integrated time series. We stack the innovations in the $(N+1)$ vector $\epsilon_t = (\epsilon_{y,t}, \epsilon_{x,t})'$.

Given that $N$ may be (very) large, we consider estimating the linear model (3) by the lasso. Define $y = (y_1, \ldots, y_T)'$ and $X = (x_1, \ldots, x_T)$. Then, the lasso estimator for $\beta$ is given by

$$\hat{\beta} = \arg\min_{\beta} \|y - X\beta\|^2 + \lambda_T \|\beta\|_1.$$  \hspace{1cm} (4)

Based on Theorem 1, we are able to derive the following error bounds on the prediction and estimation errors.

**Theorem 2.** Under Assumption 1 it holds that

$$\left\|X(\hat{\beta} - \beta)\right\|^2 + \lambda_T \|\beta - \beta\|_1 \leq \frac{4\lambda_T^2 s^3 \log^{3/2} N}{T^2 \phi_0^2},$$

with probability at least

$$1 - 2N \left[ \exp \left(-\frac{2\lambda_T^2}{T^{1+2m}}\right) + 2T \exp \left(-\frac{T^{m-\frac{1}{2}}}{2C_\sigma}\right) + 2\exp \left(-\frac{\lambda_T}{4C_\sigma}\right) \right] - (4T + C_2 s) e^{-C_1 s \log N},$$

for all $m > 0$, and universal constants $C_1, C_2 > 0$.

The probabilistic finite-sample error bound in Theorem 2 elucidates the dual role of the penalty parameter $\lambda_T$. Whilst a high degree of penalization results in a looser error bound, it simultaneously increases the probability with which this error bound holds. The attentive reader may note that the probability in (5) contains the lower bound on the probability that the REC holds from Theorem 1 minus an additional term. The latter term resembles the probability that $\lambda_T$ fails to dominate the empirical process, and is the decisive factor for determining the minimum amount of penalization required to ensure consistency. Indeed, based on Theorem 2 the consistency of the lasso estimator can be established under a suitable choice of $\lambda_T$ and appropriate restrictions on the growth rates of $T, N$, and $s$.

**Corollary 1.** Assume that, as $T, N, s \to \infty$, $\frac{T^{1+\xi} \log^{1/2} N}{\lambda_T} \to 0$ and $\frac{\log N}{T^{\xi}} \to 0$ for some $\xi > 0$, and $\frac{\log T}{s \log N} \to 0$. Then, under Assumption 1 it holds that

$$\left\|\hat{\beta} - \beta\right\|_1 = O_p \left(\frac{s^3 \log^2 N}{T^{1-\xi}}\right),$$

for any $\xi > 0$.

Corollary 1 makes the rate of consistency, as well as the sufficient restrictions on growth rates of $\lambda_T, T, N$ and $s$ explicit. Most noteworthy is that consistency may be attained when $N$ grows exponentially in $T$, thereby providing a first justification for the use of the lasso in ultra high-dimensional cointegrated applications. Furthermore, we note that when the number of time series that cointegrate with the dependent variable ($s$) grows at a slow polynomial rate in $T$, while the total number of candidate time series $N$ grows at any polynomial rate in $T$, then $\hat{\beta}$ almost attains the same $T$-consistency as in the fixed-dimensional setting.
4 Conclusion

In this paper, we validate the restricted eigenvalue condition (REC) proposed by Bickel et al. (2009) in a stationary setting to unit root nonstationary data. The gram matrix is scaled by $s^2 \log^{3/2} N$ instead of the usual $\frac{1}{N}$ that is familiar from the fixed-dimensional case. We apply our extended REC to derive finite-sample error bounds on the lasso when applied to high-dimensional cointegrating regressions. The bounds reveal that consistency can be attained, even under exponential growth of the number of time series. In addition, when the number of potentially relevant integrated regressors grow at any polynomial rate in $T$, while the number of regressors that cointegrate with the dependent variable grows at a slow polynomial rate, the estimator is almost $T$-consistent, as is the case in the classic fixed-dimensional setting.

A natural question that may arise concerns the validity of these results in the presence of multiple cointegrating relationships. We are optimistic that the results in this paper can serve useful when deriving the properties of system estimators such as penalized vector error correction models, and we consider this an interesting avenue for future research.

A Proofs

A.1 Lemmas

Lemma 1. Consider a finite sequence $\{X_t\}$ of independent, random, self-adjoint matrices with dimension $N$. Assume that each random matrix satisfies $X_t \succeq 0$, and $\lambda_{\max}(X_t) \leq R$ almost surely.

Define

$$\mu_{\min} := \lambda_{\min} \left( \sum_t \mathbb{E}(X_t) \right) \quad \text{and} \quad \mu_{\max} := \lambda_{\max} \left( \sum_t \mathbb{E}(X_t) \right).$$

Then,

$$P \left( \lambda_{\min} \left( \sum_{t=1}^T X_t \right) \leq (1 - \delta)\mu_{\min} \right) \leq N \left( \frac{e^{-\delta}}{(1 - \delta)^{1+\delta}} \right)^{\mu_{\min}/R} \quad \text{for } \delta \in [0, 1], \quad (6)$$

$$P \left( \lambda_{\max} \left( \sum_{t=1}^T X_t \right) \geq (1 + \delta)\mu_{\max} \right) \leq N \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^{\mu_{\min}/R} \quad \text{for } \delta \geq 0. \quad (7)$$

Lemma 2. Let $f_T = \frac{s \log^{3/4} N}{4}$ and define the set $\mathcal{A}(a) = \{f_T^2 \|X \epsilon_y\|_\infty \geq a\}$. Then, under Assumption [4] it holds that

$$P(\mathcal{A}(a)) \geq 1 - 2N \left[ \exp \left( \frac{2a^2T^{3-2m}}{s^4 \log^4 N} \right) + 2T \exp \left( -\frac{T^{m-\frac{1}{2}}}{2C_\sigma} \right) + 2 \exp \left( -\frac{aT^2}{4C_\sigma s^2 \log^2 N} \right) \right], \quad (8)$$

for any $m \in \mathbb{R}$.

Proof. The main argument to prove Lemma 2 relies on a truncation argument in combination with the Azuma-Hoeffding bound and is inspired by the proof of Theorem 1 in Kock and Callot (2015). First, by application of the union bound,

$$P \left( \max_{1 \leq i \leq N} \left| \sum_{t=1}^T x_{i,t} \epsilon_{y,t} \right| \geq \frac{a}{f_T^2} \right) \leq P \left( \max_{1 \leq i \leq N} \left| \sum_{t=1}^T x_{i,t-1} \epsilon_{y,t} \right| \geq \frac{a}{2f_T^2} \right) + P \left( \max_{1 \leq i \leq N} |\epsilon_{i,t} \epsilon_{y,t}| \geq \frac{a}{2f_T^2} \right) \quad (9)$$
We bound each of the two RHS terms in (9) individually. For the first term, we obtain

$$
P\left( \max_{1 \leq t \leq N} \left| \sum_{i=1}^{T} x_{i,t-1} \epsilon_{y,t} \right| \leq \frac{a}{2f_T^2} \right)$$

\[ \leq P\left( \max_{1 \leq t \leq N} \left| \sum_{i=1}^{T} x_{i,t-1} \epsilon_{y,t} \right| \leq \frac{a}{2f_T^2}, \max_{1 \leq t \leq T} x_{i,t-1} \epsilon_{y,t} \leq T^m \right) + P\left( \max_{1 \leq t \leq T} |x_{i,t-1} \epsilon_{y,t}| > T^m \right) \]

\[ \leq \sum_{i=1}^{N} P\left( \max_{1 \leq t \leq T} |x_{i,t-1} \epsilon_{y,t}| \leq \frac{a}{2f_T^2} \right) + P\left( \max_{1 \leq t \leq T} |x_{i,t-1} \epsilon_{y,t}| > T^m \right) \]

(10)

To bound the first term on the RHS of (10), we aim to apply the Azuma-Hoeffding inequality for bounded martingale difference sequences (e.g. [Wainwright, 2019, Corollary 2.20]). To this end, let $\mathcal{F}_t$ denote the natural filtration with respect to the stochastic process $(\epsilon_j)_{j=-\infty}^{\infty}$. By the proof of Lemma 4 in [Kock and Callot, 2015], it follows that

$$E(x_{i,t-1} \epsilon_{y,t} | |x_{i,t-1} \epsilon_{y,t}| \leq T^m | \mathcal{F}_{t-1}) = 0.$$  

Then, by the Azuma-Hoeffding inequality,

$$\sum_{i=1}^{N} P\left( \max_{1 \leq t \leq T} |x_{i,t-1} \epsilon_{y,t}| \leq \frac{a}{2f_T^2} \right) \leq 2N \exp \left( -\frac{a^2}{4f_T^4} \right) = 2N \exp \left( -\frac{a^2 T^4 - 2m}{s^4 \log^4 N} \right).$$

(11)

To bound the second term in (10), we first note that for any two random variables $x, y$ and constant $c > 0$, it holds that

$$P(|x| > c) \leq P(|x| > c) + P(|y| > c) \leq P(|x| > c^{1-k}) + P(|y| > c^k).$$

In addition, note that $x_{i,t-1} \sim N(0, \sigma_i^2)$ and $\epsilon_{y,t} \sim N(0, \sigma_y^2)$ by Assumption H. Then, based on a concentration inequality for sub-Gaussian variables (e.g. [Wainwright, 2019, equation (2.9)]), we bound

$$P\left( |x_{i,t-1} \epsilon_{y,t}| > T^m \right) \leq P\left( |x_{i,t-1}| > T^{mk} \right) + P\left( |\epsilon_{y,t}| > T^{m(1-k)} \right) \leq 2 \exp \left( -\frac{T^{2m(1-k)}}{2\sigma_i^2} \right) + 2 \exp \left( -\frac{T^{2m(1-k)}}{2\sigma_y^2} \right).$$

(12)

We optimize the rate at which (12) converges to zero by equating the exponents:

$$2mk - 1 = 2m(1-k) \Rightarrow k = \frac{1}{2} + \frac{1}{4m}.$$  

Plugging in this value for $k$, together with the upper bound $(\sigma_i^2, \sigma_y^2) \leq C_\sigma$ for all $1 \leq i \leq N+1$ and a union bound, it follows that

$$\sum_{i=1}^{N} P\left( \max_{1 \leq t \leq T} |x_{i,t-1} \epsilon_{y,t}| > T^m \right) \leq 4NT \exp \left( -\frac{T^{m} - \frac{1}{2}}{2C_\sigma} \right).$$

(13)

Next, again using the sub-Gaussian concentration inequality, we bound the second term of (9) as

$$P\left( \max_{1 \leq t \leq N} |\epsilon_{i,t} \epsilon_{y,t}| \geq \frac{a}{2f_T^2} \right) \leq \sum_{i=1}^{N} \left[ P\left( |\epsilon_{i,t}| \geq \left( \frac{a}{2f_T^2} \right)^{1/2} \right) + P\left( |\epsilon_{y,t}| \geq \left( \frac{a}{2f_T^2} \right)^{1/2} \right) \right] \leq 4N \exp \left( -\frac{a^2}{4f_T^2 C_\sigma} \right) = 4N \exp \left( -\frac{aT^2}{4C_\sigma s^2 \log^{3/2} N} \right).$$

(14)
Finally, combining (11), (13) and (14), we bound the RHS of (9) as
\[
\mathbb{P} \left( \max_{1 \leq t \leq N} \left| \sum_{t=1}^{T} x_{t,t} \phi_{t} \right| \geq \frac{a}{T^{2}} \right) \\
\leq 2N \left[ \exp \left( -\frac{2a^{2}T^{2}e^{-2m}}{s^{4} \log \frac{1}{N}} \right) + 2T \exp \left( -\frac{Tm-4}{2C_{\sigma}} \right) + 2\exp \left( -\frac{aT^{2}}{4C_{\sigma}^{2}e^{3}\log^{3/2} \frac{1}{N}} \right) \right].
\]

A.2 Proofs of main results

Proof of Theorem 1

We first show that the restricted minimum eigenvalue condition holds for a transformed matrix. Rewrite $S = UE$, where $U$ is a lower triangular matrix with ones on and below the diagonal and $E = (\epsilon_1, \ldots, \epsilon_T)'$. Decompose $U'U = V \Lambda V'$, where $A = \text{diag}(\lambda_1, \ldots, \lambda_T)$ are the eigenvalues of $U'U$. By Lemma 1 in Akesson and Lehoczky (1998), it holds that
\[
\lambda_{-1} = 4 \sin \left( \frac{\omega}{2} \right) = 2(1 - \cos \omega),
\]
with $\omega = \frac{2T+1}{2T}$.
Both $\lambda_{1} = O(T^2)$ and $\sum_{t=1}^{T} \lambda_{t} = O(T^2)$, which turns out to be troublesome when deriving the behaviour of the spectrum. Accordingly, we consider a transformed matrix
\[
S'(S) = \sum_{t=1}^{T} \lambda_{t} \phi_{t} \tilde{e}_t \tilde{e}_t',
\]
with
\[
\lambda_{-1} = 2(1 + \phi - \cos \omega),
\]
where $\phi$ is a positive sequence that decreases to zero as $T, N, s \to \infty$. Note that
\[
S'S - S'(S') = \sum_{t=1}^{T} (\lambda_{t} - \lambda_{\phi,t}) \tilde{e}_t \tilde{e}_t' \geq 0,
\]
thus implying that $\kappa(S, s, c_0) > \kappa(S'(S), s, c_0)$. Define the sparse minimum and maximum eigenvalues as
\[
\phi_{\min}(S'(S), f_T, u) = \min_{x \in \mathbb{R}^{N}: 1 \leq M(x) \leq u} \frac{\|f_T S'(S)x\|_2^2}{\|x\|_2^2},
\]
\[
\phi_{\max}(S'(S), f_T, u) = \max_{x \in \mathbb{R}^{N}: 1 \leq M(x) \leq u} \frac{\|f_T S'(S)x\|_2^2}{\|x\|_2^2}.
\]
By Lemma 4.1(ii) in Bickel et al. (2009), it follows that
\[
\kappa(S'(S), f_T, s, c_0) \geq \sqrt{\phi_{\min}(S'(S), f_T, s + m)} - c_0 \sqrt{\phi_{\max}(S'(S), f_T, m)} \sqrt{\frac{s}{m}},
\]
where $m \geq s$ and $s + m \leq N$. Accordingly, we aim to provide a lower bound for (13) by deriving probabilistic upper and lower bounds for the sparse minimum and maximum eigenvalues, respectively.

Let the collection of all index sets of cardinality $s$ be denoted by
\[
\mathcal{A}(s) := \{A \subseteq \{1, \ldots, N\} : |A| = s\},
\]
and note that its cardinality is upper bounded by
\[
|\mathcal{A}(s)| \leq \left( \frac{eN}{s} \right)^s \leq e^{s \log(eN/s)} \leq e^{s(1 - \log s + \log N)} \leq e^{s \log N},
\]
\(^{1}\)The second equality in (15) is based on the identity $\cos(2\alpha) = 1 - 2\sin^2(\alpha)$. 

\[\]
for $s \geq 3$. For any $A \in \mathcal{A}(s)$, let $S_A^{(\phi)} = U^{(\phi)} E_A$ represent the columns of $S^{(\phi)}$ indexed by $A$. Then, an equivalent definition of the sparse eigenvalues in (17) is given by

$$
\begin{align*}
\phi_{\min}(S^{(\phi)}_A, f_T, s) &= \min_{A \in \mathcal{A}(s)} \lambda_{\min} \left( f_T^2 S^{(\phi)}_A S^{(\phi)}_A^T \right), \\
\phi_{\max}(S^{(\phi)}_A, f_T, s) &= \max_{A \in \mathcal{A}(s)} \lambda_{\max} \left( f_T^2 S^{(\phi)}_A S^{(\phi)}_A^T \right).
\end{align*}

(20)

We proceed to derive the stochastic order of the quantities in (20). For any $A \in \mathcal{A}(s)$ it follows that

$$
S^{(\phi)}_A S^{(\phi)}_A^T = E_A V A^{(\phi)} V^T E_A = \bar{E}_A A^{(\phi)} \bar{E}_A = \sum_{t=1}^T \lambda_{\phi,s,t} \bar{e}_{A,t} \bar{e}_{A,t}^T,
$$

where $\bar{e}_{A,t} \overset{i.i.d.}{\sim} \mathcal{N}(0, \Sigma_A)$ by the rotational invariance of the multivariate normal distribution. Let

$$
\mu_{\phi,s}(A) := \lambda_{\min} \left( \sum_{t=1}^T \lambda_{\phi,s,t} \mathbb{E} \left( \bar{e}_{A,t} \bar{e}_{A,t}^T \right) \right) = \lambda_{\min} \left( \Sigma_A \right) \sum_{t=1}^T \lambda_{\phi,s,t},
$$

and the corresponding quantities

$$
\mu_{\phi,s}^* = \min_{A \in \mathcal{A}(s)} \mu_{\phi,s}(A),
$$

$$
\mu_{\phi,s}^{**} = \max_{A \in \mathcal{A}(s)} \mu_{\phi,s}(A).
$$

In addition, define

$$
R_s = C_2^s \lambda_{\phi,s,1} s \left( (C_1 + 1) \log^{1/2} N + 1 \right),
$$

(21)

for an arbitrary constant $C_1 > 0$. First, we use the truncation argument

$$
\begin{align*}
\mathbb{P} \left( \min_{A \in \mathcal{A}(s)} \lambda_{\min} \left( \frac{1}{\mu_{\phi,s}(A)} S^{(\phi)}_A S^{(\phi)}_A^T \right) \leq 1 - \delta \right)
\leq \mathbb{P} \left( \min_{A \in \mathcal{A}(s)} \lambda_{\min} \left( \frac{1}{\mu_{\phi,s}(A)} S^{(\phi)}_A S^{(\phi)}_A^T \right) \leq 1 - \delta, \sup_{A \in \mathcal{A}(s)} \lambda_{\max} \left( \lambda_{\phi,s,t} \bar{e}_{A,t} \bar{e}_{A,t}^T \right) \leq R_s \right)
\leq \mathbb{P} \left( \sup_{A \in \mathcal{A}(s)} \lambda_{\max} \left( \lambda_{\phi,s,t} \bar{e}_{A,t} \bar{e}_{A,t}^T \right) \geq R_s \right).
\end{align*}

(22)

The first term on the RHS can now be bounded with the use of the Chernoff bound in Lemma [1].

Define the constant

$$
K_\delta = \max \left( \frac{e^{-\delta}}{(1 - \delta)^{1+\delta}}, \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)
$$

and note that $0 < K_\delta < 1$ for any $0 < \delta < 1$. Then, Lemma [1] states that, for each $A \in \mathcal{A}(s)$,

$$
\mathbb{P} \left( \lambda_{\min} \left( \frac{1}{\mu_{\phi,s}(A)} S^{(\phi)}_A S^{(\phi)}_A^T \right) \leq 1 - \delta, \sup_{1 \leq t \leq T} \lambda_{\max} \left( \lambda_{\phi,s,t} \bar{e}_{A,t} \bar{e}_{A,t}^T \right) \leq R_s \right) \leq s K_\delta^{\mu_{\phi,s}(A)/R_s}.
$$

(23)

For any $s \leq N$, we can bound $\mu_{\phi,s}^*$ from below as

$$
2 \mu_{\phi,s}^* \geq \sum_{t=1}^T \frac{c_\sigma}{1 - \cos \omega_t + \phi_s} \geq \sum_{t=1}^T \frac{c_\sigma}{\left( \frac{(2t+1)\pi}{2T+1} \right)^2 + \phi_s}
$$

$$
= \sum_{t=1}^T \frac{c_\sigma (2T+1)^2}{\left( (2t+1)^2 \pi^2 + \phi_s (2T+1)^2 \right)},
$$

(24)
Since $\phi_s$ is defined as a positive sequence decreasing to zero in $T$, we have that $\phi_s < 1$ for large enough $T$. It follows that

$$
\mu_{\min}^s \geq \sum_{t=1}^{T} \frac{c_\sigma (2T + 1)^2}{2((2t + 1)^2\pi^2 + \phi_s (2T + 1)^2)} \geq \sum_{t=1}^{T} \frac{c_\sigma (2T)^2}{2((3t)^2\pi^2 + \phi_s (3T)^2\pi^2)} \geq \sum_{t=1}^{T} 4c_\sigma T^2 \geq \sum_{t=1}^{T} \frac{4c_\sigma \phi_s^{1/2}T^3}{36\pi^2 \phi_s T^2} = \frac{c_\sigma T}{9\pi^2 \phi_s^{1/2}}.
$$

(25)

Next, bound $\mu_{\min}^s / R_s$ as

$$
\frac{\mu_{\min}^s}{R_s} \geq \frac{c_\sigma T}{9\pi^2 C_\sigma^2 \phi_s^{1/2} \lambda_{\phi,1} s \left( (C_1 + 1) \log^{1/2} N + 1 \right)} \geq \frac{c_\sigma T (1 - \cos \omega_1 + \phi_s)}{9\pi^2 C_\sigma^2 \phi_s^{1/2} s \left( (C_1 + 1) \log^{1/2} N + 1 \right)} \geq \frac{c_\sigma T \phi_s^{1/2}}{9\pi^2 C_\sigma^2 (C_1 + 3)s \log^{1/2} N} \geq \frac{- (C_1 + 1)s \log N}{\log K_s}. \tag{26}
$$

where (*) follows from the observation that $2 \log^{1/2} N > 1$ for any $N \geq 2$ and (***) follows after setting

$$
\phi_s^{1/2} = \frac{- 9\pi^2 (C_1 + 1) (C_1 + 3) C_\sigma^2 s^2 \log^{3/2} N}{c_\sigma T \log K_s}. \tag{27}
$$

Note that $\phi_s^{1/2}$ is positive due to the negative term log $K_s$ in the denominator. Then, combining (25)–(27), we use the Chernoff matrix-bound from Lemma 1 to bound the first RHS term of (22) as

$$
\mathbb{P} \left( \min_{A \in \mathcal{A}(s)} \lambda_{\min} \left( \frac{1}{\mu_{\min}^s (A)} S_A^{(\phi_s)^{1/2}} S_A^{(\phi_s)^{1/2}} \right) \leq 1 - \delta, \sup_{A \in \mathcal{A}(s)} \lambda_{\max} (\lambda_{\phi,s}, \vec{e}_t \vec{\epsilon}_t) \leq R_s \right) \leq \sum_{A \in \mathcal{A}(s)} \mathbb{P} \left( \min_{A \in \mathcal{A}(s)} \lambda_{\min} \left( \frac{1}{\mu_{\min}^s (A)} S_A^{(\phi_s)^{1/2}} S_A^{(\phi_s)^{1/2}} \right) \leq 1 - \delta, \sup_{A \in \mathcal{A}(s)} \lambda_{\max} (\lambda_{\phi,s}, \vec{e}_t \vec{\epsilon}_t) \leq R_s \right) \leq s \cdot \exp \left( s \log N + \frac{\mu_{\min}^s \log K_s}{R_s} \right) \leq s \cdot \exp \left( - C_1 s \log N \right). \tag{28}
$$

Next, we proceed by bounding the second RHS term of (22). Let $v_t \sim \mathcal{N}(0, I_{|A|})$, such that $\vec{\epsilon}_{A,t} \overset{d}{=} \Sigma_A^{1/2} v_t$. By application of the union bound, the fact that the $\vec{\epsilon}_t$ are identically distributed, the definition of $R_s$ and the triangle inequality, it holds that

$$
\mathbb{P} \left( \sup_{A \in \mathcal{A}(s)} \lambda_{\max} \left( \lambda_{\phi,s}, \vec{\epsilon}_{A,t} \vec{\epsilon}_{A,t}^\top \right) \geq R_s \right) \leq \sum_{A \in \mathcal{A}(s)} \mathbb{P} \left( \lambda_{\max} \left( \lambda_{\phi,s}, \vec{\epsilon}_{A,t} \vec{\epsilon}_{A,t}^\top \right) \geq \frac{R_s}{\lambda_{\phi,s}} \right) \leq T \sum_{A \in \mathcal{A}(s)} \mathbb{P} \left( \| \vec{\epsilon}_{A,1} \vec{\epsilon}_{A,1}^\top \|_2 \geq \frac{R_s}{\lambda_{\phi,s}} \right) \leq T \sum_{A \in \mathcal{A}(s)} \mathbb{P} \left( \frac{1}{s} \sum_{j \in A} v_j^2 \geq (C_1 + 1) \log^{1/2} N + 1 \right). \tag{29}
$$

where (*) follows by the sub-exponential concentration of property of $\chi^2$ random variables (see Example 2.11 in Wainwright (2019)). Finally, plugging (25) and (27) into (22), we obtain the bound for the minimum sparse eigenvalue as

$$
\mathbb{P} \left( \phi_{\min} \left( S^{(\phi_s)}, \frac{1}{\sqrt{\mu_{\min}^s}} s \right) \leq 1 - \delta \right) \leq \mathbb{P} \left( \min_{A \in \mathcal{A}(s)} \lambda_{\min} \left( \frac{1}{\mu_{\min}^s (A)} S_A^{(\phi_s)^{1/2}} S_A^{(\phi_s)^{1/2}} \right) \leq 1 - \delta \right) \leq \left( 2T + s \right) \exp \left( - C_1 s \log N \right). \tag{30}
$$


Finally, since the lower bounds in (25) and (27) remain valid for $\mu_{\max}$ and $\frac{\mu_{\max}}{R_{\min}}$, respectively, it follows by analogous reasoning that the maximum sparse eigenvalue is bounded as

$$\mathbb{P}\left(\phi_{\max}^{\beta} \left(S(\phi_s), \frac{1}{\sqrt{\mu_{\max}}}, s\right) \geq 1 + \delta\right) \leq \mathbb{P}\left(\max_{A \in \mathcal{A}(s)} \lambda_{\max} \left(\frac{1}{\phi_{\max}(A)} S_A(\phi_s)^\beta S_A(\phi_s)\right) \geq 1 + \delta\right) \leq (2T + s) \exp\left(-C_1 s \log N\right).$$

Having derived probabilistic bounds for the minimum and maximum sparse eigenvalues, we proceed to prove the restricted eigenvalue condition in Theorem 1. Define the set

$$\mathcal{R}(s, m) = \left\{ \phi_{\min} \left(S(\phi_{s,m}), \frac{1}{\sqrt{\mu_{\min}^{s+m}}}, s + m\right) > 1 - \delta \right\} \cap \left\{ \phi_{\max} \left(S(\phi_s), \frac{1}{\sqrt{\mu_{\max}^{s+m}}}, s\right) < 1 + \delta \right\}. $$

For a positive constant $\kappa > 0$, define

$$m_{\kappa} = \left[ \frac{c_0^2 C_\sigma (1 + \delta) s}{c_\sigma \sqrt{1 - \delta - \kappa}} \right] =: [C_\kappa s],$$

with $C_\kappa = \frac{c_0^2 C_\sigma (1 + \delta)}{c_\sigma \sqrt{1 - \delta - \kappa}}$. Then, continuing from (18), it holds on $\mathcal{R}(s, m_{\kappa})$ that

$$\kappa \left(S(\phi_s), \frac{1}{\sqrt{\mu_{\min}^{s+m_{\kappa}}}}, s, c_0\right) \geq \phi_{\min} \left(S(\phi_s), \frac{1}{\sqrt{\mu_{\min}^{s+m_{\kappa}}}}, s + m_{\kappa}\right) - c_0 \sqrt{\frac{s}{m_{\kappa}}} \geq \phi_{\max} \left(S(\phi_s), \frac{1}{\sqrt{\mu_{\max}^{s+m_{\kappa}}}}, s\right) \sqrt{s} \frac{1}{m_{\kappa}} \geq \sqrt{1 - \delta} - c_0 \sqrt{\frac{C_\sigma (1 + \delta) s}{c_\sigma m_{\kappa}}} \geq \kappa,$$

where (*) uses that $\lambda_{\phi_{s,m}, t} \geq \lambda_{\phi_{s+m}, t}$ and (**) uses that

$$\frac{\mu_{\max}^{s+m_{\kappa}}}{\mu_{\min}^{s+m_{\kappa}}} \leq \frac{\mu_{\max}^{s+m_{\kappa}}}{\mu_{\min}^{s+m_{\kappa}}} \leq C_\sigma \frac{\mu_{\max}^{s+m_{\kappa}}}{\mu_{\min}^{s+m_{\kappa}}} \leq C_\sigma.$$ 

Note that by (25) and (27)

$$\frac{1}{\mu_{\max}^{s+m_{\kappa}}} \leq \frac{9\pi^2 \phi_{s+m_{\kappa}}^{1/2}}{c_\sigma T} = \frac{81\pi^4 (C_1 + 1)(C_1 + 3) C_\sigma^2 (C_1 + 1)^2 s_2 \log^{3/2} N}{c_\sigma T^2 \log K_{\phi}} = \frac{C_\mu s^2 \log^{3/2} N}{T^2},$$

with $C_\mu = -\frac{81\pi^4 (C_1 + 1)(C_1 + 3) C_\sigma^2 (C_1 + 1)^2}{c_\sigma^2 \log K_{\phi}}$. Then, we conclude that

$$\kappa \left(S, \frac{s \log^{3/4} N}{T}, s, c_0\right) \geq C_\mu^{-1/2} \kappa \left(S(\phi_s), \frac{1}{\sqrt{\mu_{\min}^{s+m_{\kappa}}}}, s, c_0\right) \geq C_\mu^{-1/2} \kappa =: \kappa_0$$

on the set $\mathcal{R}(s, m_{\kappa})$, where the fully expanded expression for $\kappa_0$ reads

$$\kappa_0 := \kappa_0(c_0, c_\sigma, C_\sigma, C_1) = \frac{\kappa c_\sigma \log^{1/2} K_{\phi}}{9\pi^2 (C_1 + 1)(C_1 + 3) C_\sigma \left(\frac{c_0^2 C_\sigma (1 + \delta)}{c_\sigma \sqrt{1 - \delta - \kappa}} + 1\right)},$$

(34)
where $\kappa > 0$ and $0 < \delta < 1$ are free constants chosen such that \[ \frac{C_2(1+\delta)s}{\kappa(\sqrt{T}-\sqrt{\delta\kappa})^2} < N. \]

Finally, based on (30) and (31),
\[
\mathbb{P}(\mathcal{R}(s, m_\kappa)) \geq 1 - (2T + s)e^{-C_1s\log N} - (2T + m_\kappa)e^{-C_1m_\kappa\log N} \\
\geq 1 - (4T + (C_k + 1)s)e^{-C_1s\log N} = 1 - (4T + C_2s)e^{-C_1s\log N},
\]
with $C_2 = C_\kappa + 1$. This completes the proof of Theorem 1.

**Proof of Theorem 2.** The proof of the error bound is based on standard arguments, see for example Theorem 6.1 in [Bühlmann and Van De Geer (2011)](https://link.springer.com/article/10.1007/978-3-540-71267-5_6), but the probability on which it holds requires new arguments. First, let $f_T = \frac{T^{1/4}}{\sqrt{N}}$ and define $\bar{T}_T = f_T^2 \lambda_T$, such that the estimator can equivalently be defined as the minimizer of
\[
\mathcal{L}(\beta) = f_T^2 \left\| y - X\beta \right\|_2^2 + \bar{T}_T \left\| \beta \right\|_1.
\]
Next, define the set
\[
\mathcal{A}_T \left( \bar{T}_T \right) = \left\{ \bar{T}_T \left\| X' \varepsilon_T \right\|_\infty \leq \frac{\bar{T}_T}{4} \right\}.
\]
By construction $\mathcal{L}(\tilde{\beta}) \leq \mathcal{L}(\beta)$. Then, on the set $\mathcal{A}_T$, it holds that
\[
f_T^2 \left\| y - X\tilde{\beta} \right\|_2^2 + \bar{T}_T \left\| \tilde{\beta} \right\|_1 \leq f_T^2 \left\| y - X\beta \right\|_2^2 + \bar{T}_T \left\| \beta \right\|_1
\]
\[
\Rightarrow f_T^2 \left\| X(\tilde{\beta} - \beta) \right\|_2^2 + \bar{T}_T \left\| \tilde{\beta} \right\|_1 \leq 2f_T^2 \left\| X' \varepsilon_T \right\|_\infty \left\| \tilde{\beta} - \beta \right\|_1 + \bar{T}_T \left\| \beta_s \right\|_1.
\]
Using that
\[
\left\| \tilde{\beta} \right\|_1 = \left\| \hat{\beta}_s \right\|_1 + \left\| \hat{\beta}_{s^c} \right\|_1 \geq \left\| \hat{\beta}_s \right\|_1 - \left\| \hat{\beta}_s - \beta_s \right\|_1 + \left\| \hat{\beta}_{s^c} \right\|_1,
\]
it follows from (37) that
\[
2\bar{T}_T \left\| \tilde{\beta} \right\|_1 \leq \bar{T}_T \left\| \tilde{\beta} - \beta \right\|_1 + 2\bar{T}_T \left\| \beta_s \right\|_1 \Rightarrow \left\| \hat{\beta}_s \right\|_1 \leq 3 \left\| \hat{\beta}_s - \beta_s \right\|_1.
\]
Define the cone
\[
\mathcal{C}_N(S, c_0) := \{ x \in \mathbb{R}^N : \| x_S \|_1 \leq c_0 \| x_s \|_1 \}
\]
and note that (38) implies that $\hat{\beta} - \beta \in \mathcal{C}_N(S, 3)$. Recall the definition of $\mathcal{R}(\hat{\Sigma}_T, s, c_0)$ in (1) and define the set
\[
\mathcal{D}_T = \left( \mathcal{R}(\hat{\Sigma}_T, s, c_0) \right)^{\mathbb{R}_+}
\]
assume that $\kappa (\hat{\Sigma}_T, s, 3) \geq \phi_0$. Then, combining (37) and (38),
\[
2 \left\| f_T X(\tilde{\beta} - \beta) \right\|_2^2 + \bar{T}_T \left\| \tilde{\beta} - \beta \right\|_1 \leq 4\bar{T}_T \left\| \hat{\beta}_s - \beta_s \right\|_1 \leq 4\sqrt{\bar{T}_T} \left\| \hat{\beta}_s - \beta_s \right\|_2
\]
\[
\leq \frac{4\bar{T}_T \sqrt{\phi_0}}{\phi_0} \left\| f_T X(\tilde{\beta} - \beta) \right\|_2 \leq \frac{4\bar{T}_T \sqrt{s}}{\phi_0} + \left\| f_T X(\tilde{\beta} - \beta) \right\|_2^2
\]
with the last inequality following from the elementary inequality $2uv \leq u^2 + v^2$. Hence, we are left to conclude that, on $\mathcal{A}_T \cap \mathcal{D}_T$,
\[
\left\| f_T X(\tilde{\beta} - \beta) \right\|_2 \leq \frac{4\bar{T}_T \sqrt{s}}{\phi_0}
\]
\[
\Rightarrow \left\| X(\tilde{\beta} - \beta) \right\|_2 \leq \frac{4f_T^2 \lambda_T \phi_0}{\phi_0} = \frac{4\lambda_T^2 s^3 \log^{3/2} N}{T^2 \phi_0^2}
\]

11
Finally, we note that $\Pr\left(\mathcal{A}T\left(\tilde{\lambda} T\right) \cap \mathcal{B} T\right) \leq 1 - \Pr\left(\mathcal{A}T\left(\tilde{\lambda} T\right)^c\right) - \Pr\left(\mathcal{B} T^c\right)$. By Lemma 2 it holds that

$$\Pr\left(\mathcal{A} T^c\right) \leq 2N \left[ \exp\left(-\frac{2\lambda^2 T^2}{T^{1+2m}}\right) + 2T \exp\left(-\frac{T m - \frac{1}{2}}{2C_\sigma}\right) + 2 \exp\left(-\frac{\tilde{\lambda} T^2}{4C_\sigma s^2 \log^{3/2} N}\right) \right]$$

$$= 2N \left[ \exp\left(-\frac{2\lambda^2 T^2}{T^{1+2m}}\right) + 2T \exp\left(-\frac{T m - \frac{1}{2}}{2C_\sigma}\right) + 2 \exp\left(-\frac{\tilde{\lambda} T}{4C_\sigma}\right) \right],$$

whereas it follows from Theorem 1 that

$$\Pr\left(\mathcal{B} T^c\right) \leq (4T + C_2 s)e^{-C_1 s \log N}.$$ Combining the latter two inequalities completes the proof.

**Proof of Corollary 1**. We show that under the the rate restrictions in Corollary 1, and Assumption 1, the error bound in Theorem 2 holds with probability converging to 1 as $T, N, s \to \infty$. This requires us to show that

$$2N \left[ \exp\left(-\frac{2\lambda^2 T^2}{T^{1+2m}}\right) + 2T \exp\left(-\frac{T m - \frac{1}{2}}{2C_\sigma}\right) + 2 \exp\left(-\frac{\tilde{\lambda} T}{4C_\sigma}\right) \right] + 4T + C_2 s)e^{-C_1 s \log N} = A_1 + A_2 + A_3 + A_4 \to 0,$$

as $T, N, s \to \infty$. For $A_1$, we set $m = 1/2 + \xi$, and note that

$$A_1 = \exp\left(-\frac{2\lambda^2 T^2}{T^{1+2m}} + \log N\right) \to 0$$

if $\frac{T^{1+\xi} \log^{1/2} N}{\lambda T} \to 0$ as $T, N, s \to \infty$. Next, note that

$$A_2 = \exp\left(\frac{T \xi}{2C_\sigma} + \log N + \log 2T\right) \to 0,$$

if $\frac{\log N}{T} \to 0$. Moving to $A_3$, it holds that

$$A_3 = 2 \exp\left(-\frac{\lambda T}{4C_\sigma} + \log N\right) \to 0,$$

if $\frac{\log N}{\lambda T} \to 0$ as $T, N, s \to \infty$, which is also implied by the combination of conditions that $\frac{T^{1+\xi} \log^{1/2} N}{\lambda T} \to 0$ and $\frac{\log N}{T} \to 0$ as $T, N, s \to \infty$. Finally, it holds that

$$A_4 = (4T + C_2 s)e^{-C_1 s \log N} = 4 \exp\left(-C_1 s \log N + \log T\right) + C_2 \exp\left(-C_1 s \log N + \log s\right) \to 0,$$

if $\frac{s \log N}{\log T} \to \infty$ and $\frac{\log N}{\log s} \to \infty$, as $T, N, s \to \infty$. The latter condition holds trivially, while the former condition is implied by the condition $\frac{\log T}{s \log N} \to 0$ as $T, N, s \to \infty$. As a result, it follows that $\mathcal{P}(T, N, s) \to 1$ as $T, N, s \to \infty$, with $\mathcal{P}(T, N, s)$ being defined in (5), thereby completing the proof.
References

Akesson, F. and Lehoczky, J. (1998). Discrete eigenfunction expansion of multidimensional Brownian motion and the Ornstein-Uhlenbeck process. Preprint.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. The Annals of Statistics, 43:1535–1567.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and dantzig selector. The Annals of Statistics, 37:1705–1732.

Bühlmann, P. and Van De Geer, S. (2011). Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer Science & Business Media.

Kasiviswanathan, S. P. and Rudelson, M. (2018). Restricted eigenvalue from stable rank with applications to sparse linear regression. In Conference On Learning Theory, pages 1011–1041. PMLR.

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. Journal of Econometrics, 186:325–344.

Koo, B., Anderson, H. M., Seo, M. H., and Yao, W. (2020). High-dimensional predictive regression in the presence of cointegration. Journal of Econometrics, 219:456–477.

Lee, J. H., Shi, Z., and Gao, Z. (2021). On lasso for predictive regression. Journal of Econometrics.

Liang, C. and Schienle, M. (2019). Determination of vector error correction models in high dimensions. Journal of Econometrics, 208(2):418–441.

Liao, Z. and Phillips, P. C. B. (2015). Automated estimation of vector error correction models. Econometric Theory, 31:581–646.

Masini, R. P., Medeiros, M. C., and Mendes, E. F. (2019). Regularized estimation of high-dimensional vector autoregressions with weakly dependent innovations. arXiv e-prints 1912.09002, arXive.

Medeiros, M. C. and Mendes, E. F. (2016). $\ell_1$-regularization of high-dimensional time series models with non-gaussian and heteroskedastic errors. Journal of Econometrics, 191:255–271.

Raskutti, G., Wainwright, M. J., and Yu, B. (2010). Restricted eigenvalue properties for correlated gaussian designs. The Journal of Machine Learning Research, 11:2241–2259.

Rudelson, M. and Zhou, S. (2012). Reconstruction from anisotropic random measurements. In Conference on Learning Theory, pages 10–1. JMLR Workshop and Conference Proceedings.

Smeekes, S. and Wijler, E. (2020). An automated approach towards sparse single-equation cointegration modelling. Journal of Econometrics.

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. Foundations of Computational Mathematics, 12:389–434.

Tropp, J. A. (2015). Convex recovery of a structured signal from independent random linear measurements. In Sampling Theory, a Renaissance, pages 67–101. Springer.

Wainwright, M. J. (2019). High-dimensional statistics: A non-asymptotic viewpoint, volume 48. Cambridge University Press.
Wong, K. C., Li, Z., and Tewari, A. (2020). Lasso guarantees for $\beta$-mixing heavy-tailed time series. *Annals of Statistics*, 48:1124–1142.

Zhou, S. (2009). Restricted eigenvalue conditions on subgaussian random matrices. arxiv e-prints, arXive.