TEXTO PARA DISCUSSÃO

No. 680

Regularized estimation of high-dimensional vector autoregressions with weakly dependent innovations

Ricardo P. Masini
Marcelo C. Medeiros
Eduardo F. Mendes

DEPARTAMENTO DE ECONOMIA
www.econ.puc-rio.br

REGULARIZED ESTIMATION OF HIGH-DIMENSIONAL VECTOR AUTOREGRESSIONS WITH WEAKLY DEPENDENT INNOVATIONS

RICARDO P. MASINI, MARCELO C. MEDEIROS, AND EDUARDO F. MENDES

ABSTRACT. There has been considerable advance in understanding the properties of sparse regularization procedures in high-dimensional models. In time series context, it is mostly restricted to Gaussian autoregressions or mixing sequences. We study oracle properties of LASSO estimation of weakly sparse vector-autoregressive models with heavy tailed, weakly dependent innovations with virtually no assumption on the conditional heteroskedasticity. In contrast to current literature, our innovation process satisfy an L^1 mixingale type condition on the centered conditional covariance matrices. This condition covers L^1-NED sequences and strong (α-) mixing sequences as particular examples. From a modeling perspective, it covers several multivariate-GARCH specifications, such as the BEKK model, and other factor stochastic volatility specifications that were ruled out by assumption in previous studies.
JEL: C32, C55, C58.
Keywords: high-dimensional time series, LASSO, VAR, mixing.

1. INTRODUCTION

Modeling multivariate time series data is an important and vibrant area of research. Applications range from economics and finance, as in Sims (1980), Bauer and Vornik (2011), Chiriac and Voev (2011), or Ramey (2016), to air pollution and ecological studies (Hoek et al., 2013; Ensor et al., 2013; Schweinberger et al., 2017). Among alternatives, the Vector Autoregressive (VAR) model is certainly one of the most successful in modeling temporal evolution of vectors, networks, and matrices. See Lütkepohl (1991) or Wilson et al. (2015) for comprehensive textbook introductions.

The advances in data collection and storage have created data sets with large numbers of time series (*Big Data*), where the number of model parameters to be estimated may exceed the number of available data observations. A common approach to dealing with high-dimensional data is to impose additional structure in the form of (approximate) sparsity and estimate the parameters by some shrinkage method. Examples of estimation techniques range from Bayesian estimation with “spike-and-slab” priors to sparsity-inducing shrinkage, such as the least absolute and shrinkage estimator (LASSO) and its many extensions. See Miranda-Agrippino and Ricco (2019) for a nice survey on Bayesian

The authors are grateful to Anders B. Kock and Giuseppe Cavaliere for helpful comments. M. C. Medeiros acknowledges partial support from CNPq/Brazil and CAPES..

VARs or Kock et al. (2020) for a review on penalized regressions applied to time-series models.

1.1. **Our Contributions.** In this paper we study non-asymptotic properties of high-dimensional VAR models and their parameter estimates using equation-wise (row-wise node-wise) LASSO. We show that, with high probability, estimated and population parameter vectors are close to each other in the Euclidean norm and discuss restrictions on the rate which the number of parameters can increase as the sample size diverges.

The importance of our results relies on the fact that our non-asymptotic guarantees serve as a fundamental ingredient for the derivation of asymptotic properties of penalized estimators in high-dimensional VAR models. In particular, our results apply with minimal restrictions on the conditional variance model, allowing, for instance, large-dimensional multivariate GARCH models. Moreover, auxiliary results proved in this paper are of independent interested and can, for instance, be used to derive finite bounds for other type of penalization such as group/structured lasso, elastic-net, SCAD or non-convex penalties.

The data are assumed to be generated from a covariance-stationary and weakly sparse VAR model, where the innovations are martingale difference with sub-Weibull tails and conditional covariance matrix satisfying a L^1 mixingale assumption. An important feature is that the resulting process $\{\mathrm{y}_t\}$ is not necessarily mixing and we avoid mixing assumptions at all in this paper, which can be notoriously difficult to show. Nevertheless, it follows that our model contemplate strong mixing innovations as a particular case.

These conditions contemplate VAR models with conditional heteroskedasticity as in Bauwens et al. (2006); Boussama et al. (2011) or stochastic volatility as in Chib et al. (2009).

1.2. **Literature review.** Some consistency results on model estimation and selection of high-dimensional VAR processes were obtained by Song and Bickel (2011), though under much stronger assumptions, such as Gaussianity. Loh and Wainwright (2012) and Basu and Michailidis (2015) developed powerful concentration inequalities that enabled them to establish consistency under weaker conditions and prove that these conditions hold with high probability. In particular, Basu and Michailidis (2015) established consistency of l_1-penalized least squares and maximum likelihood estimators of the coefficients of high-dimensional Gaussian VAR processes and related the estimation and prediction error to the complex dependence structure of VAR processes.

Other estimation approaches, including Bayesian approaches, are discussed by Davis et al. (2016). Miao et al. (2019) proposed a factor-augmented large dimensional VAR and studied finite sample properties and provide estimation results. However, they assume independent and identically distributed errors. More recently, Wong et al. (2020) derived finite-sample guarantees for the LASSO in a misspecified VAR model. Authors assume the series is either β-mixing process with sub-Weibull marginal distributions or α-mixing Gaussian processes.

1.3. **Organization of the Paper.** The paper is organized as follows. In Section 2 we define the model and the main assumptions in the paper. In Section 3 we discuss examples of applications of our results. The theoretical results are presented in Section 4, while in Section 5 we provide a discussion of our findings and conclude the paper. All technical proofs are relegated to the Appendix.

1.4. **Notation.** Throughout the paper we use the following notation. For a vector $\mathbf{b} = (b_1, ..., b_k)' \in \mathbb{R}^k$ and $p \in [1, \infty]$, $|\mathbf{b}|_p$ denotes its l_p norm, i.e. $|\mathbf{b}|_p = (\sum_{i=1}^{k} |b_i|^p)^{1/p}$ for $p \in [1, \infty)$ and $|\mathbf{b}|_\infty = \max_{1\leq i\leq k} |b_i|$. We also define $|\mathbf{b}|_0 = \sum_{i=1}^{k} I(b_i \neq 0)$. For a random variable X, $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$ for $p \in [1, \infty)$ and $\|X\|_\infty = \sup\{a : \Pr(|X| \geq a) = 0\}$. For a $m \times n$ matrix $\boldsymbol{A}$ with elements a_{ij}, we denote $|||\boldsymbol{A}|||_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}|$, $|||\boldsymbol{A}|||_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}|$, the induced l_∞ and l_1 norms respectively, and the maximum elementwise norm $|||\boldsymbol{A}|||_{\max} = \max_{i,j} |a_{ij}|$. Also $\Lambda_{\min}(\boldsymbol{A})$ and $\Lambda_{\max}(\boldsymbol{A})$ denotes the minimum and maximum eigenvalues of the matrix $\boldsymbol{A}$, respectively.

2. Model setup and Assumptions

Let $\{\mathbf{y}_t = (y_{t,1}, ..., y_{t,n})'\}$ be a vector stochastic process defined in some fixed probability space taking values on $\mathbb{R}^n$ given by

$$\mathbf{y}_t = \boldsymbol{A}_1\mathbf{y}_{t-1} + \cdots + \boldsymbol{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t, \tag{1}$$

where $\mathbf{u}_t = (u_{t,1}, ..., u_{t,n})'$ is a zero-mean vector of innovations and $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p$, are $n \times n$ parameter matrices. The dimension $n \equiv n_T$ and order $p \equiv p_T$ of the process are allowed to increase with the number of observations T. Write the vector-autoregressive (VAR) process (1) using its first-order representation:

$$\tilde{\mathbf{y}}_t = \boldsymbol{F}_T\tilde{\mathbf{y}}_{t-1} + \tilde{\mathbf{u}}_t, \tag{2}$$

where $\tilde{\mathbf{y}}_t = (\mathbf{y}_1', ..., \mathbf{y}_{p-1}')'$, $\tilde{\mathbf{u}}_t = (\mathbf{u}_t', \mathbf{0}', ..., \mathbf{0}')'$, and

$$\boldsymbol{F}_T = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots & \boldsymbol{A}_{p-1} & \boldsymbol{A}_p \\ \boldsymbol{I}_n & \mathbf{0}_n & \cdots & \mathbf{0}_n & \mathbf{0}_n \\ \mathbf{0}_n & \boldsymbol{I}_n & & \mathbf{0}_n & \mathbf{0}_n \\ \vdots & & \ddots & \vdots & \vdots \\ \mathbf{0}_n & \mathbf{0}_n & & \boldsymbol{I}_n & \mathbf{0} \end{bmatrix}.$$

Consider now the following assumptions.

Assumption (A1). *All roots of the reverse characteristic polynomial $\boldsymbol{\mathcal{A}}(z) = \boldsymbol{I}_n - \sum_{i=1}^{p} \boldsymbol{A}_j z^j$ lie outside the unit disk and there exist $\bar{c}_\Phi > 0$, $c_\phi > 0$ and $0 < \gamma_1 \leq 1$ such that*

$$\max_{\delta=1,\dots,n} \sum_{k=m}^{\infty} |\phi_{k,\delta}|_1 \leq \bar{c}_\Phi e^{-c_\phi m^{\gamma_1}}, \tag{3}$$

where $\boldsymbol{\Phi}_k := \boldsymbol{J}'\boldsymbol{F}_T^k\boldsymbol{J} = (\phi_{k,1}, ..., \phi_{k,n})'$ for all n and p, $\boldsymbol{F}_T$ denote the companion matrix and $\boldsymbol{J} = (\boldsymbol{I}_n, \mathbf{0}_n, ..., \mathbf{0}_n)'$.

Assumption (A2). *The sequence $\{\mathbf{u}_t\}$ is zero-mean, covariance stationary, martingale difference process with respect to its natural filtration $\{\mathcal{F}_t\}$. The largest and smallest eigenvalues of $\boldsymbol{\Sigma} := \mathbb{E}(\boldsymbol{u}_1\boldsymbol{u}_1')$ are bounded away from 0 and ∞ respectively, uniformly in $T \in \mathbb{N}$. Furthermore, for all $\mathbf{b}_1, \mathbf{b}_2 \in \{v \in \mathbb{R}^n : |v|_1 \leq 1\}$,*

$$\max_{t\in\mathbb{N}} \mathbb{E}\left|\mathbb{E}[\mathbf{b}_1'(\mathbf{u}_t\mathbf{u}_t' - \boldsymbol{\Sigma})\mathbf{b}_2|\mathcal{F}_{t-m}]\right| \leq a_1 e^{-a_2 m^{\gamma_2}},$$

for some $a_1, a_2 > 0$ and $0 < \gamma_2 \leq 1$.

Assumption (A3). *For all $\mathbf{b} \in \{v \in \mathbb{R}^n : |v|_1 \leq 1\}$, $\max_{t\in\mathbb{N}} \Pr(|\mathbf{b}'\mathbf{u}_t| > x) \leq 2e^{-|x/c_\alpha|^\alpha}$ for some $\alpha > 0$,$0 < c_\alpha < \infty$, and all $0 < x < \infty$.*

Assumption (A1) requires that the VAR process is stable and admits an infinite-order vector moving average, VMA(∞), representation for all n and p as

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \boldsymbol{J}'\boldsymbol{F}_T^i\boldsymbol{J}\mathbf{u}_{t-i} = \sum_{i=0}^{\infty} \boldsymbol{\Phi}_i\mathbf{u}_{t-i}. \tag{4}$$

Furthermore, the coefficients of the MA(∞) representations of each $\{y_{i,t}\}$, $i = 1, .., n$, are absolutely summable with exponentially decaying rate. This condition is satisfied in standard VAR(p) models, where n and p are fixed. In models that n is large, Lemma 4 in Appendix B.1 shows that condition (3) is satisfied if $\sum_{k=1}^{p} |||\boldsymbol{A}_k|||_\infty < 1$ and further regularity conditions on the size of the coefficients. Finally, notice that under (A1) it is also true that $\max_{k,i} |\phi_{k,i}|_\infty \leq \bar{c}_\Phi$, which means that the coefficients $\{\Phi_k\}$ are uniformly upper bounded under the maximum entry-wise norm.

Assumption (A2) requires the error process to be a martingale difference process satisfying very weak dependence condition on its conditional variance. This condition is of the weak type, L^1 projective dependence measure in Dedecker et al. (2007, section 2.2.4). Note that (1) strong mixing (or α-mixing) sequences satisfy this condition (Davidson, 1994, Theorem 14.2); and (2) uniform mixing sequences (ϕ-mixing) and β-mixing sequences are also strong mixing, but the converse is not true (Bradley, 2005, Equations (1.11) - (1.18)). If we denote the centered outer product series $\boldsymbol{v}_t = \mathsf{vech}\,(\mathbf{u}_t\mathbf{u}_t' - \boldsymbol{\Sigma})$, this assumptions requires that $\{\boldsymbol{v}_t\}$ is L^1 mixingale. It means that stochastic process with L^r bounded, L^1 near-epoch dependent, centered outer product series $\boldsymbol{v}_t$ are also contemplated in this setting Andrews (1988).

Finally, Assumptions (A1) and (A2) combined ensure that $\{\mathbf{y}_t\}$ is second order stationary for each n and p (Lütkepohl, 2006, Ch. 2). However, the process $\{\mathbf{y}_t\}$ is not mixing nor (necessarily) near-epoch dependent.

Condition (A3) imposes restrictions on the tail behavior of the innovation process $\{\mathbf{u}_t\}$ that are shared by $\{\mathbf{y}_t\}$. More precisely, we impose moment conditions on all linear combinations $\mathbf{b}'\mathbf{u}_t$ and Lemma 3, in the appendix, shows that each $\{y_{i,t}\}$ $(i = 1, ..., n)$ also share the same tail properties. This condition is essential for defining the rate in which n and p increase with T. We focus on the case the tail decays at rate $O(e^{-cx^{\alpha}})$ for some $\alpha > 0$, that is, $\{\mathbf{b}'\mathbf{u}_t\}$ is sub-Weibull with parameter α studied in Wong et al. (2020, Section 4.1). Note that when $\alpha \geq 1$ and $\alpha \geq 2$ we have the sub-exponential and sub-Gaussian tails respectively. However, when $\alpha \in (0, 1)$ the moment generating function does not exist at any point and and these variables are usually called *heavy tailed*.

Assumptions (A2) and (A3) describe the innovation process which has been shown to be satisfied by a series of models. For instance, Proposition 3 in Carrasco and Chen (2002) shows that under a set of regularity conditions our assumptions (A2) and (A3) are satisfied by the polynomial random coefficient autoregressive model; Boussama et al. (2011) derive conditions for stationarity and geometric ergodicity and geometric strong mixing for the general multivariate GARCH(p,q) model under the BEKK parametrization; and Hafner and Preminger (2009a,b) provide conditions under which (A2)–(A3) is satisfied for a multivariate GARCH specification and factor-GARCH models.

It is convenient to write the model in stacked form. Let $\mathbf{x}_t = (\mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}')'$ be the $np \times 1$ vector of regressors and $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_T)'$ the $T \times np$ matrix of covariates. Let $\mathbf{Y}_i = (y_{i,1}, ..., y_{i,T})'$ be the $T \times 1$ vector of observations for the i^{th} element of $\mathbf{y}_t$, and $\mathbf{U}_i = (u_{i,1}, ..., u_{i,T})'$ the corresponding vector of innovations. Denote $\boldsymbol{\beta}_i$ the $np \times 1$ vector of coefficients corresponding to equation i. Then, model (1) is equivalent to

$$\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta}_i + \mathbf{U}_i, \quad i = 1, \ldots, n. \tag{5}$$

We now make additional assumptions concerning model (5).

Assumption (A4). *The true parameter vectors $\boldsymbol{\beta}_i$, $i = 1, \ldots, n$, satisfy $\sum_{j=1}^{np} |\beta_{i,j}|^q \leq R_q$ for some $0 \leq q < 1$ and $0 < R_q < \infty$.*

Assumption (A5). *The smallest eigenvalue of $\boldsymbol{\Gamma} := T^{-1}\mathbb{E}(\mathbf{X}'\mathbf{X})$ is greater than a positive universal constant σ_Γ^2, uniformly on T.*

Assumption (A4) imposes *weak sparsity* of the coefficients, in a sense that most of them are small. This condition is slightly stronger than we need in a sense that we may have distinct q_i and $R_{q,i}$ for each equation. In the case $q = 0$ we have sparsity in the standard sense, meaning that $R_0 = s$, the number of non-zero coefficients. In practice, we estimate a sparse model that truncates all coefficients close to zero. This assumption is standard for *weak sparsity*, see Negahban et al. (2012)[section 4.3] and Han and Tsay (2019)[Assumption 1] for an application in time series setting.

Assumption (A5) is often used in the sparse estimation literature (e.g. Kock and Callot, 2015; Medeiros and Mendes, 2016b; Han and Tsay, 2019). Basu and Michailidis (2015) (Proposition 2.3) derived bounds for $\Lambda_{\min}(\boldsymbol{\Gamma})$ and $\Lambda_{\max}(\boldsymbol{\Gamma})$ using properties of the block Toeplitz matrix $\boldsymbol{\Gamma}$ and its generating function, the cross-spectral density of the generating VAR(p) process:

$$\frac{\Lambda_{\min}(\boldsymbol{\Sigma})}{\max_{|z|=1} \Lambda_{\max}(\boldsymbol{\mathcal{A}}^*(z)\boldsymbol{\mathcal{A}}(z))} \leq \Lambda_{\min}(\boldsymbol{\Gamma}) \leq \Lambda_{\max}(\boldsymbol{\Gamma}) \leq \frac{\Lambda_{\max}(\boldsymbol{\Sigma})}{\max_{|z|=1} \Lambda_{\min}(\boldsymbol{\mathcal{A}}^*(z)\boldsymbol{\mathcal{A}}(z))}, \tag{6}$$

where $\boldsymbol{\mathcal{A}}^*$ is the conjugate transpose of $\boldsymbol{\mathcal{A}}$, the reverse characteristic polynomial, defined in Assumption (A1). Basu and Michailidis (2015)[Proposition 2.2] shows that under (A1),

$$\max_{|z|=1} \Lambda_{\max}(\boldsymbol{\mathcal{A}}^*(z)\boldsymbol{\mathcal{A}}(z)) < \left[1 + \frac{\sum_{k=1}^{p}(\|\!|\boldsymbol{A}_k|\!\|_1 + \|\!|\boldsymbol{A}_k|\!\|_\infty)}{2}\right]^2.$$

Hence, (A5) is satisfied if, for instance, $\Lambda_{\min}(\boldsymbol{\Sigma}) > 0$, $\sum_{k=1}^{p} \|\!|\boldsymbol{A}_k|\!\|_1 < \infty$ and $\sum_{k=1}^{p} \|\!|\boldsymbol{A}_k|\!\|_\infty < \infty$.

3. Illustration

In this section we illustrate processes satisfying Assumptions (A2) and (A3).

Example 1 (Strong mixing sequences). *Let* $\{\mathbf{u}_t\}$ *denote a martingale difference, strong mixing sequence with coefficients* $\alpha_m < b_1 \exp(-b_2 m^{\gamma_2})$ *and common covariance matrix* Σ *with eigenvalues bounded away from zero and infinity, uniformly in* n*. It follows that* $r_t = \mathrm{b}_1'\mathbf{u}_t\mathbf{u}_t'\mathrm{b}_2$ *is also strong mixing of same size and, from (Davidson, 1994, Theorem 14.2),* $\mathbb{E}[r_t - \mathbb{E}(r_t)|\mathcal{F}_{t-m}] \leq a_1 \exp(-a_2 m^{\gamma_2})$*, for constants* a_1 *and* a_2*.*

Example 2 (L^1 near-epoch dependent process). *Let* $\{\mathbf{u}_t\}$ *denote a weakly stationary, martingale difference sequence. Suppose* $\mathrm{b}'\boldsymbol{v}_t = \mathrm{b}'\mathsf{vech}\,(\mathbf{u}_t\mathbf{u}_t' - \boldsymbol{\Sigma})$ *is a centered,* L^1*-NED sequence on* $\mathcal{F}_t = \sigma\langle\epsilon_t, \epsilon_{t-1}, ...\rangle$*, where* $\{\epsilon_t\}$ *is* α*-mixing with coefficients* $\alpha_m \leq c_1 \exp(c_2 m^{\gamma_1})$*, for all* $\mathbf{b} \in \{\mathbf{b} \in \mathbb{R}^{n(n+1)/2} : |\mathbf{b}|_1 \leq 1\}$*. It means that there are finite constants* $\{d_t\}$ *and* $\{\psi_m\}$ *such that*

$$\mathbb{E}\,|\mathbf{b}'(\boldsymbol{v}_t - \mathbb{E}[\boldsymbol{v}_t|\mathcal{F}_{t-m:t}])| \leq d_t\psi_m,$$

where $\mathcal{F}_{t-m:t} = \sigma\langle\epsilon_t, ..., \epsilon_{t-m}\rangle$ *and* $\psi_m \leq \exp(c_3 m^{\gamma_2})$*. Under Assumption (A3), it follows from Wong et al. (2020, Lemma 5) and Hölder inequality that for any* $r < \infty$

$$\|\mathbf{b}'\boldsymbol{v}_t\|_r \leq |\mathbf{b}|_1^r \max_{1\leq i\leq j\leq n} \|u_{it}u_{jt}\|_r \leq \max_{1\leq i\leq n} \|u_{it}\|_{2r} \leq c_4 r^{1/\alpha}.$$

Finally, it follows from Andrews (1988, Example 6) that Assumption (A2) holds with $a_1 \geq (2\max_t d_t + c_4 r^{1/\alpha})(e^{c_3/2^{\gamma_2}} + 6c_1 e^{c_2(r-1)/r2^{\gamma_2}})$ *and* $a_2 \leq (c_3 \wedge c_2(r-1)/r)/2^{\gamma_2}$*.*

Example 3 (Linear process in the variance). *Let* $\{\boldsymbol{v}_t\}$ *denote a sequence of centered independently and identically distributed, sub-Weibull random variables taking values in* $\mathbb{R}^n$ *with identity covariance matrix. Let* $\mathbf{u}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{v}_t$ *where* $\boldsymbol{H}_t^{1/2}$ *is the Cholesky decomposition of* $\boldsymbol{H}_t$ *and*

$$\boldsymbol{h}_t = \mathsf{vech}\,(\boldsymbol{H}_t) = c + \sum_{j=1}^{\infty} \Psi_j \boldsymbol{\eta}_{t-j}.$$

Here, $\mathsf{vech}\,(\boldsymbol{M})$ *stacks the lower diagonal elements of matrix* $\boldsymbol{M}$ *and* $\boldsymbol{\eta}_t = \mathsf{vech}\,(\boldsymbol{v}_t\boldsymbol{v}_t')$ *and* $\{\Psi_j\}$ *satisfy* $\sum_{j=m}^{\infty} |\tilde{\mathbf{b}}'\Psi_j|_1 \lesssim e^{-a_2 m^{\gamma_2}}$ *for all* $\tilde{\mathbf{b}} \in \{\mathbf{b} \in \mathbb{R}^{n(n+1)/2} : |\mathbf{b}|_1 \leq 1\}$*.*

We first show $\{\mathbf{u}_t\}$ *is weakly stationary martingale difference with respect to* $\mathcal{F}_{t-1} = \sigma\langle\boldsymbol{v}_{t-1}, \boldsymbol{v}_{t-2}, ...\rangle$*. First* $\mathbb{E}[\mathbf{u}_t|\mathcal{F}_{t-1}] = \boldsymbol{H}_t^{1/2}\mathbb{E}[\boldsymbol{v}_t|\mathcal{F}_{t-1}] = 0$*. The covariance is given by*

$$\mathbb{E}[\mathbf{u}_t\mathbf{u}_t'] = \mathbb{E}[\boldsymbol{H}_t^{1/2}\mathbb{E}(\boldsymbol{v}_t\boldsymbol{v}_t'|\mathcal{F}_{t-1})(\boldsymbol{H}_t^{1/2})'] = \mathbb{E}[\boldsymbol{H}_t].$$

Now, $\mathbb{E}[\boldsymbol{h}_t] = C + \sum_{j=1}^{\infty} \Psi_j \mathbb{E}\boldsymbol{\eta}_{t-j} = C + \sum_{j=1}^{\infty} \Psi_j \mathsf{vech}\,(I_n) = \boldsymbol{\Sigma}$*, where* $E(\boldsymbol{\eta}_t) = \mathsf{vech}\,(\mathbb{E}(\boldsymbol{v}_t\boldsymbol{v}_t')) = \mathsf{vech}\,(I_n)$ *for all* t*.*

For constant vectors $\mathbf{b}_1, \mathbf{b}_2 \in \{\mathbf{b} \in \mathbb{R}_n : |\mathbf{b}| \leq 1\}$,

$$\begin{aligned}\mathbb{E}[\mathbf{b}_1'(\mathbf{u}_t\mathbf{u}_t' - \mathbf{\Sigma})\mathbf{b}_2|\mathcal{F}_{t-1}] &= \mathbf{b}_1'(\boldsymbol{H}_t - \mathbb{E}\boldsymbol{H}_t)\mathbf{b}_2 \\ &= \tilde{\mathbf{b}}'(\boldsymbol{h}_t - \mathbb{E}\boldsymbol{h}_t) \\ &= \sum_{j=1}^{\infty} \tilde{\mathbf{b}}'\Psi_j(\boldsymbol{\eta}_{t-j} - \mathbb{E}\boldsymbol{\eta}_{t-j}),\end{aligned}$$

where $\tilde{\mathbf{b}} \in \{\mathbf{b} \in \mathbb{R}^{n(n+1)/2} : |\mathbf{b}|_1 \leq 1\}$*. It follows that*

$$\begin{aligned}\mathbb{E}\left|\mathbb{E}[\mathbf{b}_1'(\mathbf{u}_t\mathbf{u}_t' - \mathbf{\Sigma})\mathbf{b}_2|\mathcal{F}_{t-m}]\right| &= \mathbb{E}\left|\sum_{j=1}^{\infty} \tilde{\mathbf{b}}'\Psi_j\mathbb{E}(\boldsymbol{\eta}_{t-j} - \mathbb{E}\boldsymbol{\eta}_{t-j}|\mathcal{F}_{t-m})\right| \\ &= \left\|\sum_{j=m}^{\infty} \tilde{\mathbf{b}}'\Psi_j(\boldsymbol{\eta}_{t-j} - \mathbb{E}\boldsymbol{\eta}_{t-j})\right\|_1 \\ &\leq 2\left(\sum_{j=m}^{\infty} |\tilde{\mathbf{b}}'\Psi_j|_1\right) \max_{|\mathbf{b}|_1 \leq 1} \|\mathbf{b}'\boldsymbol{v}_t\|_2^2,\end{aligned}$$

where in the last line we use the same arguments of Lemma 3 in the appendix, followed by the triangle inequality and $\|\delta'\boldsymbol{\eta}_t\| \leq \max_{|\mathbf{b}_1|\leq 1, |\mathbf{b}_2|_1 \leq 1} \|\mathbf{b}_1'\boldsymbol{v}_t\boldsymbol{v}_t'\mathbf{b}_2\| \leq \max_{|\mathbf{b}|\leq 1} \|\mathbf{b}'\boldsymbol{v}_t\|_2^2$*. Then, Assumption (A2) is satisfied under the condition that* $\sum_{j=m}^{\infty} |\tilde{\mathbf{b}}'\Psi_j|_1 \lesssim e^{-a_2 m^{\gamma_2}}$ *and* $\|\mathbf{b}'\boldsymbol{v}_t\|_2 < \infty$*.*

As to assumption (A3), it follows from (Wong et al., 2020, Lemma 5) that we have to show that $\sup_{d\geq 1} d^{-1/\alpha}\|\mathbf{b}'\mathbf{u}_t\|_d < c_\alpha < \infty$*. For any* $d \geq 1$,

$$\begin{aligned}\|\mathbf{b}'\mathbf{u}_t\|_d &= \|\mathbf{b}'\mathbf{u}_t\mathbf{u}_t'\mathbf{b}\|_{d/2}^{1/2} \\ &= \|\mathbf{b}'\boldsymbol{H}_t^{1/2}\boldsymbol{v}_t\boldsymbol{v}_t'(\boldsymbol{H}_t^{1/2})\mathbf{b}\|_{d/2}^{1/2} \\ &= \left\|\mathbf{b}'\boldsymbol{H}_t\mathbf{b} \times \frac{\mathbf{b}'\boldsymbol{H}_t^{1/2}\boldsymbol{v}_t\boldsymbol{v}_t'(\boldsymbol{H}_t^{1/2})\mathbf{b}}{\mathbf{b}'\boldsymbol{H}_t\mathbf{b}}\right\|_{d/2}^{1/2} \\ &\leq \left\{\mathbb{E}\left(|\mathbf{b}'\boldsymbol{H}_t\mathbf{b}|^{d/2}\,\mathbb{E}\left[\left|\frac{\mathbf{b}'\boldsymbol{H}_t^{1/2}\boldsymbol{v}_t\boldsymbol{v}_t'(\boldsymbol{H}_t^{1/2})\mathbf{b}}{\mathbf{b}'\boldsymbol{H}_t\mathbf{b}}\right|^{d/2}\middle|\mathcal{F}_{t-1}\right]\right)\right\}^{1/d} \\ &\leq \left\{\mathbb{E}\left(|\mathbf{b}'\boldsymbol{H}_t\mathbf{b}|^{d/2} \sup_{\delta'\delta=1} \mathbb{E}[(\delta'\boldsymbol{v}_t\boldsymbol{v}_t'\delta)^{d/2}|\mathcal{F}_{t-1}]\right)\right\}^{1/d} \\ &= \|\mathbf{b}'\boldsymbol{H}_t\mathbf{b}\|_{d/2}^{1/2} \sup_{\delta'\delta=1} \|\delta'\boldsymbol{v}_t\|_d,\end{aligned}$$

where the last line follow by independence. Condition (A3) holds if $\|\mathbf{b}'\boldsymbol{H}_t\mathbf{b}\|_{d/2}$ *is bounded independently of* d*. It follows that there is* $\tilde{\mathbf{b}} \in \{\mathbf{b} \in \mathbb{R}^{n(n-1)/2} : |\mathbf{b}|_1 \leq 1\}$

such that

$$\begin{aligned}\|\mathbf{b}'\boldsymbol{H}_t\mathbf{b}\|_{d/2} &= \|\tilde{\mathbf{b}}'\boldsymbol{h}_t\|_{d/2} \\ &= \left\|\tilde{\mathbf{b}}'C + \sum_{j=1}^{\infty}\tilde{\mathbf{b}}'\Psi_j\boldsymbol{\eta}_{t-j}\right\|_{d/2} \\ &\le \left\|\tilde{\mathbf{b}}'C\right\|_{d/2} + \left\|\sum_{j=1}^{\infty}\tilde{\mathbf{b}}'\Psi_j\boldsymbol{\eta}_{t-j}\right\|_{d/2} \\ &\le \left\|\tilde{\mathbf{b}}'C\right\|_{d/2} + 2\left(\sum_{j=1}^{\infty}|\tilde{\mathbf{b}}'\Psi_j|_1\right)\max_{|\mathbf{b}|_1\le 1}\|\mathbf{b}'\boldsymbol{v}_t\|_{d/2}^2.\end{aligned}$$

4. LASSO ESTIMATION BOUNDS

Let $\mathcal{L}_T(\boldsymbol{\beta}_i) = \frac{1}{T}|\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}_i|_2^2$ denote the empirical squared risk, for each $i = 1, ..., n$. We estimate $\boldsymbol{\beta}_i$, $i = 1, \ldots, n$, equation-wise using the LASSO procedure

$$\widehat{\boldsymbol{\beta}}_i \in \arg\min_{\boldsymbol{\beta}_i\in\mathbb{R}^{np}}\left\{\mathcal{L}_T(\boldsymbol{\beta}_i) + \lambda_i|\boldsymbol{\beta}_i|_1\right\}, \quad i = 1, ..., n, \tag{7}$$

where λ_i are positive regularization parameters. For ease of exposition we assume $\lambda_1 = \cdots = \lambda_n = \lambda$. It is well known that $\boldsymbol{\beta}_i^* = \arg\min_{\boldsymbol{\beta}_i}\mathbb{E}\left\{\mathcal{L}_T(\boldsymbol{\beta}_i)\right\}$ are the population parameters in (5), under stated conditions.

We follow the steps in Negahban et al. (2012) to derive error bounds for the equation-wise LASSO estimator. First define the pair of subspaces $\mathcal{M}(S) = \{\boldsymbol{u} \in \mathbb{R}^{np}|u_i = 0, i \in S^c\}$ and its orthogonal complement $\mathcal{M}^{\perp}(S) = \{\boldsymbol{u} \in \mathbb{R}^{np}|u_i = 0, i \in S\}$, where $S \subseteq \{1, \ldots, np\}$. Set $\boldsymbol{u}_{\mathcal{M}}$ and $\boldsymbol{u}_{\mathcal{M}^{\perp}}$ the projection of $\boldsymbol{u}$ on $\mathcal{M}(S)$ and $\mathcal{M}^{\perp}(S)$, respectively. Clearly, for any $\boldsymbol{u} \in \mathbb{R}^{np}$, $|u|_1 = |\boldsymbol{u}_{\mathcal{M}}|_1 + |\boldsymbol{u}_{\mathcal{M}^{\perp}}|_1$. We say $|\cdot|_1$ is decomposable with respect to the pair $(\mathcal{M}(S), \mathcal{M}^{\perp}(S))$ for any set $S \subset \{1, \ldots, np\}$.

We have to show two conditions to obtain a finite sample estimation error bound for the parameter vectors. The first condition is known as *restricted strong convexity* (RSC) and restricts the geometry of the loss function around the optimum $\boldsymbol{\beta}^*$ and is related to the Restricted Eigenvalue (Van De Geer et al., 2009). The second condition is known as *deviation bound* and restricts the size of the sup-norm of the gradient $\nabla\mathcal{L}_T(\boldsymbol{\beta}^*)$.

Definition (Deviation Bound (DB)). *The* deviation bound *condition holds if the event* $\{\lambda \ge 2|\mathbf{X}'\boldsymbol{U}_i/T|_\infty\}$ *occurs with high probability for all* $i = 1, ..., n$.

Note that one may adopt individual λ_is for each equation, in which above definition should be modified adequately.

Definition (Restricted Strong Convexity (RSC))**.** *Define* $\mathbb{C}(\boldsymbol{\beta}^*, \mathcal{M}, \mathcal{M}^\perp) = \{\boldsymbol{\Delta} \in \mathbb{R}^{np} || \boldsymbol{\Delta}_{\mathcal{M}^\perp}|_1 \leq 3|\boldsymbol{\Delta}_{\mathcal{M}}|_1 + 4|\boldsymbol{\beta}^*_{\mathcal{M}^\perp}|_1\}$*. The* restricted strong convexity *holds for parameters* $\kappa_{\mathcal{L}}$ *and* $\tau_{\mathcal{L}}$ *if for any* $\boldsymbol{\Delta} \in \mathbb{C}$*,*

$$\frac{\boldsymbol{\Delta}'\mathbf{X}'\mathbf{X}\boldsymbol{\Delta}}{T} \geq \kappa_{\mathcal{L}} |\boldsymbol{\Delta}|_2^2 - \tau_{\mathcal{L}}^2(\boldsymbol{\beta}^*).$$

Negahban et al. (2012)[Section 4] show these conditions are satisfied by many loss functions and penalties. Basu and Michailidis (2015) show that both DB and RSC are satisfied by Gaussian VAR(p) models in high dimensions.

If both DB and RSC hold with large probability, Negahban et al. (2012)[Theorem 1] provides an l_2 estimation bound for $\widehat{\boldsymbol{\beta}}_i$. Our goal is to show that the error bounds are valid for each $\boldsymbol{\Delta}_i = \widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}^*_i$, $i = 1, \ldots, n$ at the same time.

Lemma 1 characterizes the solutions of the optimization program in (7). We require further notation. Define $\mathbb{C}_i := \mathbb{C}(\boldsymbol{\beta}^*_i, \mathcal{M}_{i,\eta}, \mathcal{M}^\perp_{i,\eta})$ for a pair of subsets $\mathcal{M}_{i,\eta} = \mathcal{M}(S_{i,\eta})$ and $\mathcal{M}^\perp_{i,\eta} = \mathcal{M}^\perp(S_{i,\eta})$, where $S_{i,\eta} = \{j \in \{1, ..., pn\} || \beta_{i,j}| > \eta\}$ and $S^c_{i,\eta} = \{j \in \{1, ..., pn\} || \beta_{i,j}| \leq \eta\}$. These sets represent the *active parameters* under weak sparsity. In Theorem 1 we set $\eta = \lambda / \sigma^2_\Gamma$ to derive our results.

Lemma 1. *Suppose* $\{\mathbf{y}_t\}$ *is generated from* (1) *and Assumptions (A1), (A2) and (A3) are satisfied. Let*

$$\lambda > \tau^*(\epsilon + \log(Tn^2p))^{2/\alpha} \sqrt{\frac{\epsilon + \log(n^2p)}{T}},$$

where $\tau^* > 0$ *is depends on* τ*,* α *and* $\bar{c}_\Phi$*, and any* $\epsilon > 0$*. Then, if* $T > \epsilon + \log(n^2p)$*, the event* $\left\{\forall i = 1, ..., n : \ \widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}^* \in \mathbb{C}_i\right\}$ *holds with probability at least* $1 - \pi_1(\epsilon)$ *with* $\pi_1(\epsilon) = 10e^{-\epsilon}$*.*

Lemma 1 shows that under restrictions on λ the solutions to the optimization program in (7) lie inside the star-shaped sets $\mathbb{C}_i$ with high probability, as the sample size increases. It restricts the directions in which we should control the variation of our estimators. Next result shows the deviation bound holds with high probability for appropriate choice of λ. To formalize the idea, let

$$\mathcal{D}_i(\lambda) = \left\{\lambda \geq 2\left|\frac{1}{T}\boldsymbol{X}'\mathbf{U}_i\right|_\infty\right\}, \quad i = 1, ..., n, \tag{8}$$

denote the event "*DB holds for equation* i *with regularization parameter* λ*.*"

Proposition 1 (Deviation Bound)**.** *Suppose* $\{\mathbf{y}_t\}$ *is generated from* (1)*, and Assumptions (A1), (A2) and (A3) are satisfied and* $T > \epsilon + \log(n^2p)$ *for some* $\epsilon > 0$*. Set*

$$\lambda > \tau^*(\epsilon + \log(Tn^2p))^{2/\alpha} \sqrt{\frac{\epsilon + \log(n^2p)}{T}},$$

for some $\tau^* > 0$ *and any* $\epsilon > 0$. *Then,* $\Pr(\cup_{i=1}^n \mathcal{D}_i^c) \leq 10e^{-\epsilon}$.

Suppose $\epsilon = \log(np)$, $n^2 p > T$. The regularization parameter λ satisfies

$$\lambda \gtrsim [\log(np)]^{2/\alpha} \sqrt{\frac{\log(np)}{T}},$$

and $\pi_1(\lambda) \propto 1/n^2 p$. This regularization parameter is $O([\log(np)]^{2/\alpha})$ larger, in rate, than one obtained in Wong et al. (2020, Proposition 7). Their results relied heavily in $\mathbf{y}_t$ being a β-mixing sequence in a sense that the concentration inequality derived in Merlevède et al. (2011) depends on it. In our case, the dependence is characterized by the conditional variance of the innovation process and coefficients $\Phi_1, \Phi_2, ...$, and we are not aware of "tight" concentration inequalities that hold under these assumptions. Nevertheless, for fixed n, it is possible to show that the concentration inequality for sub-Weibull martingales in Lemma 5 is tight (Fan et al., 2012b).

Let $\mathbf{\Gamma}_T = \mathbf{X}'\mathbf{X}/T$ denote the scaled Gram matrix and $\mathbf{\Gamma}$ its expected value. We show that if each element in $\mathbf{\Gamma}_T$ is sufficiently close to its expectation, and Assumptions (A4) and (A5) hold, then RSC is satisfied with high probability.

Lemma 2 (Restricted Strong Convexity). *Suppose Assumptions (A4) – (A5) hold and that* $|||\mathbf{\Gamma}_T - \mathbf{\Gamma}|||_{\max} \leq \frac{\sigma_\Gamma^2 \eta^q}{64 R_q}$. *Then, for any* $\Delta_i \in \mathbb{C}_i$,

$$\Delta_i' \Phi_T \Delta_i \geq \frac{\sigma_\Gamma^2}{2} |\Delta_i|_2^2 - \frac{\sigma_\Gamma^2}{2} R_q \eta^{2-q}. \tag{9}$$

To show RSC holds with high probability for all $i = 1, .., n$ at the same time, we have to bound the event

$$\mathcal{B}(a) = \left\{ |||\mathbf{\Gamma}_T - \mathbf{\Gamma}|||_{\max} \leq a \right\}. \tag{10}$$

where $a = \frac{\sigma_\Gamma^{2(1-q)} \lambda^q}{64 R_q}$. If we assume distinct $R_{q,i}$ and q_i for each equation, we should work with $\cap_i \mathcal{B}_i$ and $\mathcal{B}_i$ defined accordingly.

Proposition 2. *Suppose Assumptions (A1), (A2) and (A3) hold. If*

$$p < \frac{T^{\gamma_1 \wedge \gamma_2}}{(\frac{2}{\gamma_1 \wedge \gamma_2 + 1})(2 + \frac{1.4}{2\gamma_1 c_\phi \wedge a_2})},$$

and

$$a \geq \sqrt{\frac{2(1+\xi)^{1+2/\alpha} \tau^2 [\log(npT)]^{1+2/\alpha}}{T}},$$

for some $\xi > 0$, *then* $\Pr(\mathcal{B}^c(a)) \leq \pi_2(a)$, *where*

$$\begin{aligned}\pi_2(a) &:= \frac{2}{(np)^\xi T^{1+\xi}} + \frac{8}{(np)^\xi T^\xi} \\ &+ \frac{n^2}{a}\left(b_1 e^{-c_\phi \wedge a_2 (T/2)^{\gamma_2 \wedge \gamma_1}} + b_8 e^{-2\gamma_1 c_\phi (T/2)^{\gamma_1}}\right).\end{aligned}$$

This bound controls the proximity between the empirical and population covariance matrices. Similar concentration inequalities were derived by (Kock and Callot, 2015, Lemma 9), (Loh and Wainwright, 2012, Lemma 14) and Medeiros and Mendes (2016a). Their results, however, cannot be applied in our setting. Explicit expressions for the constants b_1, b_8 and τ in Proposition 2 are found in Lemma 6. Also, one may replace ϵ by its lower bound to remove dependence.

This concentration guided the choice of dependence condition used in this work. Traditionally one uses either a Hanson-Wright inequality or a Bernstein or Hoeffding type inequality to bound the empirical covariance around its mean. We write the centered Gram matrix $\mathbf{\Gamma}_T - \mathbf{\Gamma}$ as a sum of martingales and a dependence term. The martingales are handled using a Bernstein type bound and the dependence term is handled using both assumptions (A1) and (A2). Combined, they imply a sub-Weibull type decay on expected value of dependence term.

Finally, we use the bounds π_1 and π_2 in Proposition 1 and Proposition 2 to show that equation wise (nodewise) lasso regressions are close to their population counterpart in l_2.

Theorem 1. *Suppose assumptions (A1) – (A5) hold. Set* $\eta = \lambda/\sigma_\Gamma^2$. *Under conditions of Propositions 1 and 2, there exists* $T_0 > 0$ *such that for all* $T \geq T_0$,

$$|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^*|_2^2 \leq (44 + 2\lambda) R_q \left(\frac{\lambda}{\sigma_\Gamma^2}\right)^{2-q}, \quad i = 1, ..., n,$$

in a set with probability at least $1 - \pi_1(\log(np)) - \pi_2\left(\frac{\sigma_\Gamma^{2(1-q)}\lambda^q}{64 R_q}\right)$.

Theorem 1 states that, with high probability, estimated and population parameter vectors are close to each other in the Euclidean norm. It requires that Propositions 1 and 2 hold jointly, meaning that λ, R_q and σ_Γ^2 must satisfy rate conditions. We show that if the size of 'small' coefficients and smallest eigenvalue σ_Γ^2 of Γ are restricted, then the rate of λ in Proposition 1 is unaffected. For $\epsilon = \log(np)$ and $T < np^2$ Proposition 1 requires, after simplification,

$$\lambda \geq \tau^* \log(np)^{2/\alpha} \sqrt{\frac{4\log(np)}{T}},$$

for some constant τ^*. Replacing a by $\frac{\sigma_\Gamma^{2(1-q)}\lambda^q}{64R_q}$ in Proposition 2 we obtain

$$\lambda^q \gtrsim \left(\log(np)^{2/\alpha}\sqrt{\frac{\log(np)}{T}}\right) \times \left(\frac{R_q}{\sigma_\Gamma^{2(1-q)}\log(np)^{1/\alpha}}\right).$$

However, it is not necessarily a constraint in the rate of λ. Propositions 1 and 2 will hold jointly for T sufficiently large for $0 \leq q < 1$ if

$$\frac{R_q}{\sigma_\Gamma^{2(1-q)}} = o\left(\log(np)^{(2q-1)/\alpha}\left(\frac{T}{\log(np)}\right)^{(1-q)/2}\right).$$

In other words, if the *small* parameters are not too large and smallest eigenvalue of $\mathbf{\Sigma}$ is not too small as a function of T.

5. Discussion

This work provides finite sample l_2 error bounds for the equation-wise LASSO parameters estimates of a weakly sparse, high-dimensional, VAR(p) model, with dependent and heavy tailed innovation process. It covers a large collection of specifications as illustrated in section 3.

A distinctive feature this work is that the dependence structure of the innovations are characterized by a very weak projective dependence condition that is naturally verifiable in settings where one is interested in the conditional variance of the process. The series of innovations is not necessarily mixing or near-epoch dependent, nor the resulting time series $\{\mathbf{y}_t\}$.

Our bounds hold under a heavy tailed setting in a sense that we do not require the moment generating function to exist. Despite the tails in $\{\mathbf{y}_t\}$ being sub-Weibull as in Wong et al. (2020), we are not able to recover the same rates and lower bound for the regularization parameter λ. The reason is that Wong et al. (2020) bounds rely heavily on the concentration inequality for mixing sequences in Merlevède et al. (2011). Given the weak projective dependence adopted, we chose to use a martingale concentration and overcome all together the issue of using the dependence metric for deriving the concentration bound. Nevertheless, we believe the loss in efficiency is minimal. Close inspection of proof of Lemma 5 shows that the loss of efficiency is concentrated in bounding the tail. It amounts to an extra $\log(T)$ term, which does not change the rates under assumption that $T < n^2p$, eventually.

A limitation of this work is the restriction that the model is, almost, correctly specified in the mean, in a sense that innovations are martingale differences. Nevertheless, this assumption is standard in the literature and we are able to derive results covering a broad range of data generating processes and conditional dependence measures. The martingale difference condition cannot be relaxed at this moment as our deviation bound depends on it. Furthermore, we do not require strong sparsity in a sense that near zero coefficients are effectively treated as zero as long as they are concentrated in some slowly increasing l_q ball ($0 \leq q < 1$) around the origin.

Results in this paper can be easily extended to polynomial tails. The strategy is to replace the martingale concentration in Lemma 4, used to prove Propositions 1 and 2 by

$$\Pr\left(\max_{1\leq i\leq n}|\sum_{t=1}^{T}\xi_{it}| > Ta\right) \leq \frac{nK}{(a\sqrt{T})^d},$$

whenever $\|\xi_{it}\|_d < \infty$. If available under our dependence conditions, one could employ a Fuk-Nagaev type inequality. Nevertheless, if follows that under appropriate changes to concentration rates, equation-wise LASSO estimators also admit oracle bounds. A direct consequence is that moments conditions on Carrasco and Chen (2002) and Hafner and Preminger (2009a,b) are directly applicable.

Despite working with a relatively simple structure and estimation model, the machinery can be applied to more complex settings. The key points are showing that the empirical covariance concentrates around its mean in terms of its maximum entry-wise norm and the concentration inequality for large dimensional, sub-Weibull martingales. Following development of Negahban et al. (2012), the results may be naturally extended to structured regularization with node-wise regression and replacing using the Frobenius norm for system estimation. Finally, the HD VAR specification encompasses large dimensional vector-panels among other models.

Appendix A. Proof of main results

A.1. **Proof of Lemma 1.** We apply Negahban et al. (2012, Lemma 1). The empirical loss $\mathcal{L}_T(\boldsymbol{\beta}_i)$ is convex for each i. Proposition 1 ensures each (8) hold with desired probabilities. □

A.2. **Proof of Proposition 1.** Write the event $\mathcal{A}_i = \{\max_j |\mathbf{u}_i'\mathbf{x}_j| < T\lambda_0/2\}$. We shall derive probability bounds for $\Pr(\cap_{i=1}^n \mathcal{A}_i) \geq 1 - \Pr(\max_{i,j} |\mathbf{u}_i'\mathbf{x}_j| \geq T\lambda_0/2)$. We bound the probability using Corollary 1.

Under Assumption (A2), $\mathbf{u}_i'\mathbf{x}_j = \sum_{t=p}^T u_{ti}y_{t-s,j}$ ($s = 1, ..., p$ and $i, j = 1, ..., n$) is a martingale and each $u_{ti}y_{t-s,j}$ is a martingale difference process. Hence, we follow by applying Corollary 1. Conditions on T and $\lambda_0/2$ are already satisfied. We need to show that $u_{ti}y_{t-s,j}$ is sub-Weibull. For each $d \geq 1$, $\|u_{ti}y_{t-s,j}\|_d \leq \|u_{ti}\|_{2p}^{1/2}\|y_{t-s,j}\|_{2p}^{1/2} \leq \bar{c}_\Phi \max_{|\mathbf{b}|_1 \leq 1} \|\mathbf{b}'\mathbf{u}\|_{2p}$, by Lemma 3. Then, it follows from Wong et al. (2020, Lemma 5 and Lemma 6) and Assumption (A3) that $u_{ti}y_{t-s,j}$ is sub-Weibull with parameter $\alpha/2$. Hence, there is some constant τ^* depending on τ, $\bar{c}_\Phi$ and α such that $\Pr(|u_{ti}y_{t-s,j}| > x) \leq 2\exp(-|x/\tau^*|^{\alpha/2})$. Result follows.

A.3. **Proof of Lemma 2.** For notational simplicity, write $\|\boldsymbol{\Phi}_T - \Gamma\|_{\max} \leq \delta \leq \sigma_\Gamma^2/64\psi^2(\mathcal{M}_{i,\eta})$ where $\psi(\mathcal{M}_{i,\eta}) = \mathsf{sup}_{u \in \mathbb{C}_i} |u|_1/|u|_2 = \sqrt{|S_{i,\eta}|}$. Using the arguments in (Negahban et al., 2012, section 4.3), $|\boldsymbol{\beta}_{i,\mathcal{M}_{i,\eta}^\perp}|_1 \leq \sum_{j \in S_{i,\eta}^c} |\beta_{i,j}|^q |\beta_{i,j}|^{1-q} \leq \eta^{1-q}R_q$, and $R_q \geq \sum_{j \in S_{i\eta}} |\beta_{i,j}|^q \geq |S_{i,\eta}|\eta^q$. Hence $\frac{\sigma_\Gamma^2\eta^q}{64R_q} \leq \frac{\sigma_\Gamma^2}{64\psi^2(\mathcal{M}_{i,\eta})}$. It follows that

$$\begin{aligned}
\Delta_i'\boldsymbol{\Psi}_T\Delta_i &= \Delta_i'\Gamma\Delta_i + \Delta_i'[\boldsymbol{\Phi}_T - \Gamma]\Delta_i \\
&\geq |\Delta_i|_2^2 \inf_{u \in \mathbb{C}_i \setminus \{0\}} \frac{u'\Gamma u}{u'u} - |\Delta_i|_1 |[\boldsymbol{\Psi}_T - \Gamma]\Delta_i|_\infty \\
&\geq \sigma_\Gamma^2 |\Delta_i|_2^2 - |\Delta_i|_1^2 \|\boldsymbol{\Phi}_T - \Gamma\|_{\max} \\
&\geq \sigma_\Gamma^2 |\Delta_i|_2^2 - \delta|\Delta_i|_1^2 \\
&\geq \sigma_\Gamma^2 |\Delta_i|_2^2 - \delta\left(4|\Delta_{i,\mathcal{M}_{i,\eta}}|_1 + 4|\boldsymbol{\beta}_{i,\mathcal{M}_{i,\eta}^\perp}|_1\right)^2 \\
&\geq |\Delta_i|_2^2 \left(\sigma_\Gamma^2 - 32\delta\psi(\mathcal{M}_{i,\eta})^2\right) + 32\delta|\boldsymbol{\beta}_{i,\mathcal{M}_{i,\eta}^\perp}|_1^2 \\
&\geq |\Delta_i|_2^2 \frac{\sigma_\Gamma^2}{2} - \frac{\sigma_\Gamma^2}{2\psi^2(\mathcal{M}_{i,\eta})}|\boldsymbol{\beta}_{i,\mathcal{M}_{i,\eta}^\perp}|_1^2 \\
&\geq |\Delta_i|_2^2 \frac{\sigma_\Gamma^2}{2} - \frac{\sigma_\Gamma^2}{2R_q\eta^{-q}}\eta^{2(1-q)}R_q^2 \\
&= |\Delta_i|_2^2 \frac{\sigma_\Gamma^2}{2} - \frac{\sigma_\Gamma^2}{2}\eta^{2-q}R_q,
\end{aligned}$$

proving the result. □

A.4. **Proof of Proposition 2.** The proof consists on a trivial application of Lemmas 6 setting $\epsilon = \sigma_\Gamma^{2(1-q)}\lambda^q/64R_q$. □

A.5. **Proof of Theorem 1.** We apply (Negahban et al., 2012, Theorem 1). Lemma 1 ensures λ is selected accordingly, $\mathcal{L}_T(\boldsymbol{\beta}_i)$ is a convex function of $\boldsymbol{\beta}_i$, Lemma 2 ensures RSC is satisfied with $\kappa_{\mathcal{L}} = \sigma_\Gamma^2/2$ and $\tau_{\mathcal{L}}^2(\boldsymbol{\beta}_i) = \frac{\sigma_\Gamma^2 \eta^{2-q} R_q}{2}$. Define $\psi(\mathcal{M}_{i,\eta})$ as in the proof of Lemma 2 and recall $|S_{i,\eta} \leq R_q \eta^{-q}$ and $|\boldsymbol{\beta}_{i,\mathcal{M}_{i,\eta}^\perp}|_1 \leq R_q \eta^{1-q}$, and that $\eta = \lambda/\sigma_\Gamma^2$. For each i,

$$
\begin{aligned}
|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}^*|_2^2 &\leq 9\frac{\lambda}{\kappa_{\mathcal{L}}^2}\psi^2(\mathcal{M}_{i,\eta}) + \frac{\lambda}{\kappa_{\mathcal{L}}}\left[2\tau_{\mathcal{L}}^2(\boldsymbol{\beta}_i^*) + 4|\boldsymbol{\beta}_{i,\mathcal{M}_{i,\eta}^\perp}|_1\right] \\
&\leq 36\frac{\lambda}{\sigma_\Gamma^4} R_q \eta^{-q} + 2\frac{\lambda}{\sigma_\Gamma^2}\left[R_q \eta^{2-q}\sigma_\Gamma^2 + 4R_q\eta^{1-q}\right] \\
&\leq 36\frac{\lambda}{\sigma_\Gamma^4} R_q \left(\frac{\lambda}{\sigma_\Gamma^2}\right)^{-q} + 2\frac{\lambda}{\sigma_\Gamma^2}\left[R_q\left(\frac{\lambda}{\sigma_\Gamma^2}\right)^{2-q}\sigma_\Gamma^2 + 4R_q\left(\frac{\lambda}{\sigma_\Gamma^2}\right)^{1-q}\right] \\
&\leq 36R_q\left(\frac{\lambda}{\sigma_\Gamma^2}\right)^{2-q} + 2\lambda R_q\left(\frac{\lambda}{\sigma_\Gamma^2}\right)^{2-q} + 8R_q\left(\frac{\lambda}{\sigma_\Gamma^2}\right)^{2-q} \\
&= (44 + 2\lambda)R_q\left(\frac{\lambda}{\sigma_\Gamma^2}\right)^{2-q}.
\end{aligned}
$$

□

Appendix B. Auxiliary Lemmata

B.1. **Properties of $\mathbf{y}_t$.** In this section we will derive properties of the process $\{\mathbf{y}_t\}$ described in (1)

Lemma 3. *Suppose that for some norm $\|\cdot\|_\psi$ we have*

$$\max_t \max_{|\mathbf{b}|_1 \leq 1} \|\mathbf{b}'\mathbf{u}_t\|_\psi \leq c_\psi,$$

for some constant $c_\psi < \infty$ that only depends on the norm $\|\cdot\|_\psi$. Then, under conditions (A1) - (A2), for all t and $i \in \{1, \ldots, n\}$,

$$\|y_{i,t}\|_\psi \leq c_\Phi \times \sum_{j=0}^{\infty} |e_i'\Phi_j|_1.$$

Proof. Under assumption (A1) the VAR model in (1) admits the VMA(∞) representation (4) for all n and p. Let $\{e_i = (0, ..., 0, 1, 0, ..., 0)', i = 1, ..., n\}$ the canonical basis vectors. Then, for all i, $y_{i,t} = e_i'\mathbf{y}_t$ and

$$\begin{aligned}
\|e_i'\mathbf{y}_t\|_\psi &= \left\|\sum_{j=0}^{\infty} e_i'\Phi_j\mathbf{u}_{t-j}\right\|_\psi \\
&= \left\|\sum_{j=0}^{\infty}\sum_{k=1}^{n} e_i'\Phi_j e_k u_{k,t-j}\right\|_\psi \\
&= \left\|\sum_{j=0}^{\infty} |e_i'\Phi_j|_1 \sum_{k=1}^{n} \frac{e_i'\Phi_j e_k}{|e_i'\Phi_j|_*} u_{k,t-j}\right\|_\psi \\
&\le \left(\sum_{j=0}^{\infty} |e_i'\Phi_j|_1\right) \max_t \max_{|\mathbf{b}|_1\le 1} \|\mathbf{b}'\mathbf{u}_t\|_\psi \\
&\le \sum_{j=0}^{\infty} |e_i'\Phi_j|_1 \times c_\psi,
\end{aligned}$$

where $|\cdot|_* := |\cdot|_1 I(|\cdot| > 0) + I(|\cdot|_1 = 0)$. □

Due stability condition (A1), for each n and p, there exists $\bar{c}_\Phi$ such that $\sum_{i=0}^{\infty} |\phi_{i,\delta}|_1 \le \bar{c}_\Phi$ for all $\delta = 1, ..., n$. Let $\|\cdot\|_\psi$ be the Orlicz norm,

$$\|\cdot\|_\psi = \inf\{c > 0 : \psi(|\cdot|/c) \le 1\},$$

where $\psi(\cdot) : \mathbb{R}^+ \mapsto \mathbb{R}^+$ is convex, increasing function with $\psi(0) = 0$ and $\psi(x) \to \infty$ as $x \to \infty$. Traditional choices of $\psi(\cdot)$ are (a) $\psi(x) = x^p$, $p \ge 1$, (b) $\psi(x) = \exp(x^a) - 1$, $a > 1$, and (c) $\psi(x) = (ae)^{1/a} x I(x \le a^{-1/a}) + \exp(x^a) I(x > a^{-1/a})$. These choices contemplate sub-Gaussian and sub-exponential tails, as well as process with heavy-tails, such as sub-Weibull and polynomial tails. Note that by combining this result with (Wong et al., 2020, Lemma 5 and Lemma 6) if $\{\mathbf{b}'\mathbf{u}_t\}$ are sub-Weibull, so are $\{\mathbf{b}'\mathbf{y}_t\}$.

Assumption (A1) is satisfied under restrictions on the parameter space. The stability assumption is standard in the literature whereas the tail sum (3) requires further constraints on the parameter matrices. Lemma 4 presents a sufficient set of restrictions on the sparse parameter matrices $\boldsymbol{A}_1, ..., \boldsymbol{A}_p$ so that (3) is satisfied.

Lemma 4. *Suppose that for all n and p, there exists some $\rho > 0$ such that*

$$\sum_{k=1}^{p} |||\boldsymbol{A}_k|||_\infty = \sum_{k=1}^{p} \max_{j=1,...,n} |\boldsymbol{a}_{k,j}|_1 \le e^{-\rho},$$

where $\boldsymbol{A}_k = [\boldsymbol{a}_{k,1} : \cdots : \boldsymbol{a}_{k,n}]'$. Then for every $\delta = 1, ..., n$,

i. $|\phi_{k,\delta}|_1 \le \sum_{j=1}^{p\wedge k} |||\boldsymbol{A}_j|||_\infty |\phi_{k-j,\delta}|_1$, $k = 1, 2, ...$

ii. $\sum_{k=m}^{\infty} |\phi_{k,\delta}|_1 \leq c_0\, e^{-m\rho}$, $m \geq 1$, *provided that for all* p,

$$\max_{\delta=1,\ldots,n}\max_{k=1,\ldots,p} e^{k\rho} \times \sum_{j=1}^{k} \tilde{\alpha}_j |\phi_{j,\delta}|_1 \leq (1-e^{-\rho})c_0, \tag{B.1}$$

where $\alpha_i = e^{\rho}|\boldsymbol{A}_i|_\infty$ *and* $\tilde{\alpha}_i = \sum_{\boldsymbol{i}:|\boldsymbol{i}|=k-p+j}\prod_{l=1}^{k-p}\alpha_{i_l}$ *where* $\boldsymbol{i} = (i_1,\ldots,i_{k-p})$ *is a multi-index.*

Proof. Starting from the recursive definition of $\Phi_k = \sum_{j=1}^{p\wedge k}\Phi_{k-j}\boldsymbol{A}_j$,

$$|\phi_{k,\delta}|_1 = |e_\delta'\Phi_k|_1 = \left|\sum_{j=1}^{p\wedge k} e_\delta'\Phi_{k-j}\boldsymbol{A}_j\right|_1 \leq \sum_{j=1}^{p\wedge k}|\phi_{k-j,\delta}\boldsymbol{A}_j|_1 \leq \sum_{j=1}^{p\wedge k}|\phi_{k-j,\delta}|_1 |||\boldsymbol{A}_j|||_\infty.$$

Suppose $k \geq p$, let $\alpha_j = e^{\rho}|||\boldsymbol{A}_j|||_\infty$ and verify that $0 \leq \sum_{j=1}^{p}\alpha_j \leq 1$. Iterating on the previous argument $s \leq k-p$ times yields

$$\begin{aligned}
|\phi_{k,\delta}|_1 &\leq \sum_{j_1=1}^{p}\cdots\sum_{j_s=1}^{p}\left(\prod_{l=1}^{s}|\boldsymbol{A}_{j_l}|_\infty\right)|\phi_{k-\sum_{l=1}^{s}j_l,\delta}|_1\\
&= e^{-s\rho}\sum_{j_1=1}^{p}\cdots\sum_{j_s=1}^{p}\left(\prod_{l=1}^{s}\alpha_{j_l}\right)|\phi_{k-\sum_{l=1}^{s}j_l,\delta}|_1\\
&= \cdots\\
&= e^{-\rho(k-p)}\sum_{j=1}^{p}\left(\sum_{\boldsymbol{i}:|\boldsymbol{i}|_1=k-p+j}\prod_{l=1}^{k-p}\alpha_{i_l}\right)|\phi_{p-j,\delta}|_1,
\end{aligned}$$

where $\boldsymbol{i} = (i_1,\ldots,i_{k-p})$ is a multi-index and the summation is over all combinations satisfying $|\boldsymbol{i}|_1 = k-p+j$. The term inside parentheses is $\tilde{\alpha}_j$ and under the conditions of the lemma

$$|\phi_{k,\delta}|_1 \leq e^{-\rho k}\times\left[e^{\rho p}\sum_{j=1}^{p}\tilde{\alpha}_j|\phi_{p-j},\delta|_1\right] \leq (1-e^{-\rho})c_0\, e^{-\rho k}.$$

The same result follows trivially for $k < p$ under the assumptions of the lemma.

Summing over all values of $k \geq m$,

$$\sum_{k=m}^{\infty}|\phi_{k,\delta}|_1 \leq c_0(1-e^{-\rho})\sum_{k=m}^{\infty}e^{-\rho k} = c_0 e^{-m\rho}\frac{\sum_{k=0}^{\infty}e^{-\rho k}}{(1-e^{-\rho})^{-1}} = c_0 e^{-m\rho}.$$

□

B.2. **Concentration inequality for martingales.** In this section we derive concentration bounds for martingales. In the firs theorem we consider martingales with at most d finite moments, whereas in the second we allow the tails of the marginal distributions to decrease at a sub-Weibull, sub-exponential or, even sub- and super-Gaussian rate.

Lemma 5 (Concentration bounds for high dimensional martingales). *Let* $\{\xi_t\}_{t=1,\ldots,T}$ *denote a multivariate martingale difference process with respect to the filtration* $\mathcal{F}_t$ *taking values on* $\mathbb{R}^n$ *and assume* $\mathbb{E}(\xi_{it}^2)$ *is finite for all* $1 \leq i \leq t$ *and* $1 \leq t \leq T$*. Then,*

$$\Pr\left(\left|\sum_{t=1}^{T}\xi_t\right|_\infty > Tx\right) \leq 2n\exp\left(-\frac{Tx^2}{2M^2+xM}\right) + 4\Pr\left(\max_{1\leq t\leq T}|\xi_t|_\infty > M\right),$$

for all $M > 0$*.*

Proof. Write $\xi_t = (\xi_{1t}, ..., \xi_{nt})'$. The proof follows after application of (Fan et al., 2012a, Corollary 2.3).

Write $V_k^2(M) = \max_{1\leq i\leq n}\sum_{t=1}^{k}\mathbb{E}[\xi_{it}^2 I(\xi_{it} < M)|\mathcal{F}_t]$, $X_{ik} = \sum_{t=1}^{k}\xi_{it}$ and $X'_{ik}(M) = \sum_{t=1}^{k}\xi_{it}I(\xi_{it} \leq M)$. It follows that for $v > 0$ and $x > 0$,

$$\begin{aligned}
\Pr(|X_n|_\infty > x) &\leq \Pr(\exists i,k:\ X_{ik} > x \cap V_k^2(M) \leq v^2) + \Pr(V_T^2(M) > v^2) \\
&\leq \Pr(\exists i,k:\ X'_{ik}(M) > x \cap V_k^2(M) \leq v^2) + \Pr(V_T^2(M) > v^2) \\
&\quad + \Pr\left(\max_{1\leq i\leq n}\sum_{t=1}^{k}\xi_{it}I(\xi_{it} > M) > 0\right) \\
&\overset{(1)}{\leq} n\exp\left(-\frac{(Tx/M)^2}{2((v/M)^2 + \frac{T}{3}x/M)}\right) + \Pr(V_n^2(M) > v^2) \\
&\quad + \Pr\left(\max_{1\leq t\leq T}|\xi_t|_\infty > M\right) \\
&\overset{(2)}{\leq} n\exp\left(-\frac{Tx^2}{2M^2+Mx}\right) + 2\Pr\left(\max_{1\leq t\leq T}|\xi_t|_\infty > M\right).
\end{aligned}$$

In (1) we use union bound and (Fan et al., 2012a, Theorem 2.1) and in (2) we set $v^2 = T(M^2 + \frac{1}{6T}Mx)$ and the following:

$$\begin{aligned}\Pr(V_T^2(M) > v^2) &\leq \Pr\left(\max_{1\leq i\leq n}\sum_{t=1}^{T}\mathbb{E}[\xi_{it}^2 I(|\xi_{it}| \leq M)|\mathcal{F}_t] \geq v^2\right)\\ &\quad + \Pr\left(\max_{1\leq i\leq n}\sum_{t=1}^{T}\mathbb{E}[\xi_{it}^2 I(\xi_{it} < -M)|\mathcal{F}_t] > 0\right)\\ &\leq \Pr\left(\max_{1\leq i\leq n}\sum_{t=1}^{T}\mathbb{E}[\xi_{it}^2 I(|\xi_{it}| \leq M)|\mathcal{F}_t] \geq T(M^2 + \frac{1}{6T}Mx)\right)\\ &\quad + \Pr\left(\max_{1\leq t\leq T}|\xi_t|_\infty > M\right)\\ &\leq \Pr\left(\max_{1\leq t\leq T}|\xi_t|_\infty > M\right),\end{aligned}$$

where in the last line we note that $\sum_{t=1}^{T}\mathbb{E}[\xi_t^2 I(|\xi_t| \leq M)|\mathcal{F}_t] \leq TM^2$.

Finally, write $\Pr(|X_n| \geq Tx) = \Pr(X_n \geq Tx) + \Pr(-X_n \geq Tx)$ and apply above development in both terms. □

Corollary 1. *Let* $\{\xi_t = (\xi_{1t}, ..., \xi_{nt})'\}_{t\geq 1}$ *denote a multivariate martingale difference process with respect to the filtration* $\mathcal{F}_t$ *taking values on* $\mathbb{R}^n$*. Suppose that for each* $\max_{i,t}\Pr(|\xi_{it}| > x) \leq 2e^{-(x/\tau)^\alpha}$*, for all* $x > 0$*, some* $\alpha > 0$ *and* $\tau > 0$ *Then,*

$$\Pr\left(\left|\sum_{t=1}^{T}\xi_t\right| > Tx\right) \leq 2n\exp\left(-\frac{Tx^2}{2M^2 + xM}\right) + 8nT\exp\left(-\frac{M^\alpha}{\tau^\alpha}\right).$$

In particular, if $x > \tau(\epsilon + \log(nT))^{1/\alpha}\sqrt{\epsilon + \log n}/\sqrt{T}$ *and* $T > (\epsilon + \log n)$ *for any* $\epsilon > 0$*,*

$$\Pr\left(\left|\sum_{t=1}^{T}\xi_t\right| > Tx\right) \leq 10e^{-\epsilon}.$$

Proof. The first part we combine the union bound with assumption on ξ_{it}. In the second part, we will need the following bound. Let $0 < a < b/4 < \infty$. then $\sqrt{a+b} - \sqrt{a} \geq \sqrt{b}(1 - 2\sqrt{a/b})^{1/2}$. To verify that, first note that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, then $(\sqrt{a+b} - \sqrt{a})^2 = 2a + b - 2\sqrt{a^2 + ab} \geq b - 2\sqrt{ab} = b(1 - 2\sqrt{a/b})$. Now, let $a = 1/T$ and $b = 8/(\epsilon + \log n)$ and verify that the choice $M = x\sqrt{T}/\sqrt{\epsilon + \log n}$ satisfy $\log n - \frac{Tx^2}{2M^2+Mx} < -\epsilon$, then replace M and x to obtain the bound. □

B.3. **Concentration bound for empirical covariance matrix.** In this section we derive concentration bound for $\|\mathbf{\Phi}_T - \Gamma\|_{\max}$, where $\mathbf{\Phi}_T = \mathbf{X}'\mathbf{X}/T$ and $\Gamma = \mathbb{E}\mathbf{\Phi}_T$. We first split the problem into a sum of martingales and a tail dependence term. Then, we bound both individually.

Lemma 6. *Suppose Assumptions (A1), (A2) and (A3) hold and*

$$p < \frac{T^{\gamma_1 \wedge \gamma_2}}{(\frac{2}{\gamma_1 \wedge \gamma_2 + 1})(2 + \frac{1.4}{2\gamma_1 c_\phi \wedge a_2})}.$$

If for some $\xi > 0$

$$\epsilon^2 \geq \frac{2(1+\xi)^{1+2/\alpha} \tau^2 [\log(npT)]^{1+2/\alpha}}{T},$$

then

$$\begin{aligned} \Pr\left(\|\mathbf{\Gamma}_T - \mathbf{\Gamma}\|_{\max} \geq \epsilon\right) \leq & \frac{2}{(np)^\xi T^{1+\xi}} + \frac{8}{(np)^\xi T^\xi} \\ & + \frac{n^2}{\epsilon}\left(b_1 e^{-c_\phi \wedge a_2 (T/2)^{\gamma_2 \wedge \gamma_1}} + b_8 e^{-2\gamma_1 c_\phi (T/2)^{\gamma_1}}\right) \end{aligned} \tag{B.2}$$

where b_1*,* b_5 *and* τ *are constants not depending on* T*.*

Proof. Use the union bound to rewrite our probability bound in terms of $\mathbf{y}_{t-s}$:

$$\Pr(\|\mathbf{\Phi}_T - \Gamma\|_{\max} > \epsilon) \leq 2\sum_{r=0}^{p}\sum_{s=0}^{p-r} \Pr\left(\left\|\sum_{t=p+1}^{T} \mathbf{y}_{t-r}\mathbf{y}'_{t-r-s} - \mathbb{E}[\mathbf{y}_{t-r}\mathbf{y}'_{t-r-s}]\right\|_{\max} > T\epsilon\right).$$

Now, use a telescopic expansion of $\mathbf{y}_t\mathbf{y}'_{t-s}$ to obtain a sum of martingales and a dependence term:

$$\begin{aligned} \sum_{t=p+1}^{T} \mathbf{y}_t\mathbf{y}'_{t-s} - \mathbb{E}[\mathbf{y}_t\mathbf{y}'_{t-s}] = & \underbrace{\sum_{t=p+1}^{T}\sum_{l=1}^{m} \mathbb{E}[\mathbf{y}_t\mathbf{y}'_{t-s}|\mathcal{F}_{t-l+1}] - \mathbb{E}[\mathbf{y}_t\mathbf{y}'_{t-s}|\mathcal{F}_{t-l}]}_{I_1} \\ & + \underbrace{\sum_{t=p+1}^{T} \mathbb{E}[\mathbf{y}_t\mathbf{y}'_{t-s}|\mathcal{F}_{t-m}] - \mathbb{E}[\mathbf{y}_t\mathbf{y}'_{t-s}]}_{I_2} \\ = & I_1 + I_2. \end{aligned}$$

Here,

$$I_1 = \sum_{l=1}^{m}\sum_{t=p+1}^{T} V_{l,t}^{(s)} \quad \text{and} \quad I_2 = \sum_{t=p+1}^{T} \mathbb{E}[\mathbf{y}_t\mathbf{y}'_{t-s}|\mathcal{F}_{t-m}] - \mathbb{E}[\mathbf{y}_t\mathbf{y}'_{t-s}]$$

where $\{V_{l,t}^{(s)}\}_t$, $l = 1, ..., m$, are sequences of martingale differences. The same decomposition holds for all terms $\mathbf{y}_{t-s}\mathbf{y}_{t-r-s}$. Then,

$$
\begin{aligned}
\Pr(\|I_1 + I_2\|_{\max} > T\epsilon) &\leq \sum_{l=1}^{m} \Pr\left(\| \sum_{t=p+1}^{T} V_{l,t}^{(s)} \|_{\max} > \frac{T\epsilon}{2m} \right) \\
&+ \Pr\left(\left\| \sum_{t=p+1}^{T} \mathbb{E}[\mathbf{y}_t\mathbf{y}'_{t-s}|\mathcal{F}_{t-m}] - \mathbb{E}[\mathbf{y}_t\mathbf{y}'_{t-s}] \right\|_{\max} > \frac{T\epsilon}{2} \right) \\
&\leq 2mn^2 \exp\left(-\frac{T\epsilon^2}{2M^2 + M\epsilon} \right) \\
&+ 4mn^2 T \max_{l,t} \Pr\left(|V_{l,t}^{(s)}| > M \right) \quad \text{(B.3)} \\
&+ \frac{2}{T\epsilon}\mathbb{E}\left| \max_{1\leq i,j<n} \mathbb{E}| \sum_{t=p+1}^{T} \mathbb{E}[e'_i\mathbf{y}_t\mathbf{y}_{t-s}e_j|\mathcal{F}_{t-m}] - e'_i\mathbb{E}[\mathbf{y}'_t\mathbf{y}_{t-s}]e_j \right|, \quad \text{(B.4)}
\end{aligned}
$$

where $e_i = (0, ..., 0, 1, 0, ..., 0)'$ is the i^{th} canonical basis vector in $\mathbb{R}^n$.

Bounding the tail (B.3):

The martingale differences $\{V_{t,l}^{(s)}\}$ ($l = 1, .., n$ and $s = 0, ..., p$) are sub-Weibull with parameter $\alpha/2$. For any random variables (X, Y) and σ-algebras $\mathcal{F}$ and $\mathcal{G}$,

$$\|\mathbb{E}[XY|\mathcal{F}] - \mathbb{E}[XY|\mathcal{G}]\|_p \leq 2\|XY\|_p \leq 2\|Y^2\|_p^{1/2}\|X^2\|_p^{1/2}.$$

Therefore, it follows from Wong et al. (2020, Lemmas 5 and 6) that if both X and Y are sub-Weibull with parameter α, then XY is sub-Weibull with parameter $\alpha/2$. Therefore, there exists some τ^* such that $\Pr(|V_{t,l}^{(s)}| > s) \leq 2\exp(-|x/\tau^*|^{\alpha/2})$, bounding (B.3).

Bounding covariances (B.4):

Now we move toward bounding the dependence term (B.4).Write

$$
\begin{aligned}
\mathbf{y}_t\mathbf{y}'_{t-s} &= \sum_{j=0}^{s-1} \Phi_j \mathbf{u}_{t-j} \sum_{j=0}^{\infty} \mathbf{u}'_{t-s-j}\Phi'_j + \sum_{j=0}^{\infty} \Phi_{s+j}\mathbf{u}_{t-s-j} \sum_{j=0}^{\infty} \mathbf{u}'_{t-s-j}\Phi'_j \\
&= \sum_{j=0}^{s-1} \Phi_j \mathbf{u}_{t-j} \sum_{j=0}^{\infty} \mathbf{u}'_{t-s-j}\Phi'_j \\
&+ \sum_{j=0}^{\infty} \Phi_{j+s}\mathbf{u}_{t-s-j}\mathbf{u}_{t-s-j}\Phi_j \\
&+ \sum_{k=1}^{\infty}\sum_{j=0}^{\infty} \Phi_{j+s}\mathbf{u}_{t-s-j}\mathbf{u}_{t-s-j-k}\Phi_{j+k} \\
&+ \sum_{k=1}^{\infty}\sum_{j=0}^{\infty} \Phi_{j+s+k}\mathbf{u}_{t-s-j-k}\mathbf{u}_{t-s-j}\Phi_j.
\end{aligned}
$$

It follows that $\mathbb{E}[\mathbf{y}_t\mathbf{y}'_{t-s}] = \sum_{j=0}^{\infty}\Phi_{j+s}\mathbf{\Sigma}\Phi_j$. Recall that $\mathcal{F}_{t-m} = \sigma\langle \mathbf{u}_{t-i} : i = m, m+1, ...\rangle$, then, for $m > s$,

$$\begin{aligned}
\mathbb{E}[\mathbf{y}_t\mathbf{y}_{t-s}|\mathcal{F}_{t-m}] - \mathbb{E}[\mathbf{y}_t\mathbf{y}'_{t-s}] = &\sum_{j=0}^{m-s-1}\Phi_j\mathbb{E}[\mathbf{u}_{t-s-j}\mathbf{u}'_{t-s-j} - \mathbf{\Sigma}|\mathcal{F}_{t-m}]\Phi'_{j+s} \\
&+ \sum_{j=0}^{\infty}\Phi_{m+j}(\mathbf{u}_{t-m-j}\mathbf{u}_{t-m-j} - \mathbf{\Sigma})\Phi_{m-s+j} \\
&+ \sum_{k=1}^{\infty}\sum_{j=0}^{\infty}\Phi_{m+j}\mathbf{u}_{t-m-j}\mathbf{u}'_{t-m-j-k}\Phi_{m-s+j+k} \\
&+ \sum_{k=1}^{\infty}\sum_{j=0}^{\infty}\Phi_{m+j+k}\mathbf{u}_{t-m-j-k}\mathbf{u}'_{t-m-j}\Phi_{m-s+j} \\
= &\ A_1(t,s,m) + A_2(t,s,m) + A_3(t,s,m) + A_4(t,s,m).
\end{aligned} \tag{B.5}$$

We shall bound $\mathbb{E}\left|\sum_{r=0}^{p}\sum_{s=0}^{p-r}e'_kA_i(t-r,s,m)e_l\right|$ individually, for all $\{e_i, i = 1, ..., n\}$ the canonical basis vector in $\mathbb{R}^n$.

a) Bounding $\mathbb{E}\left|\sum_{r=0}^{p}\sum_{s=0}^{p-r}A_1(t-r,s,m)\right|$**:**

It follows from Assumption (A2) that for all $\mathbf{b}_1, \mathbf{b}_2 \in \{\mathbf{b} \in \mathbb{R}^n : |\mathbf{b}|_1 = 1\}$

$$\max_t \mathbb{E}\left|\mathbb{E}\left[\mathbf{b}'_1(\mathbf{u}_t\mathbf{u}'_t - \mathbf{\Sigma})'\mathbf{b}_2|\mathcal{F}_{t-m}\right]\right| \le a_1\exp(-a_2m).$$

Set $\{e_i, i = 1, ..., n\}$ the canonical basis vector in $\mathbb{R}^n$. It follows from Assumptions (A1) - (A2) that for $j \le m-s-1$:

$$\begin{aligned}
&\mathbb{E}|e'_k\Phi_j\mathbb{E}[(\mathbf{u}_{t-s-j}\mathbf{u}'_{t-s-j} - \mathbf{\Sigma})|\mathcal{F}_{t-m}]\Phi'_{j+s}e_l| \\
&\qquad \le |\phi_{j,k}|_1||\phi_{j+s,l}|_1 \max_t \mathbb{E}\mathbb{E}[\mathbf{b}'_1(\mathbf{u}_{t-s-j}\mathbf{u}'_{t-s-j} - \mathbf{\Sigma})\mathbf{b}_2|\mathcal{F}_{t-m}]| \\
&\qquad \le \bar{c}_\Phi e^{-c_\phi(j+s)^{\gamma_1}}e^{-2c_\phi j^{\gamma_1}}\left[a_1e^{-a_2(m-j-s)^{\gamma_2}}\right]
\end{aligned}$$

Let $0 < \gamma \le 1$ and $\frac{m^\gamma}{p} > (\frac{2}{\gamma+1})(2 + \frac{1.4}{c})$ then $c(m-p)^\gamma - \log(p+1) \ge c(m/2)^\gamma$. Rewriting the inequality, we have to show that $(m/p-1)^\gamma - (m/2p)^\gamma > 2\log(p+1)/cp^\gamma$. In the LHS, a second order Taylor series expansion yields $(2a-1)^\gamma - a^\gamma \ge \gamma\frac{a-1}{a}a^\gamma(1 - \frac{1-\gamma}{2}\frac{a-1}{a}) \ge \frac{\gamma+1}{2}\frac{a-1}{a}a^\gamma$, for $a > 1$. Set $a = m/2p$, so that $\frac{\gamma+1}{2}(\frac{m-2p}{m})(\frac{m}{2p})^\gamma \ge \log(p+1)/cp^\gamma$. As for the RHS, $\log(p+1)/p < 0.7$. Combining bounds above we show our claim.

Note that for $a, b \ge 0$ and $0 < \gamma \le 1$ $a^\gamma + b^\gamma = (a+b)^\gamma[x^\gamma + (1-x)^\gamma]$ where $x = a/(a+b)$, and $1 \le [x^\gamma + (1-x)^\gamma] \le 2$ for $x \in [0,1]$. Now, let $0 < \gamma \le 1$ and $c > 0$, then $\sum_{j=0}^{n}e^{-cj^\gamma} \le \int_0^n e^{-cx^\gamma}dx = c^{1/\gamma}/\gamma\Gamma(1/\gamma, n) \uparrow c^{1/\gamma}\Gamma(1/\gamma+1) < \infty$, as $n \to \infty$, where $\Gamma(a,n) = \int_0^n x^{a-1}e^{-x}dx \uparrow \int_0^\infty x^{a-1}e^{-x}dx = \Gamma(a)$ are the incomplete gamma function and gamma function, respectively.

Then,

$$\begin{aligned}
\mathbb{E}\left|\sum_{r=0}^{p}\sum_{s=0}^{p-r} e_k' A_1(t-r,s,m)e_l\right| &= \sum_{r=0}^{p}\sum_{s=0}^{p-r}\sum_{j=0}^{m-r-s-1} \mathbb{E}\left|e_k'\Phi_j \mathbb{E}[\mathbf{u}_{t-r-s-j}\mathbf{u}_{t-r-s-j}' - \mathbf{\Sigma}|\mathcal{F}_{t-m}]\Phi_{j+s}' e_l\right| \\
&\le \bar{c}_\Phi a_1 \sum_{r=0}^{p}\sum_{s=0}^{p-r}\sum_{j=0}^{m-r-s-1} e^{-c_\phi (j+s)^{\gamma_1}} e^{-2c_\phi j^{\gamma_1}} e^{-a_2(m-r-j-s)^{\gamma_2}} \\
&\le \bar{c}_\Phi a_1 \sum_{r=0}^{p} e^{-(c_\phi\wedge a_2)(m-r)^{\gamma_2\wedge\gamma_1}} \sum_{s=0}^{p-r}\sum_{j=0}^{m-r-s-1} e^{-c_\phi j^{\gamma_1}} \\
&\le \bar{c}_\Phi a_1 \frac{c_\phi^{1/\gamma_1}\Gamma(1/\gamma_1)}{2\gamma_1}(p+1)^2 e^{-(c_\phi\wedge a_2)(m-p)^{\gamma_2\wedge\gamma_1}} \\
&\le b_1 e^{-(c_\phi\wedge a_2)(m/2)^{\gamma_2\wedge\gamma_1}}
\end{aligned}$$

b) Bounding $\mathbb{E}\left|\sum_{r=0}^{p}\sum_{s=0}^{p-r} e_k' A_2(t-r,s,m)e_l\right|$**:**

Let $\max_{\mathbf{b}_1,\mathbf{b}_2,t}\mathbb{E}\left|\mathbf{b}_1'(\mathbf{u}_t\mathbf{u}_t' - \mathbf{\Sigma})\mathbf{b}_2\right| \le 2\Lambda_{\max}(\mathbf{\Sigma})$ where $\mathbf{b}_1,\mathbf{b}_2 \in \{\mathbf{b}\in\mathbb{R}^n : |\mathbf{b}|_1 = 1\}$. It follows from Lemma 7 after rearranging terms:

$$\begin{aligned}
\mathbb{E}\left|\sum_{r=0}^{p}\sum_{s=0}^{p-r} e_k' A_2(t-r,s,m)e_l\right| &\le \sum_{r=0}^{p}\sum_{s=0}^{p-r}\sum_{j=0}^{\infty}\mathbb{E}\left|e_k'\Phi_{m+j}(\mathbf{u}_{t-r-m-j}\mathbf{u}_{t-r-m-j}' - \mathbf{\Sigma})\Phi_{m-s+j}' e_l\right| \\
&\le \sum_{r=0}^{p}\sum_{s=0}^{p-r}\sum_{j=m}^{\infty} |\phi_{j,k}|_1 |\phi_{j-s,l}|_1 \max_{\mathbf{b}_1,\mathbf{b}_2,t}\mathbb{E}\left|\mathbf{b}_1'(\mathbf{u}_t\mathbf{u}_t' - \mathbf{\Sigma})\mathbf{b}_2\right| \\
&\le 2\Lambda_{\max}(\mathbf{\Sigma})\bar{c}_\Phi^2 \sum_{r=0}^{p}\sum_{s=0}^{p-r}\sum_{j=m}^{\infty} e^{-c_\phi j^{\gamma_1}} e^{-c_\phi (j-r)^{\gamma_1}} \\
&= 2b_2\left(\sum_{r=0}^{p}\sum_{s=0}^{p-r} 1\right)\sum_{j=m-p}^{\infty} e^{-2c_\phi (j-r)^{\gamma_1}} \\
&= b_3\left[(p-1)(m-p)^{(1-\gamma_1)/2} e^{-c_\phi (m-p)^{\gamma_1}}\right]^2 \\
&\le b_3 (e\gamma_1)^{-\frac{1-\gamma_1}{\gamma_1}} e^{-2\gamma_1 c_\phi (m/2)^{\gamma_1}},
\end{aligned}$$

were $b_3 = \frac{\sqrt{2}b_2}{c_\phi\gamma_1(1+\gamma_1)^2}$. In the last line, we use $\frac{m^{\gamma_1}}{p} > (\frac{2}{\gamma_1+1})(2+\frac{1.4}{2\gamma_1 c_\phi})$ for m sufficiently large:

$$\begin{aligned}
&(p-1)(m-p)^{(1-\gamma_1)/2} e^{-c_\phi (m-p)^{\gamma_1}} \\
&\qquad \le e^{-\frac{1-\gamma_1}{2}(c_\phi (m-p)^{\gamma_1} - \log(m-p))} \times e^{-\gamma_1 c_\phi (m/2)^{\gamma_1} - \log(p+1)}(p+1) \\
&\qquad \le (e\gamma_1)^{-\frac{1-\gamma_1}{2\gamma_1}} e^{-\gamma_1 c_\phi (m/2)^{\gamma_1}}
\end{aligned}$$

c) Bounding $\sum_{r=0}^{p}\sum_{s=0}^{p-r} A_j(t-r,s,m)$ **(**$j = 3,4$**):**

Under Assumption (A1)-(A3), for all $\mathbf{b} \in \mathbb{R}^n$ with $|\mathbf{b}|_1 = 1$,

$$\begin{aligned}\mathbb{E}|e_r'\Phi_{m+j}\mathbf{u}_{t-m-j}\mathbf{u}_{t-m-j-k}'\Phi_{m-s+j+k}e_s| \\ \leq |\phi_{m+j,r}|_1|\phi_{m-s+j+k,s}|_1 \max_t \|\mathbf{b}'\mathbf{u}_t\|_2^2 \\ \leq b_2 e^{-c_\phi(m+j)^{\gamma_1}-c_\phi(m-s+j+k)^{\gamma_1}},\end{aligned}$$

where $b_2 = \bar{c}_\Phi^2\Lambda_{\max}(\mathbf{\Sigma})$. As before, if we have $\mathbf{u}_{t-r-s-j}$ we must replace m my $m-r$. It follows from Lemma 7, $\frac{m^{\gamma_1}}{p} > (\frac{2}{\gamma_1+1})(2+\frac{1.4}{2\gamma_1 c_\phi})$ and m sufficiently large:

$$\begin{aligned}&\sum_{r=0}^{p}\sum_{j=0}^{\infty}e^{-c_\phi((m-r+j)^{\gamma_1}}\sum_{s=0}^{p-r}\sum_{k=0}^{\infty}e^{-c_\phi(m-r+j+k-s)^{\gamma_1}} \\ &\leq \left(\sum_{r=0}^{p}\sum_{s=0}^{p-r}1\right)\sum_{j=0}^{\infty}e^{-c_\phi((m-p+j)^{\gamma_1}}\sum_{k=0}^{\infty}e^{-c_\phi(m-p+j+k)^{\gamma_1}} \\ &\leq b_4\left((p+1)(m-p)^{1-\gamma_1}e^{-c_\phi(m-p)^{\gamma_1}}\right)^2 \\ &\leq b_4\left(e^{-(1-\gamma_1)(c_\phi(m-p)^{\gamma_1}-\log(m-p))}\times e^{-\gamma_1 c_\phi(m/2)^{\gamma_1}-\log(p+1)}(p+1)\right)^2 \\ &\leq b_5 e^{-2\gamma_1 c_\phi(m/2)^{\gamma_1}},\end{aligned}$$

where $b_4 = \frac{3}{(1+\gamma_1)^3 c_\phi^2\gamma_1^2}$, $b_5 = b_4(e\gamma_1)^{-2\frac{1-\gamma_1}{\gamma_1}}$.

Then, it follows that

$$\sum_{r=0}^{p}\sum_{s=0}^{p-r}\sum_{k=1}^{\infty}\sum_{j=0}^{\infty}\mathbb{E}|e_i'\Phi_{m+j}\mathbf{u}_{t-m-j}\mathbf{u}_{t-m-j-k}'\Phi_{m-s+j+k}e_l| \leq b_6 e^{-2\gamma_1 c_\phi(m/2)^{\gamma_1}}. \tag{B.6}$$

where $b_6 = b_2 b_5 = \frac{3\bar{c}_\Phi^2\Lambda_{\max}(\mathbf{\Sigma})(e\gamma_1)^{-2\frac{1-\gamma_1}{\gamma_1}}}{(1+\gamma_2)^3 c_\phi^2\gamma_1^2}$.

d) Combining bounds:

Finally combining the three bounds above and setting m satisfying $\frac{m^{\gamma_1\wedge\gamma_2}}{p} > (\frac{2}{\gamma_1\wedge\gamma_2+1})(2+\frac{1.4}{2\gamma_1 c_\phi\wedge a_2})$:

$$\begin{aligned}&\sum_{r=0}^{p}\sum_{s=0}^{p-r}\mathbb{E}\left|\max_{1\leq i,j<n}\frac{1}{T}\mathbb{E}|\sum_{t=p+1}^{T}\mathbb{E}[e_i'\mathbf{y}_{t-r}\mathbf{y}_{t-r-s}'e_j|\mathcal{F}_{t-m}] - \mathbb{E}[e_i'\mathbf{y}_{t-r}\mathbf{y}_{t-r-s}'e_j]\right| \\ &\qquad\leq n^2\left(b_1 e^{-(c_\phi\wedge a_2)(m/2)^{\gamma_2\wedge\gamma_1}} + b_6 e^{-2\gamma_1 c_\phi(m/2)^{\gamma_1}}\right),\end{aligned} \tag{B.7}$$

where $b_7 = b_3(e\gamma_1)^{-\frac{1-\gamma_1}{\gamma_1}} + 2b_6$

Combining Tail and Covariance:

Set $m = T$, then we require that

$$p < \frac{T^{\gamma_1\wedge\gamma_2}}{(\frac{2}{\gamma_1\wedge\gamma_2+1})(2+\frac{1.4}{2\gamma_1 c_\phi\wedge a_2})}.$$

Then, combining bounds:

$$\begin{aligned}\Pr\left(\|\boldsymbol{\Gamma}_T - \boldsymbol{\Gamma}\|_{\max} \geq \epsilon\right) \leq\ & 2\frac{1}{T}\exp\left(2\log(npT) - \frac{T\epsilon^2}{2M^2 + M\epsilon}\right) \\ & + 8e^{2\log(npT) - M^\alpha/\tau^\alpha} \\ & + \frac{n^2}{\epsilon}\left(b_1 e^{-a_2\wedge c_\phi (T/2)^{\gamma_2\wedge\gamma_1}} + b_7 e^{-2\gamma_1 c_\phi (T/2)^{\gamma_1}}\right),\end{aligned}$$

where τ, b_1 and b_5 are as above. Let $\xi > 0$ and $M^\alpha = (2+\xi)\tau^\alpha \log(npT)$, by assumption $T\epsilon^2 \geq 2(1+\xi)^{1+2/\alpha}\tau^2[\log(npT)]^{1+2/\alpha}$ which implies that $2\log(npT) - \frac{T\epsilon^2}{2M^2+M\epsilon} \leq -\xi\log(npT)$. Finally,

$$\Pr\left(\|\boldsymbol{\Gamma}_T - \boldsymbol{\Gamma}\|_{\max} \geq \epsilon\right) \leq \frac{2}{(np)^\xi T^{1+\xi}} + \frac{8}{(np)^\xi T^\xi} + \frac{n^2}{\epsilon}\left(b_1 e^{-a_2\wedge c_\phi (T/2)^{\gamma_2}} + b_8 e^{-2\gamma_1 c_\phi (T/2)^{\gamma_1}}\right).$$

□

Lemma 7. *Let* $0 < a \leq 1$ $b > 0$, $n \geq \left(\frac{ba(1+a)}{\sqrt{2}-1}\right)^{1/(1-a)}$. *Then*

$$\sum_{i=n}^{\infty}\sum_{j=0}^{\infty} e^{-bi^a - b(i+j)^a} \leq \frac{6}{(1+a)^3 b^2 a^2}\left[n^{(1-a)}e^{-bn^a}\right]^2, \tag{B.8}$$

and

$$\sum_{i=n}^{\infty} e^{-bi^a} \leq \frac{\sqrt{2}}{ba(1+a)^2} n^{1-a} e^{-bn^a}. \tag{B.9}$$

Proof. Let V denote a Weibull(a,$2b$) random variable.

$$\begin{aligned}\sum_{i=n}^{\infty}\sum_{j=0}^{\infty} e^{-bi^\alpha - b(i+j)^\alpha} &= \sum_{j=n}^{\infty}(j-n+1)e^{-2bj^\alpha} \\ &= \sum_{j=n}^{\infty}(j-n+1)\mathbb{E}[I(V \geq j)] \\ &= \sum_{j=n}^{\infty}(j-n+1)^2\mathbb{E}[I(j \leq V < j+1)] \\ &\leq \mathbb{E}[(V-n+1)^2 I(V \geq n)].\end{aligned}$$

It follows from a second order Taylor expansion that for $x \in [0,1]$,

$$(1+x)^a - x^a \geq ax^{a-1} - \frac{a(1-a)}{2}x^{a-2} \geq \frac{a(1-a)}{2}x^{a-1},$$

and

$$(1+x)^a - 1 \geq ax - \frac{a(1-a)}{2}x^2 \geq \frac{a(1-a)}{2}x.$$

Then, for $v \geq n$, set $0 < x = n/v \leq 1$

$$(v+n)^a - n^a = v^a\left[(1+x)^a - x^a\right] \geq \frac{a(1-a)}{2}\left(\frac{v}{n}\right)^{1-a} v^a = \frac{a(1-a)}{2} n^{a-1} v,$$

and for $0 \leq v \leq n$, set $0 \leq x = v/n \leq 1$

$$(v+n)^a - n^a = n^a\left[(1+x)^a - 1\right] \geq \frac{a(1-a)}{2}\frac{v}{n} n^a = \frac{a(1-a)}{2} n^{a-1} v.$$

Therefore, $(v+n)^a - n^a > \frac{a(1-a)}{2} n^{a-1} v$ for $v \geq 0$. Also $(n+v)^{a-1} \leq n^{a-1}$. Now we may bound the conditional expected value. Let X denote an exponential random variables with parameter $\lambda = ba(1+a)n^{a-1}$:

$$\begin{aligned}
\mathbb{E}[(V-n+1)^2 | I(V \geq n)] &= \int_n^\infty (v-n+1)^2 \frac{2bav^{a-1}e^{-2bv^a}}{e^{-2bn^a}} dv \\
&= \int_0^\infty (x+1)^2 2ba(x+n)^{a-1} e^{-2b[(x+n)^a - n^a]} dx \\
&\leq \int_0^\infty (x+1)^2 ban^{a-1} e^{-ba(1+a)n^{a-1}x} dx \\
&= \frac{2}{1+a}\mathbb{E}(X+1)^2 \\
&= \frac{2}{1+a}\left(\frac{2+2\lambda+\lambda^2}{\lambda^2}\right) \\
&\leq \frac{6}{(1+a)^3 b^2 a^2} n^{2(1-a)},
\end{aligned}$$

where in the last line we note that $n \geq \left(\frac{ba(1+a)}{\sqrt{2}-1}\right)^{1/(1-a)}$ implies $\lambda \leq \sqrt{2} - 1$.

Finally, the first bound follows because

$$\begin{aligned}
\mathbb{E}[(V-n+1)^2 I(V \geq n)] &= \mathbb{E}[I(V \geq n)\mathbb{E}[(V-n+1)^2 | I(V \geq n)]] \\
&\leq \frac{6}{(1+a)^3 b^2 a^2} n^{2(1-a)} e^{-2bn^a}.
\end{aligned}$$

The second bound follows after similar arguments:

$$\begin{aligned}
\sum_{i=n}^\infty e^{-bi^a} &\leq \mathbb{E}[(n+1)I(V > n)] \\
&\leq \frac{1}{1+a}\mathbb{E}(X+1)e^{-bn^a} \\
&\leq \frac{\sqrt{2}}{ba(1+a)^2} n^{1-a} e^{-bn^a}.
\end{aligned}$$

□

References

Andrews, D. W. (1988). Laws of large numbers for dependent non-identically distributed random variables. *Econometric theory*, 4(3):458 – 467.

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

Bauer, G. and Vornik, K. (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, 160:93–101.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate garch models: a survey. *Journal of applied econometrics*, 21(1):79–109.

Boussama, F., Fuchs, F., and Stelzer, R. (2011). Stationarity and geometric ergodicity of bekk multivariate garch models. *Stochastic Processes and their Applications*, 121(10):2331–2360.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability surveys*, 2(2):107–144.

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18:17–39.

Chib, S., Omori, Y., and Asai, M. (2009). Multivariate stochastic volatility. In Mikosch, T., Kreiβ, J., Davis, R., and Andersen, T., editors, *Handbook of Financial Time Series*, volume 2, chapter 2, pages 71–162. Springer.

Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26:922–947.

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

Davis, R., Zang, P., and Zhang, T. (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics*, 25:1077—-1096.

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

Ensor, K., Raun, L., and Persse, D. (2013). A case-crossover analysis of out-of-hospital cardiac arrest and air pollution. *Circulation*, 127:1192–1199.

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

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

Hafner, C. M. and Preminger, A. (2009a). Asymptotic theory for a factor garch model. *Econometric Theory*, 25(02):336–363.

Hafner, C. M. and Preminger, A. (2009b). On asymptotic theory for multivariate garch models. *Journal of Multivariate Analysis*, 100(9):2044–2054.

Han, Y. and Tsay, R. S. (2019). High-dimensional linear regression for dependent data with applications to nowcasting. *Statistica Sinica*, forthcoming.

Hoek, G., Krishnan, R., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., and Kaufman, J. (2013). Long-term air pollution exposure and cardio-respiratory mortality: a review. *Environmental Health*, 12:43.

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

Kock, A., Medeiros, M., and Vasconcelos, G. (2020). Penalized regressions. In Fuleky, P., editor, *Macroeconomic Forecasting in the Era of Big Data: Theory and Practice*, volume 52 of *Advanced Studies in Theoretical and Applied Econometrics*, pages 193–228. Springer.

Loh, P.-L. and Wainwright, M. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of Statistics*, 40:1637–1664.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

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

Medeiros, M. and Mendes, E. (2016a). Adaptive LASSO estimation for ARDL models with GARCH innovations. *Econometric Reviews*. forthcoming.

Medeiros, M. and Mendes, E. (2016b). ℓ_1-regularization of high-dimensional time-series models with non-gaussian and heteroskedastic innovations. *Journal of Econometrics*, 191:255–271.

Merlevède, F., Peligrad, M., and Rio, E. (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151(3-4):435–474.

Miao, K., Phillips, P., and Su, L. (2019). High-dimensional vars with common factors. Working paper, Singapore Management University.

Miranda-Agrippino, S. and Ricco, G. (2019). Bayesian vector autoregressions: Estimation. *Oxford Research Encyclopedia of Economics and Finance*.

Negahban, S. N., Ravikumar, P., Wainwright, M. J., and Yu, B. (2012). A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Statistical Science*, 27(4):538–557.

Ramey, V. (2016). Macroeconomic shocks and their propagation. In *Handbook of Macroeconomics*, pages 71–162. Elsevier.

Schweinberger, M., Babkin, S., and Ensor, K. (2017). High-dimensional multivariate time series with additional structure. *Journal of Computational and Graphical Statistics*, 26:610–622.

Sims, C. (1980). Macroeconomics and reality. *Econometrica*, 48:1–48.

Song, S. and Bickel, P. J. (2011). Large vector auto regressions. *arXiv preprint arXiv:1106.3915*.

Van De Geer, S. A., Bühlmann, P., et al. (2009). On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics*, 3:1360–1392.

Wilson, G., Reale, M., and Haywood, J. (2015). *Developments in Multivariate Time Series Modeling.* CRC Press, Boca Raton.

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

São Paulo School of Economics, Getulio Vargas Foundation, São Paulo (FGV-SP)
Email address: ricardo.masini@fgv.br

Department of Economics, Pontifical Catholic University of Rio de Janeiro
Email address: mcm@econ.puc-rio.br

School of Applied Mathematics (EMap), Getulio Vargas Foundation, Rio de Janeiro (FGV-RJ)
Email address: eduardo.mendes@fgv.br