Statistica Sinica Preprint No: SS-2023-0364	
Title	Concentration Inequalities for High-Dimensional Linear Processes with Dependent Innovations
Manuscript ID	SS-2023-0364
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0364
Complete List of Authors	Eduardo Fonseca Mendes and Fellipe Lopes Lima Leite
Corresponding Authors	Eduardo Fonseca Mendes
E-mails	eduardo.mendes@fgv.br
Notice: Accepted author version.	

Concentration inequalities for high-dimensional linear processes with dependent innovations

Eduardo Fonseca Mendes and Fellipe Lopes Lima Leite

Getulio Vargas Foundation

Abstract: We develop concentration inequalities for the l_∞ norm of vector linear processes with sub-Weibull, mixingale innovations. This inequality is used to obtain a concentration bound for the maximum entrywise norm of the lag-h autocovariance matrix of linear processes. We apply these inequalities to sparse estimation of large-dimensional VAR(p) systems and heterocedasticity and autocorrelation consistent (HAC) high-dimensional covariance estimation.

Key words and phrases: **Keywords**: high-dimensional time series, linear process, mixingale, sub-Weibull, autocovariance, HAC.

1. Introduction

In this paper we develop general concentration inequalities that are used in high-dimensional statistical literature. We study a class of high-dimensional time series that can be represented by linear processes with dependent innovations. Specifically, we assume that innovation process is mixingale with sub-Weibull tails. This specification covers a wide range of data-generating processes, such as factor and conditionally heteroskedastic models, as discussed in Wong et al. (2020); Bours and Steland (2021); Wilms et al. (2021); Masini et al. (2022); Adamek et al. (2023a), among many others.

Linear processes are widely used in time series analysis for their ability to represent a wide range of dependent processes. For example, the Wold decomposition theorem represents stationary nonlinear processes as a linear process with uncorrelated innovations. The Vector Autoregressive Moving Average (VARMA) model, in turn, approximates the time series by a linear process indexed by a finite number of parameters. Typically, innovations are assumed to be independent, allowing for well-understood asymptotic properties (Hall and Heyde, 1980; Phillips and Solo, 1992; Lütkepohl, 2006). As an alternative to independence, Wang et al. (2001) shows the weak convergence of partial sums for martingale difference innovations, while Wu and Min (2005) establish a central limit theorem and the invariance principle for these processes. Dedecker et al. (2011) presents maximal inequalities and a functional central limit theorem for innovations in a class of weakly dependent processes. For a comprehensive treatment of multivariate time series, linear representation of nonlinear processes, and illustrative examples, refer to Lütkepohl (2006); Brockwell and Davis (2009) and Tsay (2013).

High-dimensional statistics deal with the problem when the dimension of the random vector is large, and one is often interested in obtaining concentration bounds for averages that are uniform on the dimension. Typically, these concentrations are derived under independence and either sub-Gaussian or sub-exponential tails as thoroughly discussed in Vershynin (2018, chap. 2 and 3), Wainwright (2019, chap. 3), and Zhang and Chen (2021). In order to account for dependence, many uniform concentration results have been derived for mixing processes such as in Yu (1994); Marton (1998); Merlevède et al. (2009); Mohri and Rostamizadeh (2010); Hang and Steinwart (2017); Wong et al. (2020); Fan et al. (2023) and others. However, mixing is often difficult to verify in practice, favouring alternative forms of dependence, such as weak dependence (Dedecker et al., 2007) and functional dependence (Wu, 2005) in Doukhan and Neumann (2007); Alquier and Doukhan (2011); Adamczak (2015); Zhang and Wu (2017), among many others. Sub-Weibul tails have recently appeared as a weaker alternative to sub-Gaussian and sub-exponential tails (Vladimirova et al., 2020; Wong et al., 2020; Götze et al., 2021; Kuchibhotla and Chakrabortty, 2022; Zhang and Wei, 2022; Bong and Kuchibhotla, 2023). Sub-Weibull random variables do not have moment generating functions and are often encountered when dealing with products or powers of sub-Gaussian and subexponential random variables.

Jiang (2009) and Chen and Wu (2018) develop concentration inequalities allowing for both short- and long-range dependence and also heavy-tailed distributions. Jiang (2009) focus on a triplex inequality in which the dependence term is characterised by a mixing coefficient, whereas Chen and Wu (2018) consider a linear process on independent innovations, i.e., dependence comes from the linear weights. Both authors consider a martingale representation of the process, and the tail is a combination of an exponential and second term that can be polynomial or sub-exponential as well. We use a new concentration for martingales in Lesigne and Volný (2001); Fan et al. (2012) and Fan et al. (2012), which yield optimal bounds for sub-Weibull tails, and account for dependence and dimensionality using a mixingale dependence coefficient and Boole inequality, respectively.

Concentration bounds similar to ours admit a wide range of applications. For example, they are used to derive oracle estimation bounds for VARMA models (Wilms et al., 2021), misspecified VAR(p) models with l_1 penalty (Wong et al., 2020; Masini et al., 2022), and l_1 penalised Yule-Walker estimation (Han et al., 2015; Reuvers and Wijler, 2024; Wang and Tsay, 2023). Furthermore, these inequalities are essential for the derivation of statistical properties of methods for inference in high-dimensional time series, such as desparsified inference in VAR(p) models (Adamek et al., 2023a) and multiplier bootstrap in high-dimensional time series (Krampe et al., 2021; Adamek et al., 2023b). Finally, the concentration inequality for lag-h autocovariances is also used in the estimation of long-term covariance matrices and spectral density (Zhang and Wu, 2017; Li and Liao, 2020; Babii et al., 2022; Fan et al., 2023).

1.1 Notation

For any vector $\mathbf{b} = (b_1, ..., b_k)' \in \mathbb{R}^k$ and $p \geq 1$, $|\mathbf{b}|_p$ is the ℓ_p vector-norm with $|\mathbf{b}|_p = (\sum_{i=1}^k |b_i|^p)^{1/p}$ for $p \in [1, \infty)$ and $|\mathbf{b}|_\infty = \max_{1\leq i\leq k} |b_i|$. For a random variable X, $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$ for $p \in [1, \infty)$ and $\|X\|_\infty = \inf\{a \in \mathbb{R} : \Pr(|X| \geq a) = 0\}$. For an $(m \times n)$-dimensional matrix $\boldsymbol{A}$ with elements a_{ij}, $|||\boldsymbol{A}||| = \sqrt{\Lambda_{\max}(\boldsymbol{A'A})}$ is its spectral norm, $|||\boldsymbol{A}|||_1 = \max_{1\leq j\leq n} \sum_{i=1}^m |a_{ij}|$ and $|||\boldsymbol{A}|||_\infty = \max_{1\leq i\leq m} \sum_{j=1}^n |a_{ij}|$ are the induced l_∞ and l_1 norms, respectively. The maximum entry-wise norm of $\boldsymbol{A}$ is $|||\boldsymbol{A}|||_{\max} = |\mathsf{vec}\,(\boldsymbol{A})|_\infty = \max_{i,j} |a_{ij}|$ and its Frobenius norm is $|||\boldsymbol{A}|||_F = |\mathsf{vec}\,(\boldsymbol{A})|_2 = \sqrt{\mathsf{trace}\,(\boldsymbol{A'A})}$. The minimum and maximum eigenvalues of a square matrix A are $\Lambda_{\min}(\boldsymbol{A})$ and $\Lambda_{\max}(\boldsymbol{A})$, respectively. We shall use c, c_1, c_2, ... as generic constants that may change values each time they appear. A constant with a symbolic subscript is used to emphasise the dependence of the value on the subscript.

The vector $\boldsymbol{e}_i$ is a canonical basis vector of adequate dimension.

2. Preliminaries

In this section, we introduce the dependence and tail conditions used in the paper, and develop a concentration inequality for the ℓ_∞ vector norm (sup-norm) of sums of dependent random variables.

2.1 Sub-Weibull random variables

Sub-Weibull random variables accommodate a wide range of tail behaviour, including variables with heavy tails for which the moment-generating function does not exist, subexponential, and sub-Gaussian. Sub-Weibull random variables are studied in Wong et al. (2020); Vladimirova et al. (2020); Götze et al. (2021); Kuchibhotla and Chakrabortty (2022); Zhang and Wei (2022); Bong and Kuchibhotla (2023), and it has also appeared in the context of entries of a random matrix in Tao and Vu (2013, Condition $C0$).

Definition 1 (Sub-Weibull(α) random variable)**.** Let $\alpha > 0$. A sub-Weibull(α) random variable X satisfies $\Pr(|X| > x) \leq 2\exp\{-(x/K)^\alpha\}$,for all $x > 0$ and some $K > 0$.

There are equivalent definitions in terms of moments and moment generating functions of a $|X|^\alpha$ and Orlicz (quasi-) norms, denoted $\|\cdot\|_{\psi_\alpha} :=$

$\inf\{c > 0 : \mathbb{E}\psi_\alpha(|\cdot|/c) \leq 1\}$, with $\psi_\alpha(\cdot) = \exp(x^\alpha) - 1$ and $\alpha > 0$. In the online supplement, we discuss Orlicz norms and sub-Weibull random variables in detail. An important tail bound that follows after Markov's inequality is

$$\Pr\left(\max_{1\leq i\leq n} |X_i| > x\right) \leq \exp\left(-\frac{x^\alpha}{(c_1 \max_{1\leq i\leq n} \|X_i\|_{\psi_\alpha})^\alpha \log(1+2n)}\right). \quad (2.1)$$

Our approach differs from that of Kuchibhotla and Chakrabortty (2022); Zhang and Wei (2022) and Bong and Kuchibhotla (2023) in several key aspects. Unlike the aforementioned authors, who assume independence, we consider a dependent random sequence, which significantly alters the proof methodology. Furthermore, we employ the classical Orlicz norm approach (van der Vaart and Wellner, 1996, Sec. 2.2), in contrast to the generalised Bernstein-Orlicz norm used by the authors. Despite these differences, we demonstrate that, under our assumptions, our rate is nearly optimal, albeit slower than that for sums of independent sub-Weibull random variables.

2.2 Mixingales

We characterise the dependence in the process using the moments and conditional moments of the series. This mode of dependence is weak in the sense that it only requires the conditional moments to converge to their marginals in L_p as we have conditioned further in the past. Some classical measures of dependence, such as strong mixing, imply mixingale dependence.

Definition 2 (Mixingale)**.** Let $\{X_t\}$ be a causal stochastic process, and let $\{\mathcal{F}_t\}$ be an increasing sequence of σ fields in such a way that X_t is $\mathcal{F}_t$ measurable. The process $\{X_t\}$ is an L_p-mixingale process with respect to $\{\mathcal{F}_t\}$ if there exists a decreasing sequence $\{\rho_m\}$ and a constant c_t satisfying $\|\mathbb{E}[X_t|\mathcal{F}_{t-m}] - E[X_t]\|_p \leq c_t\rho_m$.

Mixingales can be represented as as a sum of a martingale difference terms and a conditional expectation:

$$\sum_{t=1}^{T}(\mathbf{X}_t - \mathbb{E}[\mathbf{X}_t]) = \sum_{i=1}^{m}\left(\sum_{t=1}^{T} V_{i,t}\right) + \sum_{t=1}^{T}\mathbb{E}[\mathbf{X}_t - \mathbb{E}(\mathbf{X}_t)|\mathcal{F}_{t-m}], \qquad (2.2)$$

where $V_{i,t} = \mathbb{E}[\mathbf{X}_t|\mathcal{F}_{t-i+1}] - \mathbb{E}[\mathbf{X}_t|\mathcal{F}_{t-i}]$ is a martingale difference process.

2.3 Concentration inequality

We present the triplex inequality, based on (Jiang, 2009), followed by a discussion of the concentration rates.

Theorem 1 (Concentration for sub-Weibul mixingale processes)**.** *Let* $\{\mathbf{X}_t = (X_{1t}, ..., X_{nt})'\}$ *be a causal stochastic process and* $\{\mathcal{F}_t\}$ *an increasing sequence of* σ*-algebras such that* $\mathbf{X}_t$ *is* $\mathcal{F}_t$ *measurable, and write* $\mathbf{S}_k = \sum_{t=1}^{k}(\mathbf{X}_t - \mathbb{E}[\mathbf{X}_t])$*. Suppose that each* $\{X_{jt}, \mathcal{F}_t\}$ *is* L_p*-mixingale with constants* c_{jt} *and*

$\{\rho_{jm}\}$, *and let* $\rho_m = \max_{1\le j\le n}\rho_{jm}$ *and* $\bar{c}_T = \max_{1\le j\le n} T^{-1}\sum_{t=1}^T c_{jt}$. *Furthermore, suppose that* $\max_{i,t}\|X_{it}\|_{\psi_\alpha} < c_{\psi_\alpha} < \infty$. *Then, for any natural* m *and scalar* $M > 0$*:*

$$\begin{aligned}\Pr\left(\max_{1\le k\le T}|\mathbf{S}_k|_\infty > Tx\right) &\le 2mn\exp\left(-\frac{Tx^2}{8(Mm)^2 + 2xMm}\right)\\ &\quad + 4m\exp\left(-\frac{M^\alpha}{c_1\log(3nT)}\right) + \frac{2^p}{x^p}n\rho_m^p\bar{c}_T,\end{aligned} \tag{2.3}$$

where $c_1 := (2c_{\psi_\alpha}/\log(1.5))^\alpha$.

It is natural to ask how tight these bounds are in terms of rate. Lesigne and Volný (2001) and Fan et al. (2012) show that martingales enjoy slower concentration rates. It happens because to achieve faster rates we would require further restriction on its quadratic variation process. Let $\{X_t\}$ be a strictly stationary and ergodic martingale difference sequence with $\sup_t \mathbb{E}[e^{|X_t|^\alpha}] < \infty$, then $P(\sum_{i=1}^T X_i > Tx) \ge O(e^{-cT^{\phi_\alpha}})$ where $\phi_\alpha = \alpha/(\alpha+2)$ (Theorem 2.1 Fan et al., 2012). If $\alpha = 2$, i.e., X_i has sub-Gaussian tails, the rate is $O(e^{-cT^{1/2}})$, and, similarly, if X_i has sub-exponential tails then the rate is $O(e^{-cT^{1/3}})$. It contrasts with the classical Azuma-Hoeffding inequality, valid for bounded processes, which yields a rate of $O(e^{-cT})$.

Consider the case where the dependence term has finite memory, i.e., there is some $m^* \ge 1$ such that $\rho_m = 0$ for all $m \ge m^*$, suppose $\log n \lesssim T^{\alpha/(\alpha+4)} \wedge T^{\phi_\alpha}\log(T)^{2/(\alpha+2)}$, and take $M = (Tx^2\log(nT))^{1/(\alpha+2)}$. Then the right-hand side of equation (2.3) is $O\left(e^{-cr_T^{\phi_\alpha}}\right)$ where $r_T = T/\log(nT)^{2/\alpha}$.

It means that under finite dependence the rate is nearly optimal, by a factor of $\log(nT)^{1-\phi_\alpha}$. Specifically, in the sub-Gaussian case the rate is $O(e^{-cT^{1/2}/\log(nT)^{1/2}})$, in the subexponential case the rate is $O(e^{-cT^{1/3}/\log(nT)^{2/3}})$, and in the sub-Weibull case with $\alpha = 0.5$ the rate is $O(e^{-cT^{1/5}/\log(nT)^{4/5}})$ compared to $O(e^{-cT^{1/5}})$.

If we drop the finite dependence assumption in favour of a *sub-Weibull* decay to the mixingale dependence rate $\rho_m \leq e^{m^\gamma/(pc_\rho)}$ for some $\gamma > 0$, the convergence rates will change accordingly. Let $m = M^{\alpha/\gamma}\log(3nT)^{1/\gamma}$ and $M = (Tx^2)^{\phi_{\alpha,\gamma}/\alpha}\log(3nT)^{\phi_{\alpha,\gamma}(2+\gamma)/\alpha\gamma}$ with $\phi_{\alpha,\gamma} = \alpha\gamma/(2\alpha+\gamma(2+\alpha))$. Then, if $\log(3nT) \lesssim T^{(2/\phi_{\alpha,\gamma}-(2+\gamma)/\alpha\gamma)^{-1}}$, the right hand side of equation (2.3) is $O(e^{-cT^{\phi_{\alpha,\gamma}}/\log(nT)^{1-\gamma^{-1}(2+\gamma)\phi_{\alpha,\gamma}}})$:

$$\Pr\left(\max_{1\leq k\leq T}|\mathbf{S}_k|_\infty > Tx\right) \leq 2\exp\left(-\frac{(Tx^2)^{\phi_{\alpha,\gamma}}}{16+4x^{\phi_{\alpha,\gamma}}T^{-\frac{1-\phi_{\alpha,\gamma}}{2}}\log(3nT)^{-\frac{\phi_{\alpha,\gamma}}{\alpha}}}\right) + (4+2^p x^{-p}\bar{c}_T)\exp\left(-\frac{(Tx^2)^{\phi_{\alpha,\gamma}}}{c_1\log(3nT)^{1-\frac{2+\gamma}{\gamma}\phi_{\alpha,\gamma}}}\right),$$

where $c_1 = (2c_{\psi_\alpha}/\log(1.5))^\alpha \vee 2c_\rho$.

In order to analyse this term, we have to impose rates for both α and γ. First note that as $\gamma\to\infty$, $\phi_{\alpha,\infty} = \phi_\alpha$ and $1-\phi_{\alpha,\gamma}(2+\gamma)\gamma^{-1} \to 1-\phi_\alpha$, recovering the finite dependence case, as expected. Now we consider $\gamma = 2$, that is, $\rho_m^p \leq e^{-m^2/c_\rho}$. In the sub-Gaussian case, $\phi_{\alpha,\gamma} = \phi_{2,2} = 1/3$ and the rate is $O(e^{-cT^{1/3}/\log(nT)^{1/3}})$, that is, we pay a price of $T^{1/6}\log(nT)^{1/3}$ when compared to the optimal rate for martingales and $(T/\log(nT))^{1/6}$ when compared to the rates obtained for the case of high-dimensional, limited dependence. In the subexponential case the rate is $O(e^{-cT^{1/4}/\log(nT)^{1/4}})$ which compares to $O(e^{-cT^{1/3}})$ and $O(e^{-cT^{1/3}/\log(nT)^{2/3}})$ for the martingale and finite dependence in the high-dimensional case, respectively. Increasing the dependence to $\gamma = 1$, yields a rate of $O(e^{-cT^{1/4}/\log(nT)^{1/4}})$ for the sub-Gaussian tail case and $O(e^{-cT^{1/5}/\log(nT)^{2/5}})$ for the sub-exponential tail case. Finally, considering a sub-Weibull tail with parameter $\alpha = 0.5$ yields a rate of $O(e^{-cT^{1/7}/\log(nT)^{2/7}})$, compared to $O(e^{-cT^{1/5}/\log(nT)^{4/5}})$ in the case with a high dimension and finite dependence.

In high-dimensional statistics, we are often interested in the situation where $n \to \infty$ at some rate depending on T. In the next corollary we present a useful result in which the rate is delegated to a parameter τ, which can be made dependent on n.

Corollary 1 (Sub-Weibull Concentration)**.** *According to the assumptions of Theorem 1, suppose that $\rho_m \leq e^{-m^\gamma/(pc_\rho)}$ for some $\gamma > 0$. Then, for any $\tau > 0$ and all $T \geq \log(n) + \tau$:*

$$\begin{aligned}\Pr\left(\max_{1\leq k\leq T} |\mathbf{S}_k|_\infty > Tx\right) &\leq 2c_\rho^{\frac{1}{\gamma}}(\log(n)+\tau)^{\frac{1}{\gamma}}e^{-\tau} + 2^p\bar{c}_T x^{-p}e^{-\tau} \\ &\quad + 4c_\rho^{\frac{1}{\gamma}}(\log(n)+\tau)^{\frac{1}{\gamma}} e^{\frac{-(x\sqrt{T})^\alpha}{c_1\log(3nT)(\log(n)+\tau)^{\frac{\alpha}{2}+\frac{\alpha}{\gamma}}}},\end{aligned} \tag{2.4}$$

where $c_1 := (8c_{\psi_\alpha}c_\rho^{1/\gamma}/\log(1.5))^\alpha$.

Proof. The proof follows by setting $m := c_\rho^{\frac{1}{\gamma}}(\log(n)+\tau)^{\frac{1}{\gamma}}$ and $mM := x\sqrt{T}/4\sqrt{\log(n)+\tau}$. □

The constants that appear in the inequality are not optimised. A simpler bound for high-dimensional vectors, that is, large n, follows by setting $\tau = \log(n)$. Let $c_1 = 2(2c_\rho)^{1/\gamma}$, $c_2 := \beta 2^{\frac{7}{2}\alpha+\frac{\alpha}{\gamma}}(c_{\psi_\alpha} c_\rho^{1/\gamma}/\log(1.5))^\alpha$ and suppose that $T > 2\log(n)$ and $T \le n^{\beta-1}/3$ for some $\beta > 1$. Then,

$$\Pr\left(\max_{1\le k\le T} |\mathbf{S}_k|_\infty > Tx\right) \le c_1 \frac{\log(n)^{1/\gamma}}{n} + \frac{2^p \bar{c}_T}{nx^p} + 2c_1\log(n)^{1/\gamma} e^{\frac{-(x\sqrt{T})^\alpha}{c_2\log(n)^{1+\frac{\alpha}{2}+\frac{\alpha}{\gamma}}}}.$$

This inequality sheds some light on the rate of increase in n that we can expect so that the right-hand side converges to zero. If $\alpha = 2$, the sub-Gaussian tail case, we have $\log(n) = o(T^{\gamma/(2\gamma+2)})$, in the sub-exponential tail case, $\alpha = 1$, $\log(n) = o(T^{\gamma/(3\gamma+2)})$, and in the case with a heavy tail with $\alpha = 0.5$, we have $\log(n) = o(T^{\gamma/(5\gamma+2)})$. In all cases, a strong dependence will guide the rate. If γ is very large, the rates are, respectively, close to $o(T^{1/2})$, $o(T^{1/3})$, and $o(T^{1/5})$. This concentration will be used in the next sections to handle more complex stochastic processes.

Remark 1 (Extension to other norms)**.** Denote $|\cdot|_\psi$ as some norm on $\mathbb{R}^n$ and let $l_n = \sup_{\mathbf{b}\in\mathbb{R}^n} |\mathbf{b}|_\psi/|\mathbf{b}|_\infty$ denote the compatibility constant between this norm and the supremum norm on the same space. Then, $P(|S_T|_\psi > Tx) \le P(|S_T|_\infty > Tx/l_n)$. The compatibility constant l_n will have an effect on convergence rates. Suppose we are in a finite dependence situation with subexponential tails. Then, we have $P(|S_T|_\psi > Tx) \leq O(e^{-cT^{1/3}/(l_n \log(nT))^{2/3}})$, which can be very restrictive depending on l_n. Effectively, let the ψ norm be the ℓ_p norm on R^n. Then $l_n = n^{1/p}$ and the new convergence rate is $P(|S_T|_p > Tx) \leq O(e^{-cT^{1/3}/(n^{1/p} \log(nT))^{2/3}})$, which is conservative for high-dimensional vectors.

3. Linear processes with dependent innovations

In this section, we extend the concentration inequality in Theorem 1 to linear stochastic processes with sub-Weibull tails. We first define the multivariate linear process followed by the Beveridge-Nelson (BN) decomposition (Beveridge and Nelson, 1981). The latter decomposes a linear process into simpler components that we can analyse using existing techniques and tools (Phillips and Solo, 1992). In particular, we will use Theorem 1 to bound the first term and Equation (2.1) to bound the remaining terms.

Let $\{C_j\}$ denote the $n \times n$ matrices and $\{\mathbf{X}_t\}$ denote a centered stochastic process taking values in $\mathbb{R}^n$. The linear process $\mathbf{Y}_t$ is

$$\mathbf{Y}_t = \sum_{j=0}^{\infty} C_j \mathbf{X}_{t-j} = C(L)\mathbf{X}_t, \tag{3.5}$$

where $C(L) = \sum_{j=0}^{\infty} C_j L^j$ is a lag polynomial and L is the lag operator satisfying $L\mathbf{X}_t = \mathbf{X}_{t-1}$.

The tail behaviour of $\mathbf{Y}_t$ is not directly inherited by $\mathbf{X}_t$, unless the conditions in $\{C_j\}$, discussed in Lemma 1, are satisfied.

Lemma 1. *Let $\{\boldsymbol{a}_j\}$ denote a sequence of elements in $\mathbb{R}^n$ each with a finite L_1 norm, $\{\mathbf{Z}_t\}$ a sequence of random vectors satisfying $\sup_{|\boldsymbol{a}|_1\leq 1}\|\boldsymbol{a}'\mathbf{Z}_t\|_\psi \leq c_t < \infty$ where $\|\cdot\|_\psi$ is a norm and c_ts are positive constants, and let $W_t = \sum_{j=0}^{\infty}\boldsymbol{a}_j' Z_{t-j}$. Then, $\|W_t\|_\psi \leq \sum_{j=0}^{\infty}|\boldsymbol{a}_j|_1 c_{t-j}$, provided $\sum_{j=0}^{\infty}|\boldsymbol{a}_j|_1 < \infty$.*

The BN decomposition represents the matrix polynomial $C(z)$ as

$$C(z) = C(1) - (1-z)\widetilde{C}(z), \tag{3.6}$$

where $\widetilde{C}(z) = \sum_{j=0}^{\infty}\widetilde{C}_j z^j$ and $\widetilde{C}_j = \sum_{k=j+1}^{\infty} C_k$. Then, the linear process is

$$\mathbf{Y}_t = C(L)\mathbf{X}_t = C(1)\mathbf{X}_t - \widetilde{\mathbf{X}}_t + \widetilde{\mathbf{X}}_{t-1}, \tag{3.7}$$

where $\widetilde{\mathbf{X}}_t = \widetilde{C}(L)\mathbf{X}_t$.

If $\mathbf{X}_t$ is sub-Weibull(α) then each component of $\widetilde{\mathbf{X}}_t$ is also sub-Weibull(α), provided $\sum_{j=1}^{\infty} j|\boldsymbol{e}_i' C_j|_1 < \infty$. Let $\{\boldsymbol{e}_i = (0,...,0,1,0,...,0)', i = 1,...,n\}$, with 1 on the i^{th} element of the vector, denote the canonical basis vectors for $\mathbb{R}^n$. When applying the lemma 1 to $\max_{i\leq n}\|\boldsymbol{e}_i'\widetilde{\mathbf{X}}_t\|_{\psi_\alpha}$, we substitute $\boldsymbol{a}_j = \boldsymbol{e}_i'\widetilde{C}_j$, which means that we require $\sum_{j=1}^{\infty} j|\boldsymbol{e}_i' C_j|_1 < \infty$ for all $1 \leq i \leq n$, and $\sup_{|\mathbf{b}|\leq 1}\|\mathbf{b}'\mathbf{X}_t\|_\psi < \infty$.

Theorem 2 (Concentration inequality for Linear Processes). *Let* $\{\mathbf{X}_t = (X_{1t}, ..., X_{nt})'\}$ *be a centered sub-Weibull(α) causal process taking values in* $\mathbb{R}^n$*, with sub-Weibull constant* c_{ψ_α}*, and let* $\{\mathcal{F}_t\}$ *be an increasing sequence of* σ*-algebras such that* $\mathbf{X}_t$ *is* $\mathcal{F}_t$ *measurable. Assume that, for each* $i = 1, ..., n$*,* $\{X_{it}, \mathcal{F}_t\}$ *is* L_p*-mixingale with positive constants* $\{c_{it}\}$ *and decreasing sequence* $\{\rho_{im}\}$ *and write* $\bar{c}_T = \max_{1\le i\le n} T^{-1}\sum_{t=1}^{T} c_{jt}$ *and* $\rho_m = \max_{1\le i\le n}\rho_{im}$.

Write the linear process $\mathbf{Y}_t = C(L)\mathbf{X}_t$*, where* $\{C_j\}$ *is a sequence of square matrices that satisfy* $\max_{1\le i\le n}\sum_{j=1}^{\infty} j|\boldsymbol{e}_i' C_j|_1 \le \tilde{c}_\infty < \infty$ *and denote* $c_\infty = |||C(1)|||_\infty$.

Then, for any $0 < a < 1$*,* $T > 0$*,* $M > 0$ *and* $m = 1, 2, ...,$ *we have*

$$\Pr\left(\max_{1\le k\le T}\left|\sum_{t=1}^{k}\mathbf{Y}_t\right|_\infty \ge Tx\right) \le 2mn\exp\left(-\frac{T(ax)^2}{8c_\infty^2(Mm)^2 + 2c_\infty axMm}\right) + 4m\exp\left(-\frac{M^\alpha}{c_1\log(3nT)}\right) + \frac{(2c_\infty)^p}{(ax)^p}n\rho_m^p\bar{c}_T + \exp\left(-\frac{((1-a)Tx)^\alpha}{c_2\log(1+2n)}\right) \tag{3.8}$$

where $c_1 := (2c_{\psi_\alpha}/\log(1.5))^\alpha$ *and* $c_2 := (2c_{\psi_\alpha}\tilde{c}_\infty/\log(1.5))^\alpha$

The next corollary removes the dependence of the bound on M, m, and a, replacing them with appropriate sequences and constants.

Corollary 2. *Under the assumptions of Theorem 2, let* $\rho_m \leq e^{-m^\gamma/(pc_\rho)}$ *for some* $\gamma > 0$*. Then, for any* $\tau > 0$ *and* $T > \log(n) + \tau$*,*

$$\Pr\left(\max_{1\leq k\leq T}\left|\sum_{t=1}^{k}\mathbf{Y}_t\right|_\infty \geq Tx\right) \leq c_1(\log(n)+\tau)^{1/\gamma}e^{-\tau} + e^{-\frac{(xT)^\alpha}{c_2\log(1+2n)}} + \frac{(4c_\infty)^p\bar{c}_T}{x^p}e^{-\tau} + 2c_1(\log(n)+\tau)^{1/\gamma}e^{-\frac{(x\sqrt{T})^\alpha}{c_3\log(3nT)(\log(n)+\tau)^{\frac{\alpha}{2}+\frac{\alpha}{\gamma}}}},$$

where $c_1 := 2(c_\rho)^{1/\gamma}$, $c_2 := (4c_{\psi_\alpha}\tilde{c}_\infty/\log(1.5))^\alpha$*, and* $c_3 := (8c_{\psi_\alpha}c_\infty c_\rho^{1/\gamma}/\log(1.5))^\alpha$.

Proof. The result follows after replacing $c_\infty mM = ax\sqrt{T}/4\sqrt{\log(n)+\tau}$, $m = (c_\rho(\log(n)+\tau))^{1/\gamma}$, and $a = 1/2$. The lower bound on T requires $(\log(n)+\tau)/T < 1$. □

Let $\tau = \log(n)$ and assume that $T \leq n^{\beta-1}/3$ for some $\beta > 1$. If we take $\beta = 1 + \log(3T)/\log(n)$, we have equality and as long as $\log(T)/\log(n) \not\to \infty$, as both n and T increase, β can be taken sufficiently large. This is often the case in high-dimensional statistics where n is either larger or close to T in rate. A simplified inequality is

$$\Pr\left(\max_{1\leq k\leq T}\left|\sum_{t=1}^{k}\mathbf{Y}_t\right|_\infty \geq Tx\right) \leq c_1\frac{\log(n)^{1/\gamma}}{n} + e^{-\frac{(xT)^\alpha}{c_2\log(1+2n)}} + \frac{(4c_\infty)^p\bar{c}_T}{nx^p} + 2c_1\log(n)^{1/\gamma}e^{-\frac{(x\sqrt{T})^\alpha}{c_3\log(n)^{1+\frac{\alpha}{2}+\frac{\alpha}{\gamma}}}}, \tag{3.9}$$

where the constants $c_1 := 2^{1+1/\gamma}(c_\rho)^{1/\gamma}$, $c_2 := (4c_{\psi_\alpha}\tilde{c}_\infty/\log(1.5))^\alpha$, and $c_3 := \beta 2^{\frac{7\alpha}{2}+\frac{\alpha}{\gamma}}(c_{\psi_\alpha}c_\infty c_\rho^{1/\gamma}/\log(1.5))^\alpha$. Therefore, the same rates as discussed in the previous section are recovered.

Remark 2. Some comments on the previous results are in order:

1. In Theorems 1 and 2, the stochastic process $\{\mathbf{X}_t\}$ is not assumed to be stationary, only the tail and dependence properties hold. For instance, we can have $\mathbf{X}_t$ a heteroskedastic sequence with $\mathbb{E}[\mathbf{X}_t\mathbf{X}_t'] = \Sigma_t$ with eigenvalues bounded away from zero and infinity for all $t \in \mathbf{Z}$;

2. if X_t is not centered, we must work with $\mathbf{Z}_t = \mathbf{X}_t - \mathbb{E}[\mathbf{X}_t]$, in which case $\mathbf{Y}_t = \mathbb{E}[\mathbf{Y}_t] + \sum_{j=0}^{\infty} C_j\mathbf{Z}_t$, where $\mathbb{E}[\mathbf{Y}_t] = \sum_{j=0}^{\infty} C_j\mathbb{E}[\mathbf{X}_t]$. Note that the values of $\mathbb{E}[\mathbf{X}_t]$ can also change for each t, so we can incorporate deterministic trends and seasonality into this process. Nevertheless, if $\bar{c}_T < \infty$ for all T, then $T^{-1/2}\log(n)^{1/2+1/\alpha+1/\gamma}\left|\sum_{t=1}^{T}\mathbf{Y}_t\right|_{\infty}$ is stochastically bounded;

3. the sequence $\bar{c}_T$ may increase with T, in which case limiting arguments used to obtain an expected rate of increase in n as to make probability bound converge to zero will have to accommodate it;

4. similarly, c_∞ and $\tilde{c}_\infty$ may depend on n, in which case c_2 in Theorem 2 is no longer a constant. To illustrate the effect of this change, let $a = 1/2$ and suppose that c_∞ is fixed but $\tilde{c}_\infty \leq c_0(\log(1+2n))^{1/\alpha}$: the last term on the right-hand side of equation (3.8) changes from $\exp[(Tx)^\alpha/c_2\log(1+2n)]$ to $\exp[(Tx)^\alpha/c_2'(\log(1+2n))^2]$, for two dis-

tinct constants c_2 and c_2';

Remark 3 (Martingale difference)**.** A simpler but interesting case occurs when the process $\{\mathbf{X}_t, \mathcal{F}_{t-1}\}$ is a martingale difference. We take $m = 1$ and $\rho_m = 0$ in Theorem 2, which is equivalent to taking $\gamma \to \infty$, which produces rates for n and a tighter concentration bound.

Corollary 3 (Concentration for linear processes on martingale differences)**.** *Let $\{\mathbf{X}_t = (X_{1t}, ..., X_{nt})'\}$ be a centred sub-Weibull(α) causal process taking values in $\mathbb{R}^n$, with sub-Weibull constant c_{ψ_α}, and let $\{\mathcal{F}_t\}$ be an increasing sequence of σ-algebras such that $\mathbf{X}_t$ is $\mathcal{F}_t$ measurable, and assume that $\{\mathbf{X}_t, \mathcal{F}_{t-1}\}$ is a martingale difference process.*

Write the linear process $\mathbf{Y}_t = C(L)\mathbf{X}_t$, where $\{C_j\}$ is a sequence of square matrices that satisfy $\max_{1\le i\le n}\sum_{j=1}^{\infty} j|\boldsymbol{e}_i'C_j|_1 \le \tilde{c}_\infty < \infty$ and denote $c_\infty = |||C(1)|||_\infty$.

Then, for any $T > 0$ and $M > 0$,

$$\Pr\left(\max_{1\le k\le T}\left|\sum_{t=1}^{k}\mathbf{Y}_t\right|_\infty \ge Tx\right) \le 2n\exp\left(-\frac{T(ax)^2}{8c_\infty^2M^2 + 2c_\infty axM}\right) + 4\exp\left(-\frac{M^\alpha}{c_1\log(3nT)}\right) + \exp\left(-\frac{((1-a)Tx)^\alpha}{c_2\log(1+2n)}\right)$$

where $c_1 := (2c_{\psi_\alpha}/\log(1.5))^\alpha$ and $c_2 := (2c_{\psi_\alpha}\tilde{c}_\infty/\log(1.5))^\alpha$

Proof. We take $m = 1$ and $\rho_m = 0$ in Theorem 2. □

Following the same arguments in Equation (3.9), an equivalent, simpler, probability bound is

$$\Pr\left(\left|\sum_{t=1}^{T}\mathbf{Y}_t\right|_\infty \geq Tx\right) \leq \frac{2}{n} + e^{-\frac{(xT)^\alpha}{c_1\log(1+2n)}} + 4e^{-\frac{(x\sqrt{T})^\alpha}{2^{1+\frac{\alpha}{2}}\beta c_2\log(n)^{1+\frac{\alpha}{2}}}}, \qquad (3.10)$$

where $c_1 := (4c_{\psi_\alpha}\tilde{c}_\infty/\log(1.5))^\alpha$, and $c_2 := (8c_{\psi_\alpha}c_\infty/\log(1.5))^\alpha$

Remark 4 (Integrated processes)**.** In this example we show how integrated processes violate the conditions in Theorem 2. Let $\{X_t, \mathcal{F}_{t-1}\}$ be a martingale difference process with identity covariance matrix and $C(L) = I_n$, and suppose that $\{\mathbf{b}'\mathbf{X}_t\}$ has subexponential tails for all unit vectors $\mathbf{b} \in \mathbb{R}^n$. Let $\mathbf{Y}_1 = \mathbf{X}_1$ and $(1-L)\mathbf{Y}_t = \mathbf{X}_t$, for $t = 2, \cdots, T$. Thus, $\mathbf{Y}_t = \mathbf{S}_t = \sum_{i=1}^{t}\mathbf{X}_t$ is a martingale process with respect to $\mathcal{F}_{t-1}$. It follows from the Burkhölder inequality that $\|S_{it}\|_p = O(t^{1/2p})$ and $\|\mathbb{E}[S_{it}|\mathcal{F}_{t-m}]\|_p = O((t-m)^{1/2p})$ if $m < t$ and $\|\mathbb{E}[S_{it}|\mathcal{F}_{t-m}]\|_1 = 0$ otherwise, that is, $\rho_m = \max(0,(1-m/t))^{1/2p}$ depends on t and $c_t = t^{1/2p}$. It follows from Fan et al. (2012, Theorem 2.1) and Lesigne and Volný (2001, Theorem 3.2) that $P(S_t > x) \lesssim e^{-ct^{-1/3}x^{2/3}}$, and therefore $c_\psi = O(t^{1/3})$ also depends on t.

4. Empirical lag-h autocovariance matrices

In this section, we will examine the concentration properties of the empirical lag h autocovariance matrix of sub-Weibull linear processes under the maximum entry-wise norm. This concentration property is important in the LASSO estimation of large vector auto-regressive models and in the use of the multiplier bootstrap. By studying the concentration of the empirical autocovariance matrix, we can better understand the behaviour of these statistical methods in high-dimensional settings.

Let $\{\mathbf{X}_t\}$ denote a centred, sub-Weibull, causal stochastic process taking values on $\mathbb{R}^n$, and $\mathbf{Y}_t = C(L)\mathbf{X}_t$, $t = 1, \cdots, T$, a dependent sequence of random vectors. Let

$$\widehat{\Gamma}_T(h) := \frac{1}{T}\sum_{t=h+1}^{T} \mathbf{Y}_t\mathbf{Y}'_{t-h} \quad \text{and} \quad \Gamma_T(h) := \mathbb{E}[\widehat{\Gamma}_T(h)] = \frac{1}{T}\sum_{t=h+1}^{T} \mathbb{E}[\mathbf{Y}_t\mathbf{Y}'_{t-h}]. \tag{4.11}$$

Our goal is to find a bound for

$$\Delta_T(h) = \left|\!\left|\!\left| \widehat{\Gamma}_T(h) - \Gamma_T(h) \right|\!\right|\!\right|_{\max} = \left| \mathsf{vec}\left(\widehat{\Gamma}_T(h) - \Gamma_T(h)\right) \right|_{\infty}, \tag{4.12}$$

which is the maximum element in absolute value of the matrix.

Theorem 3. *Let $\{\mathbf{X}_t = (X_{1t}, ..., X_{nt})'\}$ be a centred sub-Weibull(α), causal process taking values in $\mathbb{R}^n$, and let $\{\mathcal{F}_t\}$ be an increasing sequence of σ-algebras such that $\mathbf{X}_t$ is $\mathcal{F}_t$ measurable.*

Write $\boldsymbol{\eta}_t(k) = \mathsf{vec}\left(\mathbf{X}_t\mathbf{X}'_{t-k}\right)$ and the stochastic process $\{\boldsymbol{\eta}_t(k) = (\eta_{1t}(k), ..., \eta_{n^2t}(k))'\}$ for $k = 0, 1, ...$ The processes $\{\eta_{it}(k), \mathcal{F}_t\}$ are L_1-mixingale with constants c_{it} and decreasing sequences ρ_{im}, for each $k = 0, 1, ...$ and $i = 1, ..., n^2$. Let

$\bar{c}_T = \max_{1\le i\le n^2} T^{-1}\sum_{t=1}^T c_{it}$ *and* $\rho_m = \max_{1\le i\le n}\rho_{im} \le e^{-m^\gamma/c_\rho}$.

Finally, let the linear process $\mathbf{Y}_t = C(L)\mathbf{X}_t$*, where* $\{C_j\}$ *is a sequence of square matrices, and define the following finite constants:*

a. $\tilde{c}_{2,\infty} := \sum_{j=1}^\infty j|||C_j|||^2$;

b. $c_h := \max_{1\le k\le h}\left|\left|\left|\sum_{j=0}^\infty C_{j+k}\otimes C_j\right|\right|\right|_\infty$;

c. $c_\infty := \max_{1\le k\le h}\left|\left|\left|\sum_{i=0}^\infty (\sum_{j=i+k}^\infty C_j)\otimes C_i\right|\right|\right|_\infty$;

d. $\tilde{c}_\infty := \max_{1\le i\le n^2}\sum_{j=1}^\infty j\,|\sum_{k=0}^\infty \boldsymbol{e}_i'(C_{j+k}\otimes C_k)|_1$.

Let $\Delta_T(h)$ *be as defined in equations* (4.11) - (4.12)*. Then, for each* n *and* T *that satisfies* $T > 4\log(n)$ *and* $3T < n^{\beta-1}$ *for some* $\beta > 1$*:*

$$\begin{aligned}\Pr(\Delta_T(h) > x) \le\ & c_1\frac{\log(n)^{1/\gamma}}{n^2} + \frac{c_2\bar{c}_T}{n^2 x} \\ & + c_3\log(n)^{1/\gamma} e^{-\frac{(x\sqrt{T})^{\alpha/2}}{c_4\log(n)^{1+\frac{\alpha}{4}+\frac{\alpha}{2\gamma}}}} + 4e^{-\frac{(xT)^{\alpha/2}}{c_5\log(n)}},\end{aligned} \tag{4.13}$$

where the constants $c_1, ..., c_5$ *depend only on* β*,* $c_\rho, c_{\psi_{\alpha/2}}, \tilde{c}_{2,\infty}, c_h, c_\infty, \tilde{c}_\infty$*, but not on* n *or* T*.*

Conditions (a) - (d) regulate the persistence of the linear weight matrices $\{C_j\}$ by imposing summation conditions. Conditions (b) - (d) are satisfied by the simpler bound $\max_{1\le i\le n}\sum_{j=1}^\infty j^2|\boldsymbol{e}_i'C_j|_1 < \infty$. Specifically, applying Hölder's inequality, $c_h \le \max_j |||C_j|||_{\max} \max_{1\le i\le n}\sum_{j=1}^\infty |\boldsymbol{e}_i'C_j|_1$,

$c_\infty \leq \max_j |||C_j|||_{\max} \max_{1\leq i\leq n} \sum_{j=1}^{\infty} j|\boldsymbol{e}_i' C_j|_1$, and $\tilde{c}_\infty \leq \max_j |||C_j|||_{\max} \cdot \max_{1\leq i\leq n} \sum_{j=1}^{\infty} j^2|\boldsymbol{e}_i' C_j|_1$.

If $\{\mathbf{X}_t\}$ is centered, uncorrelated in time, and L_2-mixingale, one can show that for $k = 1, 2, ...,$ $\|\mathbb{E}[X_{it}X_{j,t-k}|\mathcal{F}_{t-m}]\|_1 \leq \tilde{c}_{k,t}e^{-m^\gamma/2c_\rho}$ for $\tilde{c}_{k,t} > (c_{it}+c_{it}c_{j,t-k}\rho_m^{3/2}+\|X_{i,t}X_{j,t-k}\|_2\|X_{j,t-k}\|_2)$. Hence, $\{\eta_{it}(k), \mathcal{F}_t\}$ is L_1-mixingale with the same "sub-Weibull" rate as $\{X_{it}, \mathcal{F}_t\}$. For $k = 0$ we directly assume that $\{X_{it}X_{jt}, \mathcal{F}_t\}$ is L_1 mixingale. Masini et al. (2022) discusses this condition, providing illustrative examples.

Remark 5 (Martingale difference)**.** If the process $\mathbf{X}_t, \mathcal{F}_{t-1}$ is a martingale difference process, $\mathbb{E}[X_{it}X_{j,t-k}|\mathcal{F}_{t-m}] = 0$ for $k = \pm 1, \pm 2, ...,$ which is equivalent to taking $\gamma \to \infty$. However, the process $\{\widetilde{\boldsymbol{\eta}}_t(0), \mathcal{F}_{t-1}\}$ is not a martingale difference and we still need the L_1-mixingale condition in the squared process. After small changes, we obtain a slightly tighter version of Theorem 3, but the leading terms remain unchanged.

Remark 6 (Accounting for centering)**.** Unless $\mathbf{Y}_1, .., \mathbf{Y}_T$ are centred, we have to account for centering. If $\mathbb{E}[\mathbf{Y}_t] = \mu_t$, $t = 1, .., T$, are known, we only have to demean $\mathbf{Y}_t$ and nothing changes as we may assume $\mu_t = 0$ without loss of generality in all derivations. However, μ_ts are typically unknown and must be estimated. Consider the simple case where $\{\mathbf{X}_t\}$ is weakly stationary, so that $\mu_t = \mu$, and define the estimators for the mean

$\bar{\mathbf{Y}}^{(1)}_{T-h} = \frac{1}{T-h}\sum_{t=1}^{T-h}\mathbf{Y}_t$ and $\bar{\mathbf{Y}}^{(h)}_{T-h} = \frac{1}{T-h}\sum_{t=h+1}^{T}\mathbf{Y}_t$. The lag-$h$ autocovariance estimator $\widehat{\Gamma}_T(h)$ is

$$\widehat{\Gamma}^*_T(h) := \frac{1}{T-h}\sum_{t=h+1}^{T}(\mathbf{Y}_t - \bar{\mathbf{Y}}^{(h)}_{T-h})(\mathbf{Y}_{t-h} - \bar{\mathbf{Y}}^{(1)}_{T-h})'. \tag{4.14}$$

It follows directly that

$$\widehat{\Gamma}^*_T(h) = \frac{1}{T-h}\sum_{t=h+1}^{T}(\mathbf{Y}_t - \mu)(\mathbf{Y}_{t-h} - \mu)' - (\bar{\mathbf{Y}}^{(h)}_{T-h} - \mu)(\bar{\mathbf{Y}}^{(1)}_{T-h} - \mu)',$$

meaning that after accounting for appropriate scaling and centering,

$$\left|\left|\left|\widehat{\Gamma}^*_T(h) - \mathbb{E}\widehat{\Gamma}^*_T(h)\right|\right|\right|_{\max} \leq \frac{T}{T-h}\Delta_T(h) + \max_{i=1,h}|\bar{\mathbf{Y}}^{(i)}_{T-h} - \mu|^2_\infty. \tag{4.15}$$

The first term on the right-hand side accounts for the covariances, and the second one for the mean.

For fixed h, the first term on the right-hand side of (4.15) is bounded as in Theorem 3, whereas the second term is bounded using Equation (3.9):

$$\Pr\left(\max_{i=1,h}|\bar{\mathbf{Y}}^{(i)}_{T-h} - \mu|^2_\infty > x\right) \leq O\left(\log(n)^{1/\gamma} e^{-\frac{(xT)^{\alpha/2}}{c_3 \log(n)^{1+\frac{\alpha}{2}+\frac{\alpha}{\gamma}}}}\right).$$

As the second term vanishes faster than the first, the convergence rate does not change.

Remark 7 (Accounting for h)**.** The concentration bound holds for each h, T and n that satisfies the conditions in Theorem 3. We have connected n and T to establish the dimension of $\mathbf{Y}_t$. Now, set $0 < l_T = (T-h)/T < 1$

and suppose that $T \cdot l_T > 4\log(n)$ and $3T \cdot l_T < n^{\beta-1}$, for some $\beta > 1$. We obtain the following bound in (4.13):

$$\begin{aligned} \Pr(\Delta_T(h) > x) \leq c_1 \frac{\log(n)^{1/\gamma}}{n^2} + \frac{c_2 \bar{c}^*_{T-h} l_T}{n^2 x} \\ + c_3 \log(n)^{1/\gamma} e^{-\frac{(x\sqrt{T/l_T})^{\alpha/2}}{c_4 \log(n)^{1+\frac{\alpha}{4}+\frac{\alpha}{2\gamma}}}} + 4e^{-\frac{(xT)^{\alpha/2}}{c_5 \log(n)}}, \end{aligned} \tag{4.16}$$

where $\bar{c}^*_{T-h} = \max_{1\leq i \leq n^2}(T-h)^{-1}\sum_{t=h+1}^{T} c_{it}$. If we take $h = (1-\nu)T$, for $0 < \nu < 1$, then $l_T = \nu$ and the bound remains unchanged in terms of rate. In other words, we may take $h \propto T$ in Theorem 3.

5. Applications

In this section, we describe two applications of concentration inequalities developed in the paper. The first application is to develop a nonasymptotic oracle bound for the regularized l_1 system estimation of the VAR(p) representations of time series. The second application is a concentration inequality for the maximum entry-wise error of the estimation of the long-run covariance of a linear process. We consider the following restrictions on the data-generating process.

Assumption (DGP)**.** The stochastic process $\{\mathbf{Y}_t\}$ taking values on $\mathbb{R}^n$ admits the linear representation:

$$\mathbf{Y}_t = \sum_{j=1}^{\infty} C_j \mathbf{u}_{t-j} = C(L)\mathbf{u}_t,$$

where (i) the innovation process $\{\mathbf{u}_t = (u_{1t}, \dots, u_{nt})'\}$ is causal, centred, uncorrelated, and sub-Weibull(α), with $\max_{i,t} \|u_{it}\|_{\psi_\alpha} \leq c_{\psi_\alpha}$; (ii) for each $i = 1, ..., n$, $\{u_{it}\}$ is L_2-mixingale, $\lim_{T\to\infty} \bar{c}_T < \infty$ and there exist positive constants c_ρ and γ so that $\rho_m \leq \exp(-m^\gamma/(2c_\rho))$; (ii) for each $i, j = 1, ..., n$, $\{u_{it}u_{jt}\}$ is L_1-mixingale, with $\lim_{T\to\infty} \bar{c}_T < \infty$ and there exist positive constants c_ρ and γ so that $\rho_m \leq \exp(-m^\gamma/(2c_\rho))$; (iv) the sequence of matrices $\{C_j\}$ satisfy for all T, n, and some $r \geq 1$, (a) $\sum_{j=1}^{\infty} j |||C_j|||^2 < \infty$; and (b) $\max_{1\leq i\leq n} \sum_{j=1}^{\infty} j^{r+1} |\boldsymbol{e}_i' C_j|_1 < \infty$.

5.1 LASSO estimation of VAR(p)

Regularised estimation of high-dimensional time series models have recently been the focus of much research, and two excellent reviews on this topic are Basu and Matteson (2021) and Masini et al. (2023). Closer to our interest, Basu and Michailidis (2015) and Kock and Callot (2015) consider LASSO estimation of Gaussian VAR(p) models, obtaining estimation and prediction bounds. Wong et al. (2020) extends this setting to VAR models on sub-Weibull strictly stationary stochastic processes with β-mixing dependence. Masini et al. (2022) derive nonasymptotic, oracle estimation bounds for linear processes with martingale difference innovations and sub-Weibull tails.

We consider processes that admit a linear representation with sub-Weibull, mixingale innovations. All previous examples are particular instances of our setup. Note that $\{\mathbf{Y}_t\}$ is not necessarily stationary, but we require the process to be centred at zero. Integrated process do not satisfy our conditions, but processes that are conditionally heteroscedastic are covered by our setup. In effect, the VAR(p) model under consideration approximates the process. We represent $\mathbf{Y}_t$ using a $VAR(p)$ model

$$\mathbf{Y}_t = A_1^* \mathbf{Y}_{t-1} + \cdots + A_p^* \mathbf{Y}_{t-p} + \mathbf{W}_t,$$

where $A_1^*, \ldots, A_p^*$ solve the quadratic program

$$(A_1^*, \ldots, A_p^*) := \underset{(A_1,\ldots,A_p)\in\mathbb{R}^{n\times pn}}{\text{arg min}} \sum_{t=p+1}^{T} \mathbb{E}\left(\left| \mathbf{Y}_t - \sum_{i=1}^{p} A_i \mathbf{Y}_{t-i} \right|_2^2 \right).$$

By construction, the error vector $\mathbf{W}_t = \mathbf{Y}_t - \sum_{i=1}^{p} A_i^* \mathbf{Y}_{t-i}$ $(t = p+1, .., T)$ is not correlated with $\mathbf{Y}_{t-1}$, ..., $\mathbf{Y}_{t-p}$.

In high-dimensional vector time series modelling, the number of series n is large compared to the number of observations T. It is assumed that the parameters A_i^* are weakly sparse, that is, most entries are (close to) zero; thus, we only estimate a small number of them.

Assumption (Identification)**.** The following identification conditions are met. (a) the covariance matrix Σ_T, with blocks $\Sigma_{T;r,s} = \Gamma_T(r-s)$, has eigenvalues bounded between $0 < \sigma_\Sigma^2 < 1$ and $1/\sigma_{\Sigma^2}$ uniformly on T; (b) the population parameters satisfy $\sum_{k=1}^p |\text{vec}\,(A_i^*)|_q^q \le R_q$, for some $0 \le q < 1$.

The first requirement ensures the least squares program has a unique solution whereas the second that this solution is weakly sparse.

One of the most popular ways for estimating sparse vector autoregressive models is the LASSO (Least Angle Selection and Shrinkage Operator), or the l_1 regularised estimator:

$$(\widehat{A}_1, \ldots, \widehat{A}_p) = \underset{(A_1,\ldots,A_p)\in\mathbb{R}^{n\times pn}}{\arg\min} \sum_{t=p+1}^{T} \left| \mathbf{Y}_t - \sum_{i=1}^{p} A_i \mathbf{Y}_{t-i} \right|_2^2 + \lambda \sum_{i=1}^{p} |\text{vec}\,(A_i)|_1.$$

The oracle estimation bound below follows as in Masini et al. (2022) Theorem 1, and its adaptation to our case is found in the online supplement. Let $a_\lambda = \min\left(\frac{\lambda}{2(1+\|\boldsymbol{A}\|_\infty)}, \frac{\sigma_\Sigma^{2(1-q)}\lambda^q}{64R_q}\right)$ and set

$$\pi(a_\lambda) = c_1 \frac{\log(n)^{1/\gamma}}{n^2} + \frac{c_2 \bar{c}_T}{n^2 a_\lambda} + c_3 \log(n)^{1/\gamma} e^{-\frac{(a_\lambda\sqrt{T})^{\alpha/2}}{c_4 \log(n)^{1+\frac{\alpha}{4}+\frac{\alpha}{2\gamma}}}} + 4e^{-\frac{(a_\lambda T)^{\alpha/2}}{c_5 \log(n)}}.$$

Then, with probability $1 - 2p\pi(a_\lambda)$ and $T > 4\log(n)$,

$$\sum_{i=1}^{p} \left\|\left| \widehat{A}_i - A_i^* \right\|\right|_F^2 \le (44 + 2\lambda) R_q \left(\frac{\lambda}{\sigma_\Sigma^2}\right)^{2-q}. \tag{5.17}$$

Suppose Assumptions DGP and Identification hold with σ_Σ^2 and R_q uniformly bounded for all n. Eventually, for λ sufficiently small, $a_\lambda = \lambda/2(1 + \|\boldsymbol{A}^*\|_\infty)$. Set

$$\lambda \ge c\,(\log\log(n) + \epsilon)^{2/\alpha} \log(n)^{2/\alpha + 1/\gamma} \sqrt{\frac{\log(n)}{T}},$$

for any $\epsilon > 0$ and some constant $c > 0$ sufficiently large. Then, the oracle bound will hold with probability at least

$$1-2p\pi(a_\lambda) = 1-4pe^{-\epsilon}-\frac{2pc_1 \log(n)^{1/\gamma}}{n^2}-\frac{(c_2/c)\bar{c}_T\sqrt{T}}{n^2 \log(n)^{1/2+2/\alpha+1/\gamma}(\log\log(n)+\epsilon)^{2/\alpha}}.$$

In practice, σ_Σ^2 and R_q can grow as a function of n, in which case the rate of decrease on λ would have to accommodate these quantities. We have for $a_\lambda = \sigma_\Sigma^{2(1-q)}\lambda^q/64R_q$

$$\lambda^q \geq c\,\frac{\sigma_\Sigma^{2(1-q)}}{R_q}\,(\log\log(n)+\epsilon)^{2/\alpha}\log(n)^{2/\alpha+1/\gamma}\sqrt{\frac{\log(n)}{T}},$$

some constant c sufficiently large. However, it is not necessarily a constraint in the rate of λ, provided that $\lambda R_q^{1/(1-q)}/\sigma_\Sigma^2 = O(1)$.

Consider the simplified setup in which R_q and σ_Σ^2 are bounded for all n and T. Basu and Michailidis (2015) obtain the estimation bound in (5.17) of $O(\log(n)/T)$ with probability $1-O(n^{-c})$. Wong et al. (2020) obtain an oracle bound $O(\log(n)/T)$ with probability $1-O(n^{-c_1}\wedge Te^{-T^{c_2}})$. Finally, in system estimation, Masini et al. (2022) obtain a bound $O(\log(n)^{(4-2q)/\alpha}(\log(n)/T)^{1-q/2})$ with probability $1-O(n^{c_1}T^{c_2}e^{-T^{c_3}} \wedge n^{-c_4})$. In our case, we obtain a bound $O(\log(n)^{(2-q)/\gamma}(\log(n)\log\log(n))^{(4-2q)/\alpha}(\log(n)/T)^{1-q/2})$ with probability $1-O(n^{-2}\wedge T^{1/2}n^{-2}\log(n)^{-c})$. Taking $q=0$ we recover the strong sparsity condition in Basu and Michailidis (2015) and Wong et al. (2020), in which case our upper bound is $O(\log(n)^{2/\gamma+4/\alpha}\log\log(n)^{4/\alpha}(\log(n)/T))$

with similar probability. In this case we pay a $O(\log(n)^c)$ price for the relaxed conditions.

5.2 Long-run covariance matrix

We present finite sample error bounds for a class of HAC estimators of the long-run covariance Ω_T, which is essential for precise inference in time series. In stationary time series, Ω_T is related to its spectral density, thus connecting the HAC estimators with the spectral density matrix estimators in statistics and econometrics (see Politis (2011) and Xiao and Wu (2012)).

Historically, estimators for (long-term) covariance matrices proposed by Newey and West, Andrews and Hansen did not take into account high dimensions. In recent years, researchers have focused on inference for high-dimensional time-series models (Zhang and Wu, 2017; Li and Liao, 2020; Babii et al., 2022; Fan et al., 2023). The most common approach is to assess the characteristics of Newey and West's HAC estimator in high dimensions and under certain dependence and tail conditions. In the case of small dimensions, Politis (2011) shows that the *flat-top* kernels can achieve faster convergence rates. We establish finite sample error bounds for the Politis's flat-top kernel, spectral density matrix estimator under maximum entrywise norm.

We follow Assumption DGP with the added condition that $\{\mathbf{u}_t\}$ is weakly stationary. Then, the autocovariances $\Gamma(j) = \mathbb{E}[\mathbf{Y}_t\mathbf{Y}_{t-j}] = \Gamma(-j)'$, for $j \geq 0$ and each t. The spectral density matrix evaluated at w is

$$\boldsymbol{F}(w) := \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\Gamma(k)e^{-ikw},$$

where $i = \sqrt{-1}$. For $\pi \leq w \leq \pi$, $\boldsymbol{F}(w)$ is positive definite and Hermitian, and $\lim_{T\to\infty}\Omega_T = 2\pi\boldsymbol{F}(0)$. For each $h = 0, \pm 1, \ldots$, the sample covariance is $\widehat{\Gamma}_T(h) = \widehat{\Gamma}_T(-h)'$ for $|h| < T$ and $\widehat{\Gamma}_T(h) = 0$ otherwise. The empirical spectral matrix estimator is

$$\widehat{\boldsymbol{F}}(w) := \frac{1}{2\pi}\sum_{h=-T+1}^{T-1}\kappa_{g,\epsilon}(h/M_T)\widehat{\Gamma}_T(h),$$

where $\kappa_{g,\epsilon}(\cdot)$ is a flat-top kernel. The typical flat-top kernel is given by $\kappa_{g,\epsilon}(u) = 1$ if $|u| \leq \epsilon$ and $\kappa_{g,\epsilon}(u) = g(u) - 1$ if $|u| > \epsilon$, where $\epsilon > 0$ is a parameter and $g : \mathbb{R} \mapsto [-1, 1]$ is a symmetric function, continuous at all but a finite number of points, satisfying $g(\epsilon) = 1$ and $\int_{\mathbb{R}} g^2(u)du < \infty$.

It follows after some algebra (see online supplement) that in a set with large probability and for any $w \in [0, 2\pi]$,

$$\left|\!\left|\!\left|\widehat{\boldsymbol{F}}(w) - \boldsymbol{F}(w)\right|\!\right|\!\right|_{\max} \lesssim \frac{2c_r}{\pi\epsilon^r M_T^r} + \frac{c_0}{\pi T} + \frac{H_T}{\sqrt{T}}\log(n)^{1+\frac{2}{\alpha}+\frac{1}{\gamma}}(\log\log(n)+\log(T)+\tau)^{2/\alpha}.$$

Naturally, we require the right-hand side to converge to zero. Therefore, H_T and M_T cannot grow too fast. Such restrictions are expected and are also observed in Li and Liao (2020) and Babii et al. (2022).

Remark 8. Let $\mathbf{Y}_t = f(\mathbf{Z}_t; \theta_0)$ where the population parameter θ_0 is estimated by $\widehat{\theta}_T$. In this case, we have access to $\widehat{\mathbf{Y}}_t = f(\mathbf{Z}_t; \widehat{\theta}_T)$ and calculate

$$\widetilde{\Gamma}_T(h) = \frac{1}{T}\sum_{t=h+1}^{T} \widehat{\mathbf{Y}}_t \widehat{\mathbf{Y}}'_{t-h} \quad \text{and} \quad \widetilde{\Gamma}_T(-h) = \widetilde{\Gamma}_T(h), \quad h = 0, 1 \ldots, T-1.$$

The spectral density estimator is

$$\widetilde{\boldsymbol{F}}(w) = \sum_{h=-T+1}^{T-1} \kappa_{g,\epsilon}(h/M_T)\widetilde{\Gamma}_T(h).$$

If the estimation error $\max_{i\le n}\sum_{t=1}^{T}\left(Y_{i,t} - \widehat{Y}_{i,t}\right)^2 < T\delta_{n,t}^2$ with high probability, we obtain

$$\left|\left|\left|\widetilde{\boldsymbol{F}}(w) - \boldsymbol{F}(w)\right|\right|\right|_{\max} \le \delta_{n,T} + \frac{2c_r}{\pi\epsilon^r M_T^r} + \frac{c_0}{\pi T} + \frac{H_T}{\sqrt{T}}\log(n)^{1+\frac{2}{\alpha}+\frac{1}{\gamma}}(\log\log(n) + \log(T) + \tau)^{2/\alpha},$$

with large probability as well.

6. Discussion

In this paper, we explore the concentration bounds for the supremum norm of averages of vector-valued linear processes. We demonstrate that, when the weights are summable, the rates obtained for sums of mixingales can be extended to sums of linear processes. This result is noteworthy for two reasons: it does not require stationarity, and the constants depending on the weights or the mixingale term can increase with the dimension n of the process or the sample size T. Additionally, the mixingale condition is a well-known condition used in the derivation of asymptotic properties of time series estimators, as well as in the analysis of properties of nonlinear models. Masini et al. (2022, Section 3) provides some examples of processes that meet the mixingale and tail conditions.

This article generalises the concentration results of Wong et al. (2020), which requires $\mathbf{Y}_t$ to be β mixed, Masini et al. (2022), which requires Y_t to be an approximately VAR(p) process, and Bours and Steland (2021), which assumes i.i.d. innovations. Masini et al. (2022) was extended to encompass misspecified linear models, where the mean is not an approximately sparse VAR model. Bours and Steland (2021) showed that factor models and generalised factor models can be represented by a linear process. This is also applicable to processes with stochastic variance, as in Masini et al. (2022, Section 3).

Kuchibhotla and Chakrabortty (2022) demonstrate that the concentration inequalities developed in this paper can be used in a variety of statistical contexts, including the derivation of estimation bounds for the maximum k-sub-matrix operator norm and the restricted isometry property. Additionally, the restricted eigenvalue condition and restricted eigenvalue condition for linear time series models were shown to hold.

We use our concentration inequalities to obtain estimation bounds for weakly sparse $VAR(p)$ autoregressions and error bounds for high-dimensional HAC estimators with flat-top kernels. We extend the work of Basu and Michailidis (2015); Kock and Callot (2015); Wong et al. (2020) and Masini et al. (2022) by considering a more general data-generating process. Additionally, we provide estimation error bounds for the HAC, flat-top kernel estimator from Politis (2011) in a high-dimensional setting. This concentration bound has been used in high-dimensional time series inference, as demonstrated in Zhang and Wu (2017); Li and Liao (2020); Babii et al. (2022) and Fan et al. (2023).

Supplementary Materials

This supplement provides a more comprehensive analysis of the innovation process and its characteristics and also includes proofs of all the results in the paper. We derive concentration bounds that are used in the proof of the triplex inequality in Section 2. Then, we present the results of Sections 3, 4 and 5 of the main paper, along with their respective proofs.

Acknowledgements

We thank professors Marcelo Fernandes and Luiz Max Carvalho, and two anonymous referees for their insightful comments. The first author acknowledge financial support from the São Paulo Research Foundation (FAPESP), grant 2023/01728-0

References

Adamczak, R. (2015). A note on the Hanson-Wright inequality for random vectors with dependencies. *Electronic Communications in Probability 20*(none), 1 – 13.

Adamek, R., S. Smeekes, and I. Wilms (2023a). Lasso inference for high-dimensional time series. *Journal of Econometrics 235*(2), 1114–1143.

Adamek, R., S. Smeekes, and I. Wilms (2023b). Sparse high-dimensional vector autoregressive bootstrap.

Alquier, P. and P. Doukhan (2011). Sparsity considerations for dependent variables. *Electronic Journal of Statistics 5*, 750–774.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society 59*(3), 817–858.

Babii, A., E. Ghysels, and J. Striaukas (2022). High-dimensional grander causality tests with an application to vix and news. *Journal of Financial Econometrics*.

Basu, S. and D. S. Matteson (2021). A survey of estimation methods for sparse high-dimensional time series models.

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

Beveridge, S. and C. R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary economics 7*(2), 151–174.

Bong, H. and A. K. Kuchibhotla (2023). Tight concentration inequality for sub-weibull random variables with generalized bernstien orlicz norm.

Bours, M. and A. Steland (2021). Large-sample approximations and change testing for high-dimensional covariance matrices of multivariate linear time series and factor models. *Scandinavian Journal of Statistics 48*(2), 610–654.

Brockwell, P. J. and R. A. Davis (2009). *Time series: theory and methods*. New York: Springer science & business media.

Chen, L. and W. B. Wu (2018). Concentration inequalities for empirical processes of linear time series. *J. Mach. Learn. Res. 18*, 231–1.

Dedecker, J., P. Doukhan, G. Lang, J. León, S. Louhichi, and C. Prieur (2007). *Weak dependence with examples and applications*. Springer.

Dedecker, J., F. Merlevède, and M. Peligrad (2011). Invariance principles for linear processes with application to isotonic regression. *Bernoulli 17*(1), 88–113.

Doukhan, P. and M. H. Neumann (2007). Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Processes and their Applications 117*(7), 878–903.

Fan, J., R. Masini, and M. C. Medeiros (2023). Bridging factor and sparse models. *Annals of Statistics 51*(4), 1692–1717.

Fan, X., I. Grama, and Q. Liu (2012). Hoeffding's inequality for supermartingales. *Stochastic Processes and their Applications 122*(10), 3545–3559.

Fan, X., I. Grama, Q. Liu, et al. (2012). Large deviation exponential inequalities for supermartingales. *Electronic Communications in Probability 17*.

Götze, F., H. Sambale, and A. Sinulis (2021). Concentration inequalities for polynomials in α-sub-exponential random variables. *Electronic Journal of Probability 26*, 1–22.

Hall, P. and C. C. Heyde (1980). *Martingale limit theory and its application.* New York: Academic press.

Han, F., H. Lu, and H. Liu (2015). A direct estimation of high dimensional stationary vector autoregressions. *The Journal of Machine Learning Research 16*(1), 3115–3150.

Hang, H. and I. Steinwart (2017). A bernstein-type inequality for some mixing processes and dynamical systems with an application to learning. *Annals of Statistics 45*(2), 708–743.

Hansen, B. E. (1992). Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica: Journal of the Econometric Society 60*(4), 967–972.

Jiang, W. (2009). On uniform deviations of general empirical risks with unboundedness, dependence and high dimensionality. *Journal of Machine Learning Research 10*, 977–996.

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

Krampe, J., J.-P. Kreiss, and E. Paparoditis (2021). Bootstrap based inference for sparse high-dimensional time series models. *Bernoulli 27*(3), 1441–1466.

Kuchibhotla, A. K. and A. Chakrabortty (2022). Moving beyond sub-Gaussianity in high-dimensional statistics: applications in covariance estimation and linear regression. *Information and Inference: A Journal of the IMA 11*(4), 1398–1456.

Lesigne, E. and D. Volný (2001). Large deviations for martingales. *Stochastic Process and their Applications 96*, 143 – 159.

Li, J. and Z. Liao (2020). Uniform nonparametric inference for time series. *Journal of Econometrics 219*(1), 38–51.

Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.

Marton, K. (1998). Measure concentration for a class of random processes. *Probability Theory and Related Fields 110*, 427–439.

Masini, R. P., M. C. Medeiros, and E. F. Mendes (2022). Regularized estimation of high-dimensional vector autoregressions with weakly dependent innovations. *Journal of Time Series Analysis 43*, 532–557.

Masini, R. P., M. C. Medeiros, and E. F. Mendes (2023). Machine learning advances for time series forecasting. *Journal of Economic Surveys 37*(1), 76–111.

Merlevède, F., M. Peligrad, and E. Rio (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In C. Houdré, V. Koltchinskii, D. Mason, and M. Peligrad (Eds.), *High Dimensional Probability V: The Luminy Volume*, pp. 273–292. Institute of Mathematical Statistics.

Mohri, M. and A. Rostamizadeh (2010). Stability bounds for stationary ϕ-mixing and β-mixing processes. *Journal of Machine Learning Research 11*(2).

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Phillips, P. C. and V. Solo (1992). Asymptotics for linear processes. *The Annals of Statistics 20*(2), 971–1001.

Politis, D. N. (2011). Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econometric Theory 27*(4), 703–744.

Reuvers, H. and E. Wijler (2024). Sparse generalized yule–walker estimation for large spatio-temporal autoregressions with an application to no2 satellite data. *Journal of Econometrics 239*(1).

Tao, T. and V. Vu (2013). Random matrices: Sharp concentration of eigenvalues. *Random Matrices: Theory and Applications 2*(03), 1350007.

Tsay, R. S. (2013). *Multivariate time series analysis: with R and financial applications*. Hoboken, New Jersey: John Wiley & Sons.

van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, Volume 47. Cambridge University Press.

Vladimirova, M., S. Girard, H. Nguyen, and J. Arbel (2020). Sub-weibull distributions: Generalizing sub-gaussian and sub-exponential properties to heavier tailed distributions. *Stat 9*(1), e318.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, Volume 48. Cambridge university press.

Wang, D. and R. S. Tsay (2023). Rate-optimal robust estimation of high-dimensional vector autoregressive models. *Annals of Statistics 51*(2), 846–877.

Wang, Q., Y.-X. Lin, and C. M. Gulati (2001). Asymptotics for moving average processes with dependent innovations. *Statistics & probability letters 54*(4), 347–356.

Wilms, I., S. Basu, J. Bien, and D. S. Matteson (2021). Sparse identification and estimation of large-scale vector autoregressive moving averages. *Journal of the American Statistical*

Association 118(541), 571–582.

Wong, K. C., Z. Li, and A. Tewari (2020). Lasso guarantees for β-mixing heavy-tailed time series. *The Annals of Statistics 48*(2), 1124–1142.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences 102*(40), 14150–14154.

Wu, W. B. and W. Min (2005). On linear processes with dependent innovations. *Stochastic Processes and Their Applications 115*(6), 939–958.

Xiao, H. and W. B. Wu (2012). Covariance matrix estimation for stationary time series. *The Annals of Statistics 40*(1), 466–493.

Yu, B. (1994, January). Rates of convergence for empirical processes of stationary mixing sequences. *The Annals of Probability 22*(1), 94–116.

Zhang, D. and W. B. Wu (2017). Gaussian approximation for high-dimensional time series. *The Annals of Statistics 45*(5), 1895–1919.

Zhang, H. and S. Chen (2021). Concentration inequalities for statistical inference. *Communications in Mathematical Research 37*(1), 1–85.

Zhang, H. and H. Wei (2022). Sharper sub-weibull concentrations. *Mathematics 10*(13), 2252.

Getulio Vargas Foundation, São Paulo School of Economics

E-mail: eduardo.mendes@fgv.br

Getulio Vargas Foundation, School of Applied Mathematics

E-mail: fellipe_293@hotmail.com