Inference for high-dimensional regressions with heteroskedasticity and autocorrelation

Andrii Babii* Eric Ghysels† Jonas Striaukas‡

May 29, 2020

Abstract

Time series regression analysis relies on the heteroskedasticity- and autocorrelation-consistent (HAC) estimation of the asymptotic variance to conduct proper inference. This paper develops such inferential methods for high-dimensional time series regressions. To recognize the time series data structures we focus on the sparse-group LASSO estimator. We establish the debiased central limit theorem for low dimensional groups of regression coefficients and study the HAC estimator of the long-run variance based on the sparse-group LASSO residuals. The treatment relies on a new Fuk-Nagaev inequality for a class of τ-dependent processes with heavier than Gaussian tails, which is of independent interest.

Keywords: HAC estimator, sparse-group LASSO, high-dimensional time series, inference for groups, Fuk-Nagaev inequality, τ-dependent sequences.

*Department of Economics, University of North Carolina–Chapel Hill - Gardner Hall, CB 3305 Chapel Hill, NC 27599-3305. Email: babii.andrii@gmail.com

†Department of Economics and Kenan-Flagler Business School, University of North Carolina–Chapel Hill. Email: eghysels@unc.edu.

‡LIDAM UC Louvain and FRS–FNRS Research Fellow. Email: jonas.striaukas@gmail.com.

1 Introduction

Modern time series analysis is increasingly faced with high-dimensional datasets sampled at different frequencies. Conventional time series are often supplemented with the non-traditional data, such as the high-dimensional data coming from the natural language processing. For instance, Bybee, Kelly, Manela, and Xiu (2020) extract 180 topic attention series from the over 800,000 daily *Wall Street Journal* news articles during 1984–2017 that have shown to be a useful supplement to more traditional macroeconomic and financial datasets for nowcasting in Babii, Ghysels, and Striaukas (2020).

The high-dimensional time series analysis in Babii, Ghysels, and Striaukas (2020) relies on the sparse-group LASSO (sg-LASSO) introduced in Simon, Friedman, Hastie, and Tibshirani (2013). The sg-LASSO allows capturing the group structures present in high-dimensional time series regressions where a single covariate with its lags constitute a group. An attractive feature of this estimator is that it encompasses the LASSO and the group LASSO as special cases, hence, improving upon the unstructured LASSO in the high-dimensional time-series setting. At the same time, the sg-LASSO can learn the distribution of time series lags in a data-driven way solving elegantly the problem first discussed in Fisher (1937).[1]

Time series analysis uses routinely the heteroskedasticity- and autocorrelation-consistent (HAC) estimation of the long-run variance; see Eicker (1963), Eicker (1967), Huber (1967), White (1980) and Gallant (1987), Newey and West (1987), Andrews (1991), among others.[2] Despite the increasing popularity of the LASSO in the empirical time series analysis, to the best of our knowledge, the HAC inference has not been formally studied in the relevant

[1]The distributed lag literature can be traced back to Fisher (1925); see also Almon (1965), Sims (1971a), Sims (1971b), Sims (1972), Shiller (1973), Haugh and Box (1977) as well as more recent mixed frequency data sampling (MIDAS) approach in Ghysels, Santa-Clara, and Valkanov (2006), Ghysels, Sinko, and Valkanov (2007), and Andreou, Ghysels, and Kourtellos (2013).

[2]For stationary time series, the HAC estimation of the long-run variance is the same problem as the estimation of the value of the spectral density at zero which itself has even longer history dating back to the smoothed periodogram estimators; see Daniell (1946), Bartlett (1948), and Parzen (1957).

literature.[3]

In this paper, we consider the HAC estimator for the sg-LASSO and study its formal statistical properties in the high-dimensional environment where the number of covariates can increase faster than the sample size. We obtain first the debiased central limit theorem with explicit bias correction under the realistic assumption that the regression errors are serially correlated, which to the best of our knowledge is new. Next, we establish formal statistical properties of the HAC estimator based on the sg-LASSO residuals. An important consequence of these results is that the optimal choice of the bandwidth parameter should scale appropriately with the number of covariates, the measure of the weak dependence, as well as the tail features of the data. This allows us conducting inference for the low-dimensional groups of coefficients, including the (mixed-frequency) Granger causality tests.

Our treatment allows for the data with heavier than exponential tails, since it is widely recognized that the economic and financial time series data are rarely sub-Gaussian. To that end, we establish a suitable version of the Fuk-Nagaev inequality, cf., Fuk and Nagaev (1971), for τ-*dependent* process with *polynomial tails*.[4]

The paper is organized as follows. We start with the large sample approximation to the distribution of the sg-LASSO estimator (and as a consequence of the LASSO and the group LASSO) with τ-dependent data in section 2. Next, we consider the HAC estimator of the asymptotic long-run variance based on the sg-LASSO residuals and study the inference for the low-dimensional groups of regression coefficients. In section 3, we establish a suitable version of the Fuk-Nagaev inequality for τ-dependent data. We report on a Monte Carlo study in section 4 which provides further insights about the validity of our theoretical anal-

[3]Previously, Chernozhukov, Härdle, Huang, and Wang (2019) consider the HAC estimator for the LASSO with the Bartlett kernel without establishing it formal properties; see also Feng, Giglio, and Xiu (2019) for an asset pricing application. Inference for the LASSO with i.i.d. data was developed in Belloni, Chernozhukov, and Hansen (2010), Belloni, Chernozhukov, and Hansen (2014), van de Geer, Bühlmann, Ritov, and Dezeure (2014), Zhang and Zhang (2014), Javanmard and Montanari (2014), and Zhang and Cheng (2017); see also Chiang and Sasaki (2019) for the inference with exchangeable arrays.

[4]The notion of τ-dependence coefficients was introduced in Dedecker and Prieur (2004) and Dedecker and Prieur (2005) as weaker than mixing coefficients allowing for a larger class of time series, such as the autoregressive processes with discrete innovations.

ysis in finite sample settings typically encountered in empirical applications. Conclusions appear in section 5. Proofs and supplementary results appear in the appendix and the supplementary material.

Notation: For a random variable $X \in \mathbf{R}$, let $\|X\|_q = (\mathbb{E}|X|^q)^{1/q}, q \geq 1$ be its L_q norm. For $p \in \mathbf{N}$, put $[p] = \{1, 2, \dots, p\}$. For a vector $\Delta \in \mathbf{R}^p$ and a subset $J \subset [p]$, let Δ_J be a vector in $\mathbf{R}^p$ with the same coordinates as Δ on J and zero coordinates on J^c. Let $\mathcal{G} = \{G_g : g \geq 1\}$ be a partition of $[p]$ defining groups. For a vector of regression coefficients $\beta \in \mathbf{R}^p$, the sparse-group structure is described by a pair $(S_0, \mathcal{G}_0)$, where $S_0 = \{j \in [p] : \beta_j \neq 0\}$ is the support of β and $\mathcal{G}_0 = \{G \in \mathcal{G} : \beta_G \neq 0\}$ is its group support. For $b \in \mathbf{R}^p$, its $\ell_q, q \geq 1$ norm is denoted $|b|_q = \left(\sum_{j\geq 1}^p |b_j|^q\right)^{1/q}, q < \infty$ and $|b|_\infty = \max_{1\leq j \leq p} |b_j|$. For $\mathbf{u}, \mathbf{v} \in \mathbf{R}^T$, the empirical inner product is defined as $\langle \mathbf{u}, \mathbf{v} \rangle_T = \frac{1}{T}\sum_{t=1}^T u_t v_t$ with the induced empirical norm $\|.\|_T^2 = \langle ., . \rangle_T = |.|_2^2/T$. For a symmetric $p \times p$ matrix A, let $\text{vech}(A) \in \mathbf{R}^{p(p+1)/2}$ be its vectorization consisting of the lower triangular and the diagonal part. Let A_G be a sub-matrix consisting of rows of A corresponding to indices in $G \subset [p]$. If $G = \{j\}$ for some $j \in [p]$, then we simply put $A_G = A_j$. For a $p \times p$ matrix A, put $\|A\|_\infty = \max_{j\in[p]} |A_j|_1$. For $a, b \in \mathbf{R}$, we put $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. Lastly, we write $a_n \lesssim b_n$ if there exists a (sufficiently large) absolute constant C such that $a_n \leq C b_n$ for all $n \geq 1$ and $a_n \sim b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$.

2 HAC inference for sg-LASSO

2.1 Debiased central limit theorem

Consider a generic dynamic linear regression

$$y_t = \mathbb{E}[y_t|\mathcal{F}_t] + u_t, \qquad \mathbb{E}[u_t|\mathcal{F}_t] = 0, \qquad t \in \mathbf{Z},$$

where $(y_t)_{t\in\mathbf{Z}}$ is a real-valued stochastic process and $(\mathcal{F}_t)_{t\in\mathbf{Z}}$ is some filtration. We approximate the conditional mean $\mathbb{E}[y_t|\mathcal{F}_t]$ with its best linear approximation with respect to the L_2 norm, denoted $X_t^\top \beta$, where $(X_t)_{t\in\mathbf{Z}}$ is a stochastic process in $\mathbf{R}^p$ that may include some covariates, lags of covariates up to a certain order, as well as lags of the dependent variable. For a sample of size T, in the vector notation

$$\mathbf{y} = \mathbf{m} + \mathbf{u},$$

where $\mathbf{y} = (y_1, \dots, y_T)^\top$, $\mathbf{m} = (\mathbb{E}[y_1|\mathcal{F}_1], \dots, \mathbb{E}[y_T|\mathcal{F}_T])^\top$, and $\mathbf{u} = (u_1, \dots, u_T)^\top$. The best linear approximation is denoted $\mathbf{X}\beta$, where $\mathbf{X}$ is $T \times p$ design matrix and $\beta \in \mathbf{R}^p$ is the unknown regression parameter. The linear approximation $\mathbf{X}\beta$ can be constructed from lagged values of y_t, some covariates, as well as lagged values of covariates measured at a higher frequency, in which case, we obtain the autoregressive distributed lag mixed frequency data sampling model (ARDL-MIDAS)

$$\phi(L) y_t = \sum_{k=1}^{K} \psi(L^{1/m}; \beta_k) x_{t,k} + u_t,$$

where $\phi(L) = I - \rho_1 L - \rho_2 L^2 - \cdots - \rho_J L^J$ is a low frequency lag polynomial and $\psi(L^{1/m}; \beta_k) x_{t,k} = \frac{1}{m} \sum_{j=0}^{m-1} \beta_{k,j} x_{t-j/m,k}$ is a high-frequency lag polynomial; see Andreou, Ghysels, and Kourtellos (2013) and Babii, Ghysels, and Striaukas (2020). Note that with $m = 1$ we have all data sampled at the same frequency and we recover a standard autoregressive distributed lag (ARDL) model. The ARDL-MIDAS regression has a group structure where a single group is defined as all lags of $x_{t,k}$ or all lags of y_t and following Babii, Ghysels, and Striaukas (2020), we focus on the sparse-group LASSO (sg-LASSO) estimator.[5] The sg-LASSO, denoted $\hat{\beta}$, solves the penalized least-squares problem

$$\min_{b \in \mathbf{R}^p} \|\mathbf{y} - \mathbf{X}b\|_T^2 + 2\lambda\Omega(b) \tag{1}$$

with

$$\Omega(b) = \alpha|b|_1 + (1-\alpha)\|b\|_{2,1},$$

where $|b|_1 = \sum_{j=1}^{p} |b_j|$ is the ℓ_1 norm corresponding to the LASSO penalty and $\|b\|_{2,1} = \sum_{G \in \mathcal{G}} |b_G|_2$ is the group LASSO penalty.

Our first result describes the large sample approximation to the distribution of the bias-corrected sg-LASSO estimator (and as a consequence of the LASSO and the group

[5]The sg-LASSO estimator allows selecting groups and important group members at the same time. Since the sparsity at the lag level is questionable in empirical applications.

LASSO) with serially correlated regression errors. Let $B = \hat{\Theta}\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\hat{\beta})/T$ denote that bias-correction for the sg-LASSO estimator, where $\hat{\Theta}$ is the nodewise LASSO estimator of the precision matrix Θ, cf., Meinshausen and Bühlmann (2006).[6] We measure the time series dependence with τ-dependence coefficients. For a σ-algebra $\mathcal{M}$ and a random vector $\xi \in \mathbf{R}^l$, the τ coefficient is defined as

$$\tau(\mathcal{M}, \xi) = \sup_{f \in \Lambda(\mathbf{R}^l)} \int_{\mathbf{R}} \|F_{f(\xi)|\mathcal{M}}(t) - F_{f(\xi)}(t)\|_1 \mathrm{d}t,$$

where $\Lambda(\mathbf{R}^l) = \{f : \mathbf{R}^l \to \mathbf{R} : \ |f(x) - f(y)| \leq |x - y|_2\}$ is a set of 1-Lipschitz functions, $F_{f(\xi)}$ is the CDF of $f(\xi)$, and $F_{f(\xi)|\mathcal{M}}$ is the CDF of $f(\xi)$ conditionally on $\mathcal{M}$; see also Dedecker and Prieur (2005), Lemma 1 for an equivalent variational characterization. Let $(\xi_t)_{t\in\mathbf{Z}}$ be a stochastic process and let $\mathcal{M}_t = \sigma(\xi_t, \xi_{t-1}, \dots)$ be its natural filtration. The τ-dependence coefficient is defined as

$$\tau_k = \sup_{j\geq 1} \max_{1\leq l\leq j} \frac{1}{l} \sup_{t+k\leq t_1 < \cdots < t_l} \tau(\mathcal{M}_t, (\xi_{t_1}, \dots, \xi_{t_l})), \qquad k \geq 0.$$

The process is called τ-dependent if its τ-dependence coefficients tend to zero. The τ-dependence coefficients were introduced in Dedecker and Prieur (2004) and Dedecker and Prieur (2005) as dependence measures weaker than mixing. In particular, they provide sharper bounds on autocovariances than mixing coefficients, see also Dedecker and Doukhan (2003).

The following assumptions impose several mild regularity conditions on the DGP.

Assumption 2.1 (Data). *$(y_t, X_t)_{t\in\mathbf{Z}}$ is a stationary process such that (i) $\max_{j\in[p]} \|u_t X_{t,j}\|_q = O(1)$ for some $q > 2$; (ii) $\max_{j,k\in[p]} \|X_{t,j}X_{t,k}\|_{\tilde{q}} = O(1)$ for some $\tilde{q} > 2$; (iii) $(u_t X_t)_{t\in\mathbf{Z}}$ is a vector of τ-dependent processes with $\tau_k \leq ck^{-a}$ for some $a > (q-1)/(q-2)$ and $c > 0$; (iv) $(X_t X_t^\top)_{t\in\mathbf{Z}}$ is a matrix of τ-dependent processes with $\tilde{\tau}_k \leq \tilde{c}k^{-\tilde{a}}$ for some $\tilde{a} > (\tilde{q}-1)/(\tilde{q}-2)$ and $\tilde{c} > 0$.*

Note that we do not require the sub-Gaussianity and allow for the temporal dependence to vanish at a polynomial rate.

[6] We assume that Θ exists, see assumptions below and the following subsection for more details.

Assumption 2.2 (Covariance matrix)**.** *There exists a universal constant* $\gamma > 0$ *such that the smallest eigenvalue of* $\Sigma = \mathbb{E}[X_t X_t^\top]$ *is bounded away from zero by* γ*.*

This assumption ensures in particular that the precision matrix $\Theta = \Sigma^{-1}$ exists.

Assumption 2.3 (Regularization parameter)**.** *The regularization parameter satisfies*

$$\lambda \sim \left(\frac{p}{\delta T^{\kappa-1}}\right)^{1/\kappa} \vee \sqrt{\frac{\log(8p/\delta)}{T}}$$

for some $\delta \in (0,1)$ *and* $\kappa = \frac{(a+1)q-1}{a+q-1}$*.*

The choice of the regularization parameter is governed by the Fuk-Nagaev inequality; see Theorem 3.1 and Eq. 4 following the discussion after that theorem in the next section.

Assumption 2.4. *(i)* $\|\mathbf{m}-\mathbf{X}\beta\|_T^2 = O_P(s_\alpha \lambda^2)$*; (ii)* $s_\alpha^{\tilde{\kappa}/2} p = o(T^{(\tilde{\kappa}-1)/2})$ *and* $p^2 \exp(-A_2 T/s_\alpha^2) = o(1)$*, where* $s_\alpha^{1/2} = \alpha\sqrt{|S_0|} + (1-\alpha)\sqrt{|\mathcal{G}_0|}$ *is the effective sparsity of* β *and* $\tilde{\kappa} = \frac{(\tilde{a}+1)\tilde{q}-1}{\tilde{a}+\tilde{q}-1}$*.*

This assumption in conjunction with the previous assumptions is needed for the consistency of the sg-LASSO estimator; see Theorem A.1 in the supplementary material. The following assumption provides an additional set of conditions needed for establishing the debiased central limit theorem for the sg-LASSO estimator.

Assumption 2.5. *Let* $G \subset [p]$ *be a group of fixed size and suppose that (i)* $\sup_x \mathbb{E}[u_0^2 | X_0 = x] \leq C < \infty$*; (ii)* $\max_{j\in G} |\Theta_j|_1 = O(1)$*; (iii) coordinates of* $(\xi_t)_{t\in\mathbf{Z}}$ *and* $(\xi_t \xi_k \top)_{t\in\mathbf{Z}}$ *have* τ*-dependence coefficients satisfying* $\sum_{t=1}^{\infty} \tau_{T,t}^{\frac{q-2}{q-1}} = O(1)$ *for some* $q > 2$ *and* $\sum_{t=1}^{\infty} \tilde{\tau}_{T,t} = O(1)$ *for all* $k \geq t$*, where* $\xi_t = u_t \Theta_G X_t$*; (iv) the long run variance of* $(v_{t,j}^2)_{t\in\mathbf{Z}}$ *exists for every* $j \in G$*, where* $v_{t,j}$ *is the regression error in* j^{th} *nodewise LASSO regression; (v)* $s \log p = o(T^{1/2})$ *and* $s^{\kappa/2} p = o(T^{3\kappa/4-1})$*, where* $s = s_\alpha \vee S$*,* $S = \max_{j\in G} S_j$*, and* S_j *is the number of non-zero coefficients in the* j^{th} *row of the precision matrix; (vi)* $\|\mathbf{m} - \mathbf{X}\beta\|_T = o_P(T^{-1/2})$*.*

Under the maintained assumptions, we obtain the following approximation to the large sample distribution of the debiased sg-LASSO estimator with serially correlated regression errors.

Theorem 2.1. *Suppose that Assumptions 2.1, 2.2, 2.3, 2.4, and 2.5 are satisfied for the sg-LASSO regression and for each nodewise LASSO regression $j \in G$. Then*

$$\sqrt{T}(\hat{\beta}_G + B_G - \beta_G) \xrightarrow{d} N(0, \Xi_G)$$

with $\Xi_G = \lim_{T\to\infty} \operatorname{Var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T u_t \Theta_G X_t\right)$.

Our debiased CLT extends van de Geer, Bühlmann, Ritov, and Dezeure (2014) to the weakly dependent data with serially correlated regression errors and describes the long run asymptotic variance for the low-dimensional group of regression coefficients estimated with the sg-LASSO. One could also consider Gaussian approximations for groups of increasing size, which requires an appropriate high-dimensional Gaussian approximation result for τ-dependent processes and is left for future research; see Chernozhukov, Chetverikov, Kato, et al. (2013) for such a result in the i.i.d. case.

2.2 HAC estimator

The bias-correction term B and the expression of the long-run variance in Theorem 2.1 depend on the estimator of the precision matrix $\Theta = \Sigma^{-1}$. There exist several approaches to the estimation of the precision matrix in the high-dimensional setting: the nodewise LASSO, see Bühlmann and van de Geer (2011); the weighted graphical LASSO see Janková and van de Geer (2018); or the ridge regression. We focus on nodewise LASSO regressions introduced in Meinshausen and Bühlmann (2006) and used subsequently in van de Geer, Bühlmann, Ritov, and Dezeure (2014) in the i.i.d. setting. The estimator is based on the observation that the covariance matrix of the partitioned vector $X = (X_j, X_{-j}^\top)^\top \in \mathbf{R} \times \mathbf{R}^{p-1}$ can be written as

$$\Sigma = \mathbb{E}[XX^\top] = \begin{pmatrix} \Sigma_{j,j} & \Sigma_{j,-j} \\ \Sigma_{-j,j} & \Sigma_{-j,-j} \end{pmatrix},$$

where $\Sigma_{j,j} = \mathbb{E}[X_j^2]$ and all other elements defined similarly. Then by the partitioned inverse formula, the 1^{st} row of the precision matrix $\Theta = \Sigma^{-1}$ is computed as

$$\Theta_j = \sigma_j^{-2}\begin{pmatrix} 1 & -\gamma_j^\top \end{pmatrix},$$

where $\gamma_j = \Sigma_{-j,-j}^{-1}\Sigma_{-j,j}$ is the projection coefficient in the regression of X_j on X_{-j}

$$X_j = X_{-j}^\top \gamma_j + v_j, \qquad \mathbb{E}[X_{-j}v_j] = 0, \tag{2}$$

and $\sigma_j^2 = \Sigma_{j,j} - \Sigma_{j,-j}\gamma_j = \mathbb{E}[v_j^2]$ is the variance of the regression error. Therefore, we can estimate the 1^{st} row of the precision matrix as $\hat{\Theta}_j = \hat{\sigma}_j^{-2}\begin{pmatrix} 1 & -\hat{\gamma}_j^\top \end{pmatrix}$ with $\hat{\gamma}_j$ solving

$$\min_{\gamma \in \mathbf{R}^{p-1}} \|\mathbf{X}_j - \mathbf{X}_{-j}\gamma\|_T^2 + 2\lambda_j|\gamma|_1$$

and

$$\hat{\sigma}_j^2 = \|\mathbf{X}_j - \mathbf{X}_{-j}\hat{\gamma}_j\|_T^2 + \lambda_j|\hat{\gamma}_j|,$$

where $\mathbf{X}_j \in \mathbf{R}^T$ is the column vector of observations of X_j and $\mathbf{X}_{-j}$ is the $T \times (p-1)$ matrix of observations of $X_{-j}^\top$. More generally, the nodewise LASSO estimator of the precision matrix can be written then as $\hat{\Theta} = \hat{B}^{-1}\hat{C}$ with

$$\hat{C} = \begin{pmatrix} 1 & -\hat{\gamma}_{1,1} & \dots & -\hat{\gamma}_{1,p-1} \\ -\hat{\gamma}_{2,1} & 1 & \dots & -\hat{\gamma}_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ -\hat{\gamma}_{p-1,1} & \dots & -\hat{\gamma}_{p-1,p-1} & 1 \end{pmatrix} \qquad \text{and} \qquad \hat{B} = \operatorname{diag}(\hat{\sigma}_1^2, \dots, \hat{\sigma}_p^2).$$

To the best of our knowledge the HAC estimator for the LASSO has not been studied in the relevant literature.[7] We focus on the HAC estimator for the sparse-group LASSO, covering the LASSO and the group LASSO as special cases. For a group $G \subset [p]$ of a fixed size, the HAC estimator of the long-run variance is

$$\hat{\Xi}_G = \sum_{|k|<T} K\left(\frac{k}{M_T}\right)\hat{\Gamma}_k, \tag{3}$$

where $\hat{\Gamma}_k = \hat{\Theta}_G\left(\frac{1}{T}\sum_{t=1}^{T-k}\hat{u}_t\hat{u}_{t+k}X_tX_{t+k}^\top\right)\hat{\Theta}_G^\top$, and $\hat{\Gamma}_{-k} = \hat{\Gamma}_k^\top$, where the kernel function $K : \mathbf{R} \to [-1, 1]$ with $K(0) = 1$ gives less weight to more distant noisy covariances, and M_T is a bandwidth parameter, see Parzen (1957) and Andrews (1991). Several choices of

[7]Convergence rates of the HAC estimator for non-sparse high-dimensional least-squares under weak dependence conditions have been obtained previously, e.g., in Li and Liao (2019).

the kernel function are possible, for example, the Parzen kernel is

$$K_{PR}(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for } 0 \leq |x| \leq 1/2, \\ 2(1 - |x|)^3 & \text{for } 1/2 \leq |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases},$$

see appendix for more details on the choice of the kernel.

Assumption 2.6. *Suppose that $(u_t^2)_{t\in\mathbf{Z}}$ and $(v_{t,j}^2)_{t\in\mathbf{Z}}$ for each $j \in G$ have finite long run variances.*

The following result derives the convergence rate of the HAC estimator for a group of coefficients G estimated using the LASSO to the long-run variance, which under the stationarity simplifies to

$$\Xi_G = \sum_{k\in\mathbf{Z}} \Gamma_k$$

with $\Gamma_k = \Theta_G \mathbb{E}[u_t u_{t+k} X_t X_{t+k}^\top]\Theta_G^\top$ and $\Gamma_{-k} = \Gamma_k^\top$.

Theorem 2.2. *Suppose that assumptions of Theorem 2.1 and Assumption 2.6 are satisfied with $\kappa \geq \tilde{q}$. Suppose also that Assumptions A.1.1, and A.1.2 in the appendix are satisfied with $V_t = (u_t v_{t,j}/\sigma_j^2)_{j\in G}$. Then if $s\sqrt{\frac{\log p}{T}} = o(1)$ and $s^\kappa p = o(T^{4\kappa/5-1})$ as $M_T \to \infty$ and $T \to \infty$*

$$\|\hat{\Xi}_G - \Xi_G\| = O_P\left(M_T \left(\frac{sp^{1/\kappa}}{T^{1-1/\kappa}} \vee s\sqrt{\frac{\log p}{T}} + \frac{s^2 p^{2/\kappa}}{T^{2-3/\kappa}} + \frac{s^3 p^{5/\kappa}}{T^{4-5/\kappa}} \right) + M_T^{-\varsigma} + T^{-(\varsigma\wedge 1)} \right).$$

The first term in the inner parentheses is of the same order as the estimation error of the maximum between the estimation errors of the sg-LASSO and the nodewise LASSO. The optimal choice of the bandwidth parameter depends on the number of covariates p and the dependence-tails exponent κ.[8]

It is worth stressing that this result does not follow from previous results that provide a comprehensive treatment of the kernel HAC estimation based on residuals in the fixed-dimensional case with $\sqrt{T}$-consistent estimators; see Andrews (1991).

[8]A comprehensive study of the optimal bandwidth choice based on higher-order asymptotic expansions is beyond the scope of this paper and is left for the future research, see, e.g., Lazarus, Lewis, Stock, and Watson (2018) for the recent literature review and practical recommendations in the low-dimensional case.

2.3 Inference for low-dimensional groups

In the (mixed-frequency) distributed lag setting, testing the statistical significance of a single covariate amounts to using the Wald test, which has the interpretation of the Granger causality test in the econometrics literature. Therefore, we are interested in testing

$$H_0: \; R\beta_G = 0 \qquad \text{against} \qquad H_1: \; R\beta_G \neq 0$$

for some low-dimensional group of regression coefficients $G \subset [p]$, where R is $r \times |G|$ matrix, e.g., $R = I_{|G|}$ if we want to test that all coefficients are jointly zero. Assuming that R is a full row rank matrix and that $\hat{\Xi}_G$ is positive definite, the Wald statistics is

$$W_T = T\left[R(\hat{\beta}_G + B_G - \beta_G)\right]^\top \left(R\hat{\Xi}_G R^\top\right)^{-1}\left[R(\hat{\beta}_G + B_G - \beta_G)\right].$$

It follows from Theorems 2.1 and 2.2 that under H_0, $W_T \xrightarrow{d} \chi^2_r$. The Wald test rejects when $W_T > q_{1-\alpha}$, where $q_{1-\alpha}$ is the quantile of order $1-\alpha$ of χ^2_r. This can be extended to nonlinear restrictions by the usual Delta method argument.

For testing hypotheses on the increasing set of regression coefficients, it is preferable to use the non-pivotal sup-norm based statistics instead of the one based on the quadratic form, see e.g., Ghysels, Hill, and Motegi (2020). The sup-norm statistics is known to perform remarkably in the high-dimensional setting and allows the exponential dependence on the dimension in the i.i.d. setting, cf., Chernozhukov, Chetverikov, Kato, et al. (2013). Such an extension for approximately sparse regression models is left for future research.

3 Fuk-Nagaev inequality

In this section, we derive a suitable for us version of the Fuk-Nagaev concentration inequality for the maximum of high-dimensional sums. The inequality allows for the data with polynomial tails and τ-dependence coefficients decreasing at a polynomial rate. Our main result does not require that the time series is stationary.

Theorem 3.1. *Let $(\xi_t)_{t\in\mathbf{Z}}$ be a centered stochastic process in $\mathbf{R}^p$ such that (i) for some $q > 2$, $\max_{j\in[p],t\in[T]} \|\xi_{t,j}\|_q = O(1)$; (ii) for every $j \in [p]$, τ-dependence coefficients of*

$\xi_{t,j}$ *satisfy* $\tau_k^{(j)} \leq ck^{-a}$ *for some universal constants* $a, c > 0$. *Then there exist universal constants* $c_1, c_2 > 0$ *such that for every* $u > 0$

$$\Pr\left(\left|\sum_{t=1}^{T}\xi_t\right|_\infty > u\right) \leq c_1 pTu^{-\kappa} + 4p\exp\left(-\frac{c_2u^2}{B_T^2}\right),$$

where $\kappa = \frac{(a+1)q-1}{a+q-1}$, $B_T^2 = \max_{j\in[p]}\sum_{t=1}^{T}\sum_{k=1}^{T}|\mathrm{Cov}(\xi_{t,j},\xi_{k,j})|$.

The inequality describes the mixture of the polynomial and Gaussian tails for the maximum of high-dimensional sums. In the limiting case of the i.i.d. data, as $a \to \infty$, the dependence-tails exponent κ approaches q and we recover the inequality for the independent data stated in Fuk and Nagaev (1971), Corollary 4 for $p = 1$. In this sense, our inequality is sharp. It is well-known that the Fuk-Nagaev inequality delivers sharper estimates of tail probabilities in contrast to Markov's bound in conjunction with Rosenthal's moment inequality, cf., Nagaev (1998). Our proof relies on the blocking technique, see Bosq (1993), and the coupling inequality for τ-dependent sequences, see Dedecker and Prieur (2004), Lemma 5. In contrast to previous results, e.g., Dedecker and Prieur (2004), Theorem 2, our inequality reflects the mixture of the polynomial and the exponential tails and can readily be applied to the LASSO-type estimators.

For stationary processes, by Lemma A.1.2 in the appendix, $B_T^2 = O(T)$ as long as $a > \frac{q-1}{q-2}$, whence we obtain from Theorem 3.1 that for every $\delta \in (0,1)$

$$\Pr\left(\left|\frac{1}{T}\sum_{t=1}^{T}\xi_t\right|_\infty \leq C\left(\frac{p}{\delta T^{\kappa-1}}\right)^{1/\kappa} \vee \sqrt{\frac{\log(8p/\delta)}{T}}\right) \geq 1-\delta, \tag{4}$$

where $C > 0$ is some finite universal constant.

4 Monte Carlo experiments

In this section, we aim to assess the debiased HAC inferences in finite samples. To assess the small sample properties of the HAC estimator for the low-dimensional parameter in a high dimensional data setting we draw covariates $\{x_{t,j}, j \in [p]\}$ independently from the AR(1) process

$$x_{t,j} = \rho x_{t-1,j} + \epsilon_{t,j},$$

where ρ is the persistence parameter. The regression error follows the AR(1) process

$$u_t = \rho u_{t-1} + \nu_t,$$

where errors are either both $\epsilon_{t,j}, \nu_t \sim_{i.i.d.} N(0,1)$ or $\epsilon_{t,j}, \nu_t \sim_{i.i.d.}$ student-$t(5)$. The vector of population regression coefficients β has the first five non-zero entries $(5,4,3,2,1)$ and all remaining entries are zero. The sample size is $T = 500$ and the number of covariates is $p \in \{20, 200\}$. We set the persistence parameter $\rho = 0.8$ and focus on the LASSO estimator to estimate coefficients $\hat{\beta}$. Throughout the experiment, we choose the LASSO tuning parameters using the rule-of-thumb $\lambda = \hat{\sigma}\sqrt{T}\Phi^{-1}(1 - 0.1/2p)$, where Φ is a CDF of the standard normal distribution, $\hat{\sigma}$ is a preliminary estimate of $\sigma = \sqrt{\mathbb{E}u^2}$, see Chernozhukov, Hansen, and Spindler (2016).[9]

We report the average coverage (av. cov) and the average length of confidence intervals for the nominal coverage of 0.95 and on a grid of values of the bandwidth parameter $M_T \in \{10, 20, 40, \dots, 160\}$, using the Parzen and the Quadratic spectral kernels. We estimate the long run covariance matrix $\hat{\Xi}$ based on the LASSO residuals $\hat{u}_t$ and the precision matrix $\hat{\Theta}$ using nodewise LASSO regressions. The first step is to compute scores $\hat{V}_t = \hat{u}_t X_t$, where $\hat{u}_t = y_t - X_t^\top \hat{\beta}$, and $\hat{\beta}$ is the LASSO estimator. Then we compute the high-dimensional HAC estimator using the formuala in equation (3). We compute the pivotal statistics for each MC experiment $i \in [N]$ and each coefficient $j \in [p]$ as $\text{pivot}_j^{(i)} \triangleq (\hat{\beta}_j^{(i)} + B_j^{(i)} - \beta)/\sqrt{\hat{\Xi}_{j,j}^{(i)}/T}$, where $B_j^{(i)} = \hat{\Theta}_j^{(i)} \mathbf{X}^{\top(i)} \hat{\mathbf{u}}^{(i)}/T$, $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\beta}$. Then we compute the empirical coverage as

$$\text{av.cov}_j = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{\text{pivot}_j^i \in [-1.96, 1.96]\}$$

and the average confidence interval length as $\text{length}_j = \frac{1}{N}\sum_{i=1}^{N} 2 \times 1.96 \times \sqrt{\hat{\Xi}_{j,j}^{(i)}/T}$. The number of Monte Carlo experiments was set at $N = 5000$.

We report average results over the active and inactive sets of the vector of coefficients. Table 1 shows results for the Gaussian data and Table 2 for the student-$t(5)$ data for the Parzen kernel. We find that the optimal bandwidth parameter M_T appears to be smaller

[9] An alternative, less conservative, but computationally more intensive bootstrap-based algorithm is discussed in Chernozhukov, Härdle, Huang, and Wang (2019).

when the number of regressors p is larger. Indeed, since the number of lags in the HAC estimator is M_T, fewer lags should be taken into account as p increases. For Gaussian data and the active set of coefficients, the optimal bandwidth parameter is around 30 for $p = 200$ and 80 for $p = 20$. A similar pattern is found for the inactive set, although with slightly smaller optimal bandwidth parameters – 20 for large p case and 60 for small p. For heavy-tailed data simulated from the Student's t-distribution, the optimal bandwidth parameter is lower. Table 3 and 4 report similar results for the Quadratic spectral kernel. Overall, the simulation results confirm our theoretical findings.

	Average coverage (av. cov)				Confidence interval length			
	Active set of β		Inactive set of β		Active set of β		Inactive set of β	
$M_T \backslash p$	20	200	20	200	20	200	20	200
10	0.912	0.930	0.918	0.942	0.330	0.320	0.321	0.304
20	0.932	0.944	0.940	0.951	0.358	0.340	0.344	0.312
40	0.942	0.956	0.947	0.956	0.375	0.364	0.350	0.312
60	0.947	0.965	0.950	0.960	0.386	0.384	0.349	0.311
80	0.950	0.971	0.952	0.964	0.395	0.402	0.348	0.309
100	0.953	0.976	0.954	0.968	0.403	0.419	0.346	0.308
120	0.956	0.980	0.957	0.972	0.411	0.435	0.344	0.306
140	0.959	0.983	0.958	0.976	0.419	0.451	0.342	0.305
160	0.962	0.984	0.960	0.980	0.427	0.465	0.340	0.303

Table 1: HAC inference simulation results – The table reports average coverage (first four columns) and average length of confidence intervals (last four columns) for active and inactive sets of β. The data is generated using Gaussian distribution. We report results for a set of bandwidth parameter M_T values.

5 Conclusion

This paper develops valid inferential methods for high-dimensional time series regressions estimated with the sparse-group LASSO (sg-LASSO) estimator that encompasses the LASSO and the group LASSO as special cases. We derive the debiased central limit theorem with the explicit bias correction for the sg-LASSO with serially correlated regres-

	Average coverage (av. cov)				Confidence interval length			
	Active set of β		Inactive set of β		Active set of β		Inactive set of β	
$M_T \backslash p$	20	200	20	200	20	200	20	200
10	0.926	0.951	0.928	0.953	0.348	0.347	0.332	0.319
20	0.946	0.963	0.948	0.959	0.379	0.369	0.355	0.324
40	0.956	0.972	0.955	0.963	0.402	0.397	0.361	0.323
60	0.960	0.978	0.959	0.966	0.417	0.423	0.361	0.322
80	0.965	0.983	0.962	0.970	0.431	0.446	0.359	0.320
100	0.969	0.986	0.965	0.974	0.443	0.467	0.357	0.318
120	0.973	0.988	0.968	0.978	0.455	0.487	0.355	0.316
140	0.975	0.991	0.972	0.982	0.466	0.506	0.353	0.314
160	0.977	0.992	0.974	0.986	0.477	0.524	0.351	0.313

Table 2: HAC inference simulation results – The table reports average coverage (first four columns) and average length of confidence intervals (last four columns) for active and inactive sets of β. The data is generated using student-t(5) distribution. We report results for a set of bandwidth parameter M_T values.

	Average coverage (av. cov)				Confidence interval length			
	Active set of β		Inactive set of β		Active set of β		Inactive set of β	
$M_T \backslash p$	20	200	20	200	20	200	20	200
10	0.929	0.941	0.937	0.950	0.355	0.336	0.343	0.312
20	0.940	0.953	0.946	0.954	0.372	0.357	0.350	0.313
40	0.947	0.967	0.949	0.959	0.389	0.390	0.349	0.310
60	0.952	0.976	0.952	0.965	0.403	0.420	0.345	0.307
80	0.957	0.981	0.955	0.972	0.417	0.446	0.341	0.304
100	0.962	0.985	0.958	0.978	0.429	0.471	0.338	0.301
120	0.965	0.987	0.961	0.983	0.441	0.493	0.334	0.299
140	0.969	0.989	0.964	0.985	0.452	0.514	0.331	0.296
160	0.972	0.991	0.966	0.987	0.463	0.534	0.328	0.294

Table 3: HAC inference simulation results table – The table reports average coverage (first four columns) and average length of confidence intervals (last four columns) for active and inactive sets of β. The data is generated using Gaussian distribution. We report results for a set of bandwidth parameter M_T values. The long run variance is computed using Quadratic spectral kernel.

	Average coverage (av. cov)				Confidence interval length			
	Active set of β		Inactive set of β		Active set of β		Inactive set of β	
$M_T \backslash p$	20	200	20	200	20	200	20	200
10	0.944	0.961	0.947	0.958	0.376	0.364	0.354	0.325
20	0.953	0.970	0.954	0.961	0.397	0.388	0.362	0.324
40	0.961	0.979	0.958	0.966	0.422	0.431	0.360	0.320
60	0.968	0.986	0.963	0.971	0.443	0.468	0.356	0.317
80	0.973	0.990	0.967	0.978	0.463	0.501	0.352	0.314
100	0.977	0.992	0.972	0.984	0.481	0.531	0.348	0.311
120	0.980	0.993	0.975	0.988	0.498	0.558	0.345	0.308
140	0.983	0.995	0.977	0.990	0.514	0.584	0.342	0.305
160	0.984	0.995	0.979	0.991	0.529	0.608	0.339	0.302

Table 4: HAC inference simulation results table – The table reports average coverage (first four columns) and average length of confidence intervals (last four columns) for active and inactive sets of β. The data is generated using student-t(5) distribution. We report results for a set of bandwidth parameter M_T values. The long run variance is computed using Quadratic spectral kernel.

sion errors. Furthermore, we also study HAC estimators of the long-run variance for low dimensional groups of regression coefficients and characterize how the optimal bandwidth parameter should scale with the sample size, the temporal dependence, as well as tails of the data. These results lead to the valid t- and Wald tests for the low-dimensional subset of parameters, including Granger causality tests. Our treatment relies on a new suitable variation of the Fuk-Nagaev inequality for τ-dependent processes which allows us to handle the time series data with polynomial tails.

An interesting avenue for future research is to study more carefully the problem of the optimal data-driven bandwidth choice, see, e.g., Sun, Phillips, and Jin (2008) for some steps in this direction in low dimensional settings.

References

Almon, S. (1965): "The distributed lag between capital appropriations and expenditures," *Econometrica*, 33(1), 178–196.

Andreou, E., E. Ghysels, and A. Kourtellos (2013): "Should macroeconomic forecasters use daily financial data and how?," *Journal of Business and Economic Statistics*, 31, 240–251.

Andrews, D. W. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59(3), 817–858.

Babii, A., E. Ghysels, and J. Striaukas (2020): "Machine learning time series regressions with an application to nowcasting," arXiv preprint arXiv:2005.

Bartlett, M. (1948): "Smoothing periodograms from time-series with continuous spectra," *Nature*, 161(4096), 686–687.

Belloni, A., V. Chernozhukov, and C. Hansen (2010): "LASSO Methods for Gaussian Instrumental Variables Models," arXiv preprint arXiv:1012.1297.

——— (2014): "Inference on treatment effects after selection among high-dimensional controls," *Review of Economic Studies*, 81(2), 608–650.

Bosq, D. (1993): "Bernstein-type large deviations inequalities for partial sums of strong mixing processes," *Statistics*, 24(1), 59–70.

Bühlmann, P., and S. van de Geer (2011): *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Bybee, L., B. T. Kelly, A. Manela, and D. Xiu (2020): "The structure of economic news," *National Bureau of Economic Research*, and http://structureofnews.com.

Chernozhukov, V., D. Chetverikov, K. Kato, et al. (2013): "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors," *Annals of Statistics*, 41(6), 2786–2819.

Chernozhukov, V., C. Hansen, and M. Spindler (2016): "hdm: High-dimensional metrics," *arXiv preprint arXiv:1608.00354*.

Chernozhukov, V., W. K. Härdle, C. Huang, and W. Wang (2019): "Lasso-driven inference in time and space," *Annals of Statistics* (forthcoming).

Chiang, H. D., and Y. Sasaki (2019): "Lasso under multi-way clustering: estimation and post-selection inference," *arXiv preprint arXiv:1905.02107*.

Daniell, P. (1946): "Discussion of paper by M.S. Bartlett," *Journal of the Royal Statistical Society Supplements*, 8, 88–90.

Dedecker, J., and P. Doukhan (2003): "A new covariance inequality and applications," *Stochastic Processes and their Applications*, 106(1), 63–80.

Dedecker, J., and C. Prieur (2004): "Coupling for τ-dependent sequences and applications," *Journal of Theoretical Probability*, 17(4), 861–885.

——— (2005): "New dependence coefficients. Examples and applications to statistics," *Probability Theory and Related Fields*, 132(2), 203–236.

Eicker, F. (1963): "Asymptotic normality and consistency of the least squares estimators for families of linear regressions," *Annals of Mathematical Statistics*, pp. 447–456.

——— (1967): "Limit theorems for regressions with unequal and dependent errors," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 59–82. Berkeley, CA: University of California Press.

Feng, G., S. Giglio, and D. Xiu (2019): "Taming the factor zoo: a test of new factors," *Journal of Finance* (forthcoming).

Fisher, I. (1925): "Our unstable dollar and the so-called business cycle," *Journal of the American Statistical Association*, 20(150), 179–202.

——— (1937): "Note on a short-cut method for calculating distributed lags," *Bulletin de l'Institut international de Statistique*, 29(3), 323–328.

Fuk, D. K., and S. V. Nagaev (1971): "Probability inequalities for sums of independent random variables," *Theory of Probability and Its Applications*, 16(4), 643–660.

Gallant, A. R. (1987): *Nonlinear statistical models.* John Wiley, New York.

Ghysels, E., J. B. Hill, and K. Motegi (2020): "Testing a large set of zero restrictions in regression models, with an application to mixed frequency Granger causality," *Journal of Econometrics* (forthcoming).

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006): "Predicting volatility: getting the most out of return data sampled at different frequencies," *Journal of Econometrics*, 131, 59–95.

Ghysels, E., A. Sinko, and R. Valkanov (2007): "MIDAS regressions: Further results and new directions," *Econometric Reviews*, 26(1), 53–90.

Haugh, L. D., and G. E. Box (1977): "Identification of dynamic regression (distributed lag) models connecting two time series," *Journal of the American Statistical Association*, 72(357), 121–130.

Huber, P. J. (1967): "The behavior of maximum likelihood estimates under nonstandard conditions," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 221–233. University of California Press.

Janková, J., and S. van de Geer (2018): "Inference in high-dimensional graphical models," in *Handbook of Graphical Models*, ed. by M. Maathuis, M. Drton, S. Lauritzen, and M. Wainwright, pp. 325–346. CRC Press.

Javanmard, A., and A. Montanari (2014): "Confidence intervals and hypothesis testing for high-dimensional regression," *The Journal of Machine Learning Research*, 15(1), 2869–2909.

Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson (2018): "HAR Inference: Recommendations for Practice," *Journal of Business and Economic Statistics*, 36(4), 541–559.

Li, J., and Z. Liao (2019): "Uniform nonparametric inference for time series," *Journal of Econometrics* (forthcoming).

Meinshausen, N., and P. Bühlmann (2006): "High-dimensional graphs and variable selection with the LASSO," *Annals of Statistics*, 34(3), 1436–1462.

Nagaev, S. V. (1998): "Some refinements of probabilistic and moment inequalities," *Theory of Probability and Its Applications*, 42(4), 707–713.

Neumann, M. H. (2013): "A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics," *ESAIM: Probability and Statistics*, 17, 120–134.

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

Parzen, E. (1957): "On consistent estimates of the spectrum of a stationary time series," *Annals of Mathematical Statistics*, 28(2), 329–348.

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

Shiller, R. J. (1973): "A distributed lag estimator derived from smoothness priors," *Econometrica*, 41, 775–788.

Simon, N., J. Friedman, T. Hastie, and R. Tibshirani (2013): "A sparse-group LASSO," *Journal of Computational and Graphical Statistics*, 22(2), 231–245.

Sims, C. A. (1971a): "Discrete approximations to continuous time distributed lags in econometrics," *Econometrica*, 39, 545–563.

——— (1971b): "Distributed lag estimation when the parameter space is explicitly infinite-dimensional," *Annals of Mathematical Statistics*, 42(5), 1622–1636.

——— (1972): "The role of approximate prior restrictions in distributed lag estimation," *Journal of the American Statistical Association*, 67(337), 169–175.

Sun, Y., P. C. Phillips, and S. Jin (2008): "Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing," *Econometrica*, 76(1), 175–194.

van de Geer, S., P. Bühlmann, Y. Ritov, and R. Dezeure (2014): "On asymptotically optimal confidence regions and tests for high-dimensional models," *Annals of Statistics*, 42(3), 1166–1202.

White, H. (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, pp. 817–838.

Zhang, C.-H., and S. S. Zhang (2014): "Confidence intervals for low dimensional parameters in high dimensional linear models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 217–242.

Zhang, X., and G. Cheng (2017): "Simultaneous inference for high-dimensional linear models," *Journal of the American Statistical Association*, 112(518), 757–768.

APPENDIX

A.1 Proofs

Proof of Theorem 2.1. By Fermat's rule, the sg-LASSO satisfies

$$\mathbf{X}^\top(\mathbf{X}\hat{\beta} - \mathbf{y})/T + \lambda z^* = 0$$

for some $z^* \in \partial\Omega(\hat{\beta})$, where $\partial\Omega(\hat{\beta})$ is the sub-differential of $b \mapsto \Omega(b)$ at $\hat{\beta}$. Rearranging this expression and multiplying by $\hat{\Theta}$

$$\hat{\beta} - \beta + \hat{\Theta}\lambda z^* = \hat{\Theta}\mathbf{X}^\top\mathbf{u}/T + (I - \hat{\Theta}\hat{\Sigma})(\hat{\beta} - \beta) + \hat{\Theta}\mathbf{X}^\top(\mathbf{m} - \mathbf{X}\beta)/T.$$

Plugging in λz^* and multiplying by $\sqrt{T}$

$$\begin{aligned}\sqrt{T}(\hat{\beta} - \beta + B) &= \hat{\Theta}\mathbf{X}^\top\mathbf{u}/\sqrt{T} + \sqrt{T}(I - \hat{\Theta}\hat{\Sigma})(\hat{\beta} - \beta) + \hat{\Theta}\mathbf{X}^\top(\mathbf{m} - \mathbf{X}\beta)/\sqrt{T} \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^T u_t\Theta X_t + \frac{1}{\sqrt{T}}\sum_{t=1}^T u_t(\hat{\Theta} - \Theta)X_t + \sqrt{T}(I - \hat{\Theta}\hat{\Sigma})(\hat{\beta} - \beta) \\ &\quad + \hat{\Theta}\mathbf{X}^\top(\mathbf{m} - \mathbf{X}\beta)/\sqrt{T}.\end{aligned}$$

Next, we look at coefficients corresponding to $G \subset [p]$

$$\begin{aligned}\sqrt{T}(\hat{\beta}_G - \beta_G + B_G) &= \frac{1}{\sqrt{T}}\sum_{t=1}^T u_t\Theta_G X_t + \frac{1}{\sqrt{T}}\sum_{t=1}^T u_t(\hat{\Theta}_G - \Theta_G)X_t + \sqrt{T}(I - \hat{\Theta}\hat{\Sigma})_G(\hat{\beta} - \beta) \\ &\quad + \hat{\Theta}_G\mathbf{X}^\top(\mathbf{m} - \mathbf{X}\beta)/\sqrt{T} \\ &\triangleq I_T + II_T + III_T + IV_T.\end{aligned}$$

We will show that $I_T \xrightarrow{d} N(0, \Xi_G)$ by the triangular array CLT, cf., Neumann (2013), Theorem 2.1. To that end, by the Crámer-Wold theorem, it is sufficient to show that $z^\top I_T \xrightarrow{d} z^\top N(0, \Xi_G)$ for every $z \in \mathbf{R}^{|G|}$. Note that under Assumptions 2.1 and 2.5 (i)-(ii)

$$\begin{aligned}\sum_{t=1}^T \mathbb{E}\left|\frac{z^\top\xi_t}{\sqrt{T}}\right|^2 &= \mathbb{E}|u_t z^\top\Theta_G X_t|^2 \\ &\leq Cz^\top\Theta_G\Sigma\Theta_G^\top z \\ &= Cz^\top(\Theta_G)_G z \\ &= O(1)\end{aligned}$$

and that under Assumption 2.1 (i), for every $\epsilon > 0$

$$\sum_{t=1}^{T} \mathbb{E}\left[\left|\frac{z^\top \xi_t}{\sqrt{T}}\right|^2 \mathbf{1}\left\{\left|z^\top \xi_t\right| > \epsilon\sqrt{T}\right\}\right] \leq \frac{\mathbb{E}\left|z^\top \xi_t\right|^q}{(\epsilon\sqrt{T})^{q-2}}$$

$$= o(1),$$

where we use the fact that by the Minkowski inequality and Assumptions 2.1 (i) and 2.5 (ii) to show that

$$\begin{aligned}\mathbb{E}\|z^\top \xi_t\|_q &\leq \sum_{k\in G}|z_k|\sum_{j\in[p]}|\Theta_{k,j}|\|u_t X_{t,j}\|_q \\ &= O\left(|z|_1 \max_{k\in G}|\Theta_k|_1\right) \\ &= O(1).\end{aligned} \tag{A.1}$$

Next, under stationarity

$$\operatorname{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z^\top \xi_t\right) = \operatorname{Var}(z^\top \xi_0) + 2\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\operatorname{Cov}(z^\top \xi_0, z^\top \xi_t)$$

and under Assumptions 2.1 (i) and 2.5 (iii), by Proposition A.1.1 and the Minkowski inequality

$$\begin{aligned}\sum_{t=1}^{\infty}|\operatorname{Cov}(z^\top \xi_0, z^\top \xi_t)| &\leq \|z^\top \xi_0\|_q^{q/(q-1)}\sum_{t=1}^{\infty}\|\mathbb{E}(z^\top \xi_t|\mathcal{M}_0) - \mathbb{E}(z^\top \xi_t)\|_1^{\frac{q-2}{q-1}} \\ &\leq \|z^\top \xi_0\|_q^{q/(q-1)}|z|_1^{\frac{q-2}{q-1}}\sum_{t=1}^{\infty}\tau_{T,t}^{\frac{q-2}{q-1}} \\ &= O(1).\end{aligned}$$

By the Lebesgue dominated convergence, this shows that the long run variance exists

$$\lim_{T\to\infty}\operatorname{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z^\top \xi_t\right) = z^\top \Xi_G z.$$

Let $\mathcal{M}_t = \sigma(\xi_t, \xi_{t-1}, \xi_{t-2}, \dots)$. Then, for every measurable function $g : \mathbf{R}^h \to \mathbf{R}$ with $\sup_x |g(x)| \leq 1$, by Dedecker and Doukhan (2003), Proposition 1, for all $h \in \mathbf{N}$ and all indices $1 \leq t_1 < t_2 < \cdots < t_h < t_h + r \leq t_h + s \leq T$

$$\begin{aligned}
&\left|\mathrm{Cov}\left(g(z^\top \xi_{t_1}/\sqrt{T}, \dots, z^\top \xi_{t_h}/\sqrt{T}) z^\top \xi_{t_h}/\sqrt{T}, z^\top \xi_{t_h+r}/\sqrt{T}\right)\right| \\
&\leq \frac{1}{T}\int_0^{\gamma(\mathcal{M}_0, z^\top \xi_r)} Q_{z^\top \xi_0} \circ G_{z^\top \xi_r}(u)\mathrm{d}u \\
&\leq \frac{1}{T}\|\mathbb{E}(z^\top \xi_r|\mathcal{M}_0) - \mathbb{E}(z^\top \xi_r)\|_1^{\frac{q-2}{q-1}} \|z^\top \xi_0\|_q^{q/(q-1)} \\
&\leq \frac{1}{T}|z|_1^{\frac{q-2}{q-1}} \|z^\top \xi_0\|_q^{q/(q-1)} \tau_r^{\frac{q-2}{q-1}},
\end{aligned}$$

where the second line follows by stationarity and $\sup_x |g(x)| \leq 1$, the third since by Hölder's inequality and the change of variables

$$\int_0^{\|z^\top \xi_0\|_1} Q_{z^\top \xi_0}^{q-1} \circ G_{z^\top \xi_r}(u)\mathrm{d}u = \int_0^1 Q_{z^\top \xi_0}^{q}(u)\mathrm{d}u = \|z^\top \xi_0\|_q^q,$$

and the last by Proposition A.1.1. Similarly,

$$\begin{aligned}
&\left|\mathrm{Cov}\left(g(z^\top \xi_{t_1}/\sqrt{T}, \dots, z^\top \xi_{t_h}/\sqrt{T}), z^\top \xi_{t_h+r}/\sqrt{T} z^\top \xi_{t_h+s}/\sqrt{T}\right)\right| \\
&\leq \frac{1}{T}\int_0^{\gamma(\mathcal{M}_0, z^\top \xi_r z^\top \xi_s)} Q_g \circ G_{z^\top \xi_r z^\top \xi_s}(u)\mathrm{d}u \\
&\leq \frac{1}{T}\left\|\mathbb{E}(z^\top \xi_r z^\top \xi_s|\mathcal{M}_0) - \mathbb{E}(z^\top \xi_r z^\top \xi_s)\right\|_1 \\
&\leq \frac{1}{T}|z|_1^2 \tilde{\tau}_{T,r}.
\end{aligned}$$

Therefore, by Neumann (2013), Theorem 2.1, $z^\top I_T \xrightarrow{d} z^\top N(0, \Xi_G)$ for every $z \in \mathbf{R}^{|G|}$.

Next,

$$\begin{aligned}
|II_T|_\infty &= \left|(\hat{\Theta} - \Theta)_G\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T u_t X_t\right)\right|_\infty \\
&\leq \|\hat{\Theta}_G - \Theta_G\|_\infty \left|\frac{1}{\sqrt{T}}\sum_{t=1}^T u_t X_t\right|_\infty \\
&= O_P\left(\frac{S p^{1/\kappa}}{T^{1-1/\kappa}} \vee S\sqrt{\frac{\log p}{T}}\right) O_P\left(\frac{p^{1/\kappa}}{T^{1/2-1/\kappa}} + \sqrt{\log p}\right) \\
&= o_P(1),
\end{aligned}$$

where the second line follows by $|Ax|_\infty \leq \|A\|_\infty |x|_\infty$, the third line by Proposition A.1.2 and the inequality in Eq. 4, and the last under Assumption 2.5 (v).

Likewise, using $|Ax|_\infty \leq \max_{j,k}|A_{j,k}|_\infty |x|_1$, by Proposition A.1.2 and Theorem A.1

$$\begin{aligned}|(I-\hat{\Theta}\hat{\Sigma})_G(\hat{\beta}-\beta)|_\infty &\leq \max_{j\in G}|(I-\hat{\Theta}\hat{\Sigma})_j|_\infty|\hat{\beta}-\beta|_1\\ &= O_P\left(\frac{p^{1/\kappa}}{T^{1-1/\kappa}}\vee\sqrt{\frac{\log p}{T}}\right)O_P\left(\frac{s_\alpha p^{1/\kappa}}{T^{1-1/\kappa}}\vee s_\alpha\sqrt{\frac{\log p}{T}}\right).\end{aligned}$$

Under Assumption 2.5 (v), this shows that

$$\begin{aligned}|III_T|_\infty &= O_P\left(\frac{p^{1/\kappa}}{T^{1/2-1/\kappa}}\vee\sqrt{\log p}\right)O_P\left(\frac{s_\alpha p^{1/\kappa}}{T^{1-1/\kappa}}\vee s_\alpha\sqrt{\frac{\log p}{T}}\right)\\ &= o_P(1).\end{aligned}$$

Lastly, under Assumption 2.5 (vi)

$$\begin{aligned}|IV_T|_\infty &\leq \max_{j\in G}|\mathbf{X}\hat{\Theta}_j^\top|_2\|\mathbf{m}-\mathbf{X}\beta\|_T\\ &= \max_{j\in G}\sqrt{\hat{\Theta}_j\hat{\Sigma}\hat{\Theta}_j^\top}\,o_P(1)\\ &= o_P(1),\end{aligned}$$

where the last line follows since $\hat{\Theta}_j$ is consistent in the ℓ_1 norm while $\hat{\Sigma}$ is consistent in the ℓ_∞ norm under maintained assumptions. □

Next, we focus on the HAC estimator based on LASSO residuals. Note that by construction of the precision matrix $\hat{\Theta}$, its j^{th} row is $\hat{\Theta}_j X_t = \hat{v}_{t,j}/\hat{\sigma}_j^2$, where $\hat{v}_{t,j}$ is the regression residual from the j^{th} nodewise LASSO regression and $\hat{\sigma}_j^2$ is the corresponding estimator of the variance of the regression error. Therefore, the HAC estimator based on the LASSO residuals in Eq. 3 can be written as

$$\hat{\Xi}_G = \sum_{|k|<T} K\left(\frac{k}{M_T}\right)\hat{\Gamma}_k,$$

where $\hat{\Gamma}_k$ has generic (j,h)-entry $\frac{1}{T}\sum_{t=1}^{T-k}\hat{u}_t\hat{u}_{t+k}\hat{v}_{t,j}\hat{v}_{t+k,h}\hat{\sigma}_j^{-2}\hat{\sigma}_h^{-2}$.

Similarly, we define

$$\tilde{\Xi}_G = \sum_{|k|<T} K\left(\frac{k}{M_T}\right)\tilde{\Gamma}_k,$$

where $\tilde{\Gamma}_k$ has generic (j,h)-entry $\frac{1}{T}\sum_{t=1}^{T-k}u_t u_{t+k}v_{t,j}v_{t+k,h}\sigma_j^{-2}\sigma_h^{-2}$ and note that the long-run variance Ξ_G has generic (j,h)-entry $\mathbb{E}[u_t u_{t+k}v_{t,j}v_{t+k,h}]\sigma_j^{-2}\sigma_h^{-2}$.

Assumption A.1.1. *Suppose that uniformly over $k \in \mathbf{Z}$ and $j, h \in G$ (i) $\mathbb{E}|u_0 u_k v_{0,j} v_{k,h}| < \infty$; (ii) $\mathbb{E}|v_{0,j} u_k v_{k,h}|^2 < \infty$, $\mathbb{E}|u_0 u_k v_{k,h}|^2 < \infty$, $\mathbb{E}|u_0 v_{0,j} u_k|^2 < \infty$, and $\mathbb{E}|u_0 v_{0,j} v_{k,h}|^2 < \infty$; (iii) $\mathbb{E}|u_0|^{2q} < \infty$ and $\mathbb{E}|v_{0,j}|^{2q} < \infty$ for some $q \geq 1$.*

Proof of Theorem 2.2. By Proposition A.1.3 with $V_t = (u_t v_{t,j}/\sigma_j^2)_{j \in G}$

$$\|\hat{\Xi}_G - \Xi_G\| \leq \|\hat{\Xi}_G - \tilde{\Xi}_G\| + O_P\left(\sqrt{\frac{M_T}{T}} + M_T^{-\varsigma} + T^{-(\varsigma \wedge 1)}\right). \tag{A.2}$$

Next,

$$\begin{aligned}
\|\hat{\Xi}_G - \tilde{\Xi}_G\| &\leq \sum_{|k|<T} \left|K\left(\frac{k}{M_T}\right)\right| \|\hat{\Gamma}_k - \tilde{\Gamma}_k\| \\
&\leq |G| \sum_{|k|<T} \left|K\left(\frac{k}{M_T}\right)\right| \max_{j,h \in G} \left|\frac{1}{\hat{\sigma}_j^2 \hat{\sigma}_h^2 T} \sum_{t=1}^{T-k} \hat{u}_t \hat{u}_{t+k} \hat{v}_{t,j} \hat{v}_{t+k,h} - \frac{1}{\sigma_j^2 \sigma_h^2 T} \sum_{t=1}^{T-k} u_t u_{t+k} v_{t,j} v_{t+k,h}\right| \\
&\leq |G| \sum_{|k|<T} \left|K\left(\frac{k}{M_T}\right)\right| \max_{j,h \in G} \frac{1}{\hat{\sigma}_j^2 \hat{\sigma}_h^2} \left|\frac{1}{T} \sum_{t=1}^{T-k} \hat{u}_t \hat{u}_{t+k} \hat{v}_{t,j} \hat{v}_{t+k,h} - \frac{1}{T} \sum_{t=1}^{T-k} u_t u_{t+k} v_{t,j} v_{t+k,h}\right| \\
&\quad + |G| \max_{j,h \in G} \left|\frac{1}{\hat{\sigma}_j^2 \hat{\sigma}_h^2} - \frac{1}{\sigma_j^2 \sigma_h^2}\right| \sum_{|k|<T} \left|K\left(\frac{k}{M_T}\right)\right| \left|\frac{1}{T} \sum_{t=1}^{T-k} u_t u_{t+k} v_{t,j} v_{t+k,h}\right| \\
&\triangleq S_T^a + S_T^b.
\end{aligned}$$

By Proposition A.1.1, since $s_\alpha \sqrt{\frac{\log p}{T}} = o(1)$ and $s_\alpha^\kappa p = o(T^{4\kappa/5-1})$, under stated assumptions, we obtain $\max_{j \in G} |\hat{\sigma}_j^2 - \sigma_j^2| = o_P(1)$, and whence $\max_{j \in G} \hat{\sigma}_j^{-2} = O_P(1)$. Using $\hat{a}\hat{b} - ab = (\hat{a} - a)b + a(\hat{b} - b) + (\hat{a} - a)(\hat{b} - b)$, by Proposition A.1.1

$$S_T^b = O_P\left(\frac{s_\alpha p^{1/\kappa}}{T^{1-1/\kappa}} \vee s_\alpha \sqrt{\frac{\log p}{T}}\right) \sum_{|k|<T} \left|K\left(\frac{k}{M_T}\right)\right| \max_{j,h \in G} \left|\frac{1}{T} \sum_{t=1}^{T-k} u_t u_{t+k} v_{t,j} v_{t+k,h}\right|$$

Under Assumption A.1.2 (i) and A.1.1 (i)

$$\begin{aligned}
\mathbb{E}\left[\sum_{|k|<T} \left|K\left(\frac{k}{M_T}\right)\right| \max_{j,h \in G} \left|\frac{1}{T} \sum_{t=1}^{T-k} u_t u_{t+k} v_{t,j} v_{t+k,h}\right|\right] &\leq O(M_T) \sup_{k \in \mathbf{Z}} \sum_{j,h \in G} \mathbb{E}\left|\frac{1}{T} \sum_{t=1}^{T-k} u_t u_{t+k} v_{t,j} v_{t+k,h}\right| \\
&\leq O(M_T) |G|^2 \sup_{k \in \mathbf{Z}} \max_{j,h \in G} \mathbb{E}|u_t u_{t+k} v_{t,j} v_{t+k,h}| \\
&= O(M_T),
\end{aligned}$$

and whence $S_T^b = O_P\left(M_T\left(\frac{s_\alpha p^{1/\kappa}}{T^{1-1/\kappa}} \vee s_\alpha \sqrt{\frac{\log p}{T}}\right)\right)$.

Next, we evaluate uniformly over $|k| < T$

$$\begin{aligned}
&\left|\frac{1}{T}\sum_{t=1}^{T-k}\hat{u}_t\hat{u}_{t+k}\hat{v}_{t,j}\hat{v}_{t+k,h} - \frac{1}{T}\sum_{t=1}^{T-k}u_tu_{t+k}v_{t,j}v_{t+k,h}\right| \\
&\leq \left|\frac{1}{T}\sum_{t=1}^{T-k}(\hat{u}_t\hat{v}_{t,j} - u_tv_{t,j})u_{t+k}v_{t+k,h}\right| + \left|\frac{1}{T}\sum_{t=1}^{T-k}u_tv_{t,j}(\hat{u}_{t+k}\hat{v}_{t+k,h} - u_{t+k}v_{t+k,h})\right| \\
&\quad + \left|\frac{1}{T}\sum_{t=1}^{T-k}(\hat{u}_t\hat{v}_{t,j} - u_tv_{t,j})(\hat{u}_{t+k}\hat{v}_{t+k,h} - u_{t+k}v_{t+k,h})\right| \triangleq I_T + II_T + III_T.
\end{aligned}$$

We bound the first term as

$$\begin{aligned}
I_T &\leq \left|\frac{1}{T}\sum_{t=1}^{T-k}(\hat{u}_t - u_t)v_{t,j}u_{t+k}v_{t+k,h}\right| + \left|\frac{1}{T}\sum_{t=1}^{T-k}u_t(\hat{v}_{t,j} - v_{t,j})u_{t+k}v_{t+k,h}\right| \\
&\quad + \left|\frac{1}{T}\sum_{t=1}^{T-k}(\hat{u}_t - u_t)(\hat{v}_{t,j} - v_{t,j})u_{t+k}v_{t+k,h}\right| \triangleq I_T^a + I_T^b + I_T^c.
\end{aligned}$$

By the Cauchy-Schwartz inequality, under Assumptions of Theorem A.1 for the sg-LASSO and Assumption A.1.1 (ii)

$$\begin{aligned}
I_T^a &= \left|\frac{1}{T}\sum_{t=1}^{T-k}\left(X_t^\top(\beta - \hat{\beta}) + m_t - X_t^\top\beta\right)v_{t,j}u_{t+k}v_{t+k,h}\right| \\
&\leq (\|\mathbf{X}(\hat{\beta} - \beta)\|_T + \|\mathbf{m} - \mathbf{X}\beta\|_T)\sqrt{\frac{1}{T}\sum_{t=1}^{T-k}v_{t,j}^2u_{t+k}^2v_{t+k,h}^2} \\
&= O_P\left(\frac{s_\alpha p^{1/\kappa}}{T^{1-1/\kappa}} \vee \sqrt{\frac{s_\alpha \log p}{T}}\right).
\end{aligned}$$

Similarly, under Assumptions of Theorem A.1 for the nodewise LASSO and Assumption A.1.1 (ii)

$$I_T^b \leq \left(\|\mathbf{X}_{-j}(\hat{\gamma}_j - \gamma_j)\|_T + o_P(T^{-1/2})\right)\sqrt{\frac{1}{T}\sum_{t=1}^{T-k}u_t^2u_{t+k}^2v_{t+k,h}^2} = O_P\left(\frac{S_jp^{1/\kappa}}{T^{1-1/\kappa}} \vee \sqrt{\frac{S_j\log p}{T}}\right).$$

Note that for arbitrary $(\xi_t)_{t\in\mathbf{Z}}$ and $q \geq 1$, by Jensen's inequality

$$\mathbb{E}\left[\max_{t\in[T]}|\xi_t|\right] \leq \left(\mathbb{E}\left[\max_{t\in[T]}|\xi_t|^q\right]\right)^{1/q} \leq \left(\mathbb{E}\left[\sum_{t=1}^{T}|\xi_t|^q\right]\right)^{1/q} = T^{1/q}\left(\mathbb{E}|\xi_t|^q\right)^{1/q}.$$

Then by the Cauchy-Schwartz inequality under Assumption A.1.1 (iii) and Theorem A.1

$$\begin{aligned}
I_T^c &\leq (\|\mathbf{X}(\hat{\beta} - \beta)\|_T + o_P(T^{-1/2}))(\|\mathbf{X}_{-j}(\hat{\gamma}_j - \gamma_j)\|_T + o_P(T^{-1/2}))\max_{t\in[T]}|u_tv_{t,h}| \\
&= O_P\left(\frac{s^2p^{2/\kappa}}{T^{2-3/\kappa}} \vee \frac{s\log p}{T^{1-1/\kappa}}\right),
\end{aligned}$$

where we use the fact that $\kappa \leq q$. Therefore, under maintained assumptions

$$I_T = O_P\left(\frac{sp^{1/\kappa}}{T^{1-1/\kappa}} \vee \sqrt{\frac{s\log p}{T}} + \frac{s^2p^{2/\kappa}}{T^{2-3/\kappa}} \vee \frac{s\log p}{T^{1-1/\kappa}}\right)$$

and by symmetry

$$II_T = O_P\left(\frac{sp^{1/\kappa}}{T^{1-1/\kappa}} \vee \sqrt{\frac{s\log p}{T}} + \frac{s^2p^{2/\kappa}}{T^{2-3/\kappa}} \vee \frac{s\log p}{T^{1-1/\kappa}}\right).$$

Lastly, by the Cauchy-Schwartz inequality

$$\begin{aligned} III_T &\leq \sqrt{\frac{1}{T}\sum_{t=1}^{T-k}(\hat{u}_t\hat{v}_{t,j} - u_tv_{t,j})^2\frac{1}{T}\sum_{t=1}^{T-k}(\hat{u}_{t+k}\hat{v}_{t+k,h} - u_{t+k}v_{t+k,h})^2} \\ &\leq \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t\hat{v}_{t,j} - u_tv_{t,j})^2\frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t\hat{v}_{t,h} - u_tv_{t,h})^2}. \end{aligned}$$

For each $j \in G$

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t\hat{v}_{t,j} - u_tv_{t,j})^2 &\leq \frac{3}{T}\sum_{t=1}^{T}|\hat{u}_t - u_t|^2v_{t,j}^2 + \frac{3}{T}\sum_{t=1}^{T}|\hat{v}_{t,j} - v_{t,j}|^2u_t^2 \\ &\quad + \frac{3}{T}\sum_{t=1}^{T}|\hat{u}_t - u_t|^2|\hat{v}_{t,j} - v_{t,j}|^2 \\ &\triangleq III_T^a + III_T^b + III_T^c. \end{aligned}$$

Since under Assumption A.1.1 (iii), $\mathbb{E}|v_{t,j}|^{2q} < \infty$ and $\mathbb{E}|u_t|^{2q} < \infty$,

$$III_T^a \leq 3\max_{t\in[T]}|v_{t,j}|^2(\|\mathbf{X}(\hat{\beta} - \beta)\|_T^2 + o_P(T^{-1/2})) = O_P\left(\frac{s_\alpha p^{2/\kappa}}{T^{2-3/\kappa}} \vee \frac{s_\alpha \log p}{T^{1-1/\kappa}}\right)$$

and

$$III_T^b \leq 3\max_{t\in[T]}|u_t|^2(\|\mathbf{X}_{-j}(\hat{\gamma}_j - \gamma_j)\|_T^2 + o_P(T^{-1/2})) = O_P\left(\frac{S_jp^{2/\kappa}}{T^{2-3/\kappa}} \vee \frac{S_j\log p}{T^{1-1/\kappa}}\right).$$

For the last term, since under Assumption 2.1 (ii), $\sup_k \mathbb{E}|X_{t,k}|^{2\tilde{q}} < \infty$ and $\kappa \geq \tilde{q}$, by

Theorem A.1

$$
\begin{aligned}
III_T^c &\leq 3(\|\mathbf{X}(\hat{\beta}-\beta)\|_T^2 + o_P(T^{-1/2}))\max_{t\in[T]}|X_{t,-j}^\top(\hat{\gamma}_j-\gamma_j)+m_t-X_t^\top\beta|^2 \\
&\leq O_P\left(\frac{s_\alpha p^{2/\kappa}}{T^{2-2/\kappa}}\vee\frac{s_\alpha\log p}{T}\right)\left(2\max_{t\in[T]}|X_t|_\infty^2|\hat{\gamma}_j-\gamma_j|_1^2+2T\|\mathbf{m}-\mathbf{X}^\top\beta\|_T^2\right) \\
&= O_P\left(\left(\frac{s_\alpha p^{2/\kappa}}{T^{2-2/\kappa}}\vee\frac{s_\alpha\log p}{T}\right)\left(\frac{S^2p^{2/\kappa}}{T^{2-2/\kappa}}\vee S^2\frac{\log p}{T}\right)(pT)^{1/\kappa}\right) \\
&= O_P\left(\frac{s^3p^{5/\kappa}}{T^{4-5/\kappa}}+\frac{s^3p^{3/\kappa}\log p}{T^{3-3/\kappa}}+\frac{s^3p^{1/\kappa}\log^2 p}{T^{2-1/\kappa}}\right) \\
&= O_P\left(\frac{s^3p^{5/\kappa}}{T^{4-5/\kappa}}+\frac{s^3p^{3/\kappa}\log p}{T^{3-3/\kappa}}\right),
\end{aligned}
$$

where we use the fact that $\kappa > 2$, $s = s_\alpha \vee S$, $s^\kappa p = o(T^{4\kappa/5-1})$, and $\frac{s^2\log p}{T} = o(1)$ as $T\to\infty$. Then for every $j \in G$

$$
\frac{1}{T}\sum_{t=1}^T(\hat{u}_t\hat{v}_{t,j}-u_tv_{t,j})^2 = O_P\left(\frac{sp^{2/\kappa}}{T^{2-3/\kappa}}\vee\frac{s\log p}{T^{1-1/\kappa}}+\frac{s^3p^{5/\kappa}}{T^{4-5/\kappa}}+\frac{s^3p^{3/\kappa}\log p}{T^{3-3/\kappa}}\right),
$$

and whence

$$
III_T = O_P\left(\frac{sp^{2/\kappa}}{T^{2-3/\kappa}}\vee\frac{s\log p}{T^{1-1/\kappa}}+\frac{s^3p^{5/\kappa}}{T^{4-5/\kappa}}+\frac{s^3p^{3/\kappa}\log p}{T^{3-3/\kappa}}\right).
$$

Therefore, since $\hat{\sigma}_j^2 \xrightarrow{P} \sigma_j^2$, we obtain

$$
\begin{aligned}
S_T^a &= O_P\left(M_T\left(\frac{sp^{1/\kappa}}{T^{1-1/\kappa}}\vee\sqrt{\frac{s\log p}{T}}+\frac{s^2p^{2/\kappa}}{T^{2-3/\kappa}}\vee\frac{s\log p}{T^{1-1/\kappa}}+\frac{s^3p^{5/\kappa}}{T^{4-5/\kappa}}+\frac{s^3p^{3/\kappa}\log p}{T^{3-3/\kappa}}\right)\right) \\
&= O_P\left(M_T\left(\frac{sp^{1/\kappa}}{T^{1-1/\kappa}}\vee\sqrt{\frac{s\log p}{T}}+\frac{s^2p^{2/\kappa}}{T^{2-3/\kappa}}+\frac{s^3p^{5/\kappa}}{T^{4-5/\kappa}}\right)\right),
\end{aligned}
$$

where the last line follows since $s^\kappa p/T^{4\kappa/5-1} = o(1)$. Combining this estimate with previously obtained estimate for S_T^b

$$
\|\hat{\Xi}_G-\tilde{\Xi}_G\| = O_P\left(M_T\left(\frac{sp^{1/\kappa}}{T^{1-1/\kappa}}\vee s\sqrt{\frac{\log p}{T}}+\frac{s^2p^{2/\kappa}}{T^{2-3/\kappa}}+\frac{s^3p^{5/\kappa}}{T^{4-5/\kappa}}\right)\right).
$$

The result follows from combining this estimate with the estimate in equation (A.2). □

Proof of Theorem 3.1. Suppose first that $p = 1$. For $a \in \mathbf{R}$, with some abuse of notation, let $[a]$ denote its integer part. We split partial sums into blocks $V_k = \xi_{(k-1)J+1}+\cdots+\xi_{kJ}, k = 1,2,\ldots,[T/J]$ and $V_{[T/J]+1} = \xi_{[T/J]J+1}+\cdots+\xi_T$, where we set $V_{[T/J]+1} = 0$ if T/J is an integer. Let $\{U_t : t = 1, 2, \ldots, [T/J] + 1\}$ be i.i.d. random variables drawn from the uniform distribution on $(0, 1)$ independently of $\{V_t : t = 1, 2, \ldots, [T/J] + 1\}$. Put $\mathcal{M}_t = \sigma(V_1, \ldots, V_{t-2})$ for every $t = 3, \ldots, [T/J] + 1$. Next, for $t = 1, 2$, set $V_t^* = V_t$, while for $t \geq 3$, by Dedecker and Prieur (2004), Lemma 5, there exist random variables $V_t^* =_d V_t$ such that:

1. V_t^* is $\sigma(V_1, \ldots, V_{t-2}) \vee \sigma(V_t) \vee \sigma(U_t)$-measurable;
2. $V_t^* \perp\!\!\!\perp (V_1, \ldots, V_{t-2})$;
3. $\|V_t - V_t^*\|_1 = \tau(\mathcal{M}_t, V_t)$.

It follows from properties 1. and 2. that $(V_{2t}^*)_{t\geq 1}$ and $(V_{2t-1}^*)_{t\geq 1}$ are sequences of independent random variables. Then

$$\begin{aligned}\left|\sum_{t=1}^{T} \xi_t\right| &\leq \left|\sum_{t\geq 1} V_{2t}^*\right| + \left|\sum_{t\geq 1} V_{2t-1}^*\right| + \sum_{t=3}^{[T/J]+1} |V_t - V_t^*| \\ &\triangleq I_T + II_T + III_T.\end{aligned}$$

By Fuk and Nagaev (1971), Corollary 4, there exist constants $c_q^{(j)}, j = 1, 2$ such that

$$\begin{aligned}\Pr(I_T \geq x) &\leq \frac{c_q^{(1)}}{x^q} \sum_{t\geq 1} \mathbb{E}|V_{2t}^*|^q + 2\exp\left(-\frac{c_q^{(2)} x^2}{\sum_{t\geq 1} \operatorname{Var}(V_{2t}^*)}\right) \\ &\leq \frac{c_q^{(1)}}{x^q} \sum_{t\geq 1} \mathbb{E}|V_{2t}|^q + 2\exp\left(-\frac{c_q^{(2)} x^2}{B_T^2}\right),\end{aligned}$$

where the second inequality follows since $\sum_{t\geq 1} \operatorname{Var}(V_{2t}^*) = \sum_{t\geq 1} \operatorname{Var}(V_{2t}) \leq B_T^2$. Similarly

$$\Pr(II_T \geq x) \leq \frac{c_q^{(1)}}{x^q} \sum_{t\geq 1} \mathbb{E}|V_{2t-1}|^q + 2\exp\left(-\frac{c_q^{(2)} x^2}{B_T^2}\right).$$

Lastly, since $\mathcal{M}_t$ and V_t are separated by $J + 1$ lags of $(\xi_t)_{t\geq 1}$, we have $\tau(\mathcal{M}_t, V_t) \leq J\tau_J(J + 1)$. Then by Markov's inequality and property 3.

$$\begin{aligned}\Pr(III_T \geq x) &\leq \frac{1}{x} \sum_{t=3}^{[T/J]+1} \tau(\mathcal{M}_t, V_t) \\ &\leq \frac{T}{x} \tau_{J+1}.\end{aligned}$$

Combining all the estimates together

$$\begin{aligned}\Pr\left(\left|\sum_{t=1}^T \xi_t\right| \geq 3x\right) &\leq \Pr(I_T \geq x) + \Pr(II_T \geq x) + \Pr(III_T \geq x) \\ &\leq \frac{c_q^{(1)}}{x^q}\sum_{t=1}^{[T/J]+1}\mathbb{E}|V_t|^q + 4\exp\left(-\frac{c_q^{(2)}x^2}{B_T^2}\right) + \frac{T}{x}\tau_{J+1} \\ &\leq \frac{c_q^{(1)}}{x^q}J^{q-1}\sum_{t=1}^T\|\xi_t\|_q^q + \frac{T}{x}c(J+1)^{-a} + 4\exp\left(-\frac{c_q^{(2)}x^2}{B_T^2}\right).\end{aligned}$$

To balance the first two terms, we shall set $J \sim x^{\frac{q-1}{q+a-1}}$, in which case we obtain the result under maintained assumptions. The result for $p > 1$ follows by the union bound. □

The following covariance inequality follows from Dedecker and Doukhan (2003) and Dedecker and Prieur (2004) and provides a sharp way to control autocovariances with τ-dependence coefficients. Note that τ-dependence coefficients can in turn be controlled by the α-mixing coefficient, see Dedecker and Prieur (2004), Lemma 6 and Remark 2.

Lemma A.1.1. *Let $(\xi_t)_{t\in\mathbf{Z}}$ be a centered stationary stochastic process with $\|\xi_0\|_q < \infty$ for some $q > 2$. Then*

$$|\mathrm{Cov}(\xi_0, \xi_t)| \leq \gamma_t^{\frac{q-2}{q-1}}\|\xi_0\|_q^{q/(q-1)}$$

and

$$\gamma_t \leq \tau_t,$$

where $\gamma_t = \|\mathbb{E}(\xi_t|\mathcal{M}_0) - \mathbb{E}(\xi_t)\|_1$ is the L_1 mixingale coefficient with respect to the canonical filtration $\mathcal{M}_0 = \sigma(\xi_0, \xi_{-1}, \xi_{-2}, \dots)$.

Proof. Let Q be the quantile function of $|\xi_0|$ and let G be the inverse of $x \mapsto \int_0^x Q(u)\mathrm{d}u$. By Dedecker and Doukhan (2003), Proposition 1

$$\begin{aligned}|\mathrm{Cov}(\xi_0, \xi_t)| &\leq \int_0^{\gamma_t}(Q\circ G)(u)\mathrm{d}u \\ &\leq \gamma_t^{\frac{q-2}{q-1}}\left(\int_0^{\|\xi_0\|_1}(Q\circ G)^{q-1}(u)\mathrm{d}u\right)^{1/(q-1)} \\ &= \gamma_t^{\frac{q-2}{q-1}}\|\xi_0\|_q^{q/(q-1)},\end{aligned}$$

where the second line follows by Hölder's inequality and the last equality by the change of variables $\int_0^{\|\xi_0\|_1}(Q\circ G)^{q-1}(u)\mathrm{d}u = \int_0^1 Q^q(u)\mathrm{d}u = \mathbb{E}|\xi_0|^q$. The second statement follows from Dedecker and Doukhan (2003), Lemma 1 and Dedecker and Prieur (2004), Lemma 6. □

Lemma A.1.2. *Let $(\xi_t)_{t\in\mathbf{Z}}$ be a centered stationary stochastic process such that $\|\xi_t\|_q < \infty$ for some $q > 2$ and $\tau_k = O(k^{-a})$ for some $a > \frac{q-1}{q-2}$. Then*

$$\sum_{t=1}^{T}\sum_{k=1}^{T}|\mathrm{Cov}(\xi_{t,j},\xi_{k,j})| = O(T).$$

Proof. Under stationarity

$$\begin{aligned}\sum_{t=1}^{T}\sum_{k=1}^{T}|\mathrm{Cov}(\xi_{t,j},\xi_{k,j})| &= T\mathrm{Var}(\xi_0) + 2\sum_{k=1}^{T-1}(T-k)\,\mathrm{Cov}(\xi_0,\xi_k)\\ &\leq T\mathrm{Var}(\xi_0) + 2T\|\xi_t\|_q^{q/(q-1)}\sum_{k=1}^{T-1}\tau_k^{\frac{q-2}{q-1}}\\ &= O(T),\end{aligned}$$

where the second line follows by Proposition A.1.1 and the last since the series $\sum_{k=1}^{\infty}k^{-a\frac{q-2}{q-1}}$ converges under the maintained assumptions. □

SUPPLEMENTARY MATERIAL

Additional results and proofs: This file contains supplementary results with proofs.

We recall first the convergence rates for the sg-LASSO with weakly dependent data that will be needed throughout the paper from Babii, Ghysels, and Striaukas (2020), Corollary 3.1.

Theorem A.1. *Suppose that Assumptions 2.1, 2.2, 2.3, and 2.4 are satisfied. Then*

$$\|\mathbf{X}(\hat{\beta}-\beta)\|_T^2 = O_P\left(\frac{s_\alpha p^{2/\kappa}}{T^{2-2/\kappa}} \vee \frac{s_\alpha \log p}{T}\right).$$

and

$$\Omega(\hat{\beta}-\beta) = O_P\left(\frac{s_\alpha p^{1/\kappa}}{T^{1-1/\kappa}} \vee s_\alpha\sqrt{\frac{\log p}{T}}\right).$$

Next, we consider the regularized estimator of the variance of the regression error

$$\hat{\sigma}^2 = \|\mathbf{y}-\mathbf{X}\hat{\beta}\|_T^2 + \lambda\Omega(\hat{\beta}),$$

where $\hat{\beta}$ is the sg-LASSO estimator. While the regularization is not needed to have a consistent variance estimator, the LASSO version of the regularized estimator ($\alpha = 1$) is needed to establish the CLT for the debiased sg-LASSO estimator. The following result describes the converges of this variance estimator to its population counterpart $\sigma^2 = \mathbb{E}\|\mathbf{u}\|_T^2$.

Proposition A.1.1. *Suppose that Assumptions 2.1, 2.2, 2.3, and 2.4 are satisfied and that $(u_t^2)_{t\in\mathbf{Z}}$ has a finite long run variance. Then*

$$\hat{\sigma}^2 = \sigma^2 + O_P\left(\frac{s_\alpha p^{1/\kappa}}{T^{1-1/\kappa}} \vee s_\alpha\sqrt{\frac{\log p}{T}}\right)$$

provided that $\frac{s_\alpha p^{1/\kappa}}{T^{1-1/\kappa}} \vee s_\alpha\sqrt{\frac{\log p}{T}} = o(1)$.

Proof. We have

$$\begin{aligned}
|\hat{\sigma}^2-\sigma^2| &= \left|\|\mathbf{u}\|_T^2 + 2\langle \mathbf{u}, \mathbf{m}-\mathbf{X}\hat{\beta}\rangle_T - \|\mathbf{m}-\mathbf{X}\hat{\beta}\|_T^2 + \lambda\Omega(\hat{\beta}) - \sigma^2\right| \\
&\leq |\sigma^2 - \|\mathbf{u}\|_T^2| + 2\|\mathbf{u}\|_T\|\mathbf{m}-\mathbf{X}\hat{\beta}\|_T + 2\|\mathbf{X}(\hat{\beta}-\beta)\|_T^2 + 2\|\mathbf{m}-\mathbf{X}\beta\|_T^2 + \lambda\Omega(\hat{\beta}) \\
&\triangleq I_T + II_T + III_T + IV_T + V_T.
\end{aligned}$$

By the Chebychev's inequality since the long-run variance exists, for every $\varepsilon > 0$

$$\Pr\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2 - \sigma^2)\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\sum_{t\in\mathbf{Z}}\mathrm{Cov}(u_0^2, u_t^2),$$

whence $I_T = O_P\left(\frac{1}{\sqrt{T}}\right)$. Therefore, by the triangle inequality and Theorem A.1

$$\begin{aligned} II_T &= O_P(1)\|\mathbf{m} - \mathbf{X}\hat{\beta}\|_T \\ &\leq O_P(1)\left(\|\mathbf{m} - \mathbf{X}\beta\|_T + \|\mathbf{X}(\hat{\beta} - \beta)\|_T\right) = O_P\left(s_\alpha^{1/2}\lambda + \frac{s_\alpha^{1/2}p^{1/\kappa}}{T^{1-1/\kappa}} \vee \sqrt{\frac{s_\alpha \log p}{T}}\right). \end{aligned}$$

By Theorem A.1 we also have

$$III_T + IV_T = O_P\left(\frac{s_\alpha p^{2/\kappa}}{T^{2-2/\kappa}} \vee \frac{s_\alpha \log p}{T} + s_\alpha\lambda^2\right).$$

Lastly, another application of Theorem A.1 gives

$$\begin{aligned} V_T &= \lambda\Omega(\hat{\beta} - \beta) + \lambda\Omega(\beta) \\ &= O_P\left(\lambda\left(\frac{s_\alpha p^{1/\kappa}}{T^{1-1/\kappa}} \vee s_\alpha\sqrt{\frac{\log p}{T}}\right) + \lambda s_\alpha\right). \end{aligned}$$

The result follows from combining all estimates together. □

Next, we look at the estimator of the precision matrix. Consider nodewise LASSO regressions in equation (2) for each $j \in [p]$. Put $S = \max_{j\in G} S_j$, where S_j is the support of γ_j.

Proposition A.1.2. *Suppose that Assumptions 2.1, 2.2, 2.3, and 2.4 are satisfied for each nodewise regression $j \in G$ and that $(V_{t,j}^2)_{t\in\mathbf{Z}}$ has a finite long-run variance for each $j \in G$. Then if $S^\kappa p = o(T^{\kappa-1})$ and $S = o(T^{1/2}/\log^{1/2} p)$*

$$\|\hat{\Theta}_G - \Theta_G\|_\infty = O_P\left(\frac{Sp^{1/\kappa}}{T^{1-1/\kappa}} \vee S\sqrt{\frac{\log p}{T}}\right)$$

and

$$\max_{j\in G}|(I - \hat{\Theta}\hat{\Sigma})_j|_\infty = O_P\left(\frac{p^{1/\kappa}}{T^{1-1/\kappa}} \vee \sqrt{\frac{\log p}{T}}\right).$$

Proof. By Theorem A.1 and Proposition A.1.1 with $\alpha = 1$ (corresponding to the LASSO estimator of γ_j and σ_j^2)

$$\begin{aligned}\|\hat{\Theta}_G - \Theta_G\|_\infty &= \max_{j\in G} |\hat{\Theta}_j - \Theta_j|_1 \\ &\leq \max_{j\in G}\left\{|\hat{\gamma}_j|_1 \left|\hat{\sigma}_j^{-2} - \sigma_j^{-2}\right| + |\hat{\gamma}_j - \gamma_j|_1 |\sigma_j^{-2}|\right\} \\ &= O_P\left(\frac{Sp^{1/\kappa}}{T^{1-1/\kappa}} \vee S\sqrt{\frac{\log p}{T}}\right),\end{aligned}$$

where we use the fact that $|G|$ is fixed and that $\hat{\sigma}_j^2 \xrightarrow{p} \sigma_j^2$.

Second, for each $j \in G$, by Fermat's rule,

$$\mathbf{X}_{-j}^\top(\mathbf{X}_j - \mathbf{X}_{-j}\hat{\gamma}_j)/T = \lambda_j z^*, \qquad z^* \in \partial|\hat{\gamma}_j|_1,$$

where $\hat{\gamma}_j^\top z^* = |\hat{\gamma}_j|_1$ and $|z^*|_\infty \leq 1$. Then

$$\begin{aligned}\mathbf{X}_j^\top(\mathbf{X}_j - \mathbf{X}_{-j}\hat{\gamma}_j)/T &= \|\mathbf{X}_j - \mathbf{X}_{-j}\hat{\gamma}_j\|_T^2 + \hat{\gamma}_j^\top \mathbf{X}_{-j}^\top(\mathbf{X}_j - \mathbf{X}_{-j}\hat{\gamma}_j)/T \\ &= \|\mathbf{X}_j - \mathbf{X}_{-j}\hat{\gamma}_j\|_T^2 + \lambda_j \hat{\gamma}_j^\top z^* = \hat{\sigma}_j^2,\end{aligned}$$

and whence

$$\begin{aligned}|(I - \hat{\Theta}\hat{\Sigma})_j|_\infty &= |I_j - (\mathbf{X}_j - \mathbf{X}_{-j}\hat{\gamma}_j)^\top \mathbf{X}/(T\hat{\sigma}_j^2)|_\infty \\ &= \max\left\{|1 - \mathbf{X}_j^\top(\mathbf{X}_j - \mathbf{X}_{-j}\hat{\gamma}_j)/(T\hat{\sigma}_j^2)|, |\mathbf{X}_{-j}^\top(\mathbf{X}_j - \mathbf{X}_{-j}\hat{\gamma}_j)/(T\hat{\sigma}_j^2)|_\infty\right\} \\ &= \lambda_j |z^*|_\infty / \hat{\sigma}_j^2 = O_P\left(\frac{p^{1/\kappa}}{T^{1-1/\kappa}} \vee \sqrt{\frac{\log p}{T}}\right),\end{aligned}$$

where the last line follows since $\hat{\sigma}_j^{-2} = O_P(1)$ and $|z^*|_\infty \leq 1$. The conclusion follows from the fact that $|G|$ is fixed. □

Next, we first derive the non-asymptotic Frobenius norm bound with explicit constants for a generic HAC estimator of the sample mean that holds uniformly over a class of distributions. We focus on the p-dimensional centered stochastic process $(V_t)_{t\in\mathbf{Z}}$ and put

$$\Xi = \sum_{k\in\mathbf{Z}} \Gamma_k \qquad \text{and} \qquad \tilde{\Xi} = \sum_{|k|<T} K\left(\frac{k}{M_T}\right)\tilde{\Gamma}_k,$$

where $\Gamma_k = \mathbb{E}[V_t V_{t+k}^\top]$ and $\tilde{\Gamma}_k = \frac{1}{T}\sum_{t=1}^{T-k} V_t V_{t+k}^\top$. Put also $\Gamma = (\Gamma_k)_{k\in\mathbf{Z}}$ and let $\langle ., . \rangle$ be the Frobenius inner product with corresponding Frobenius norm $\|.\|$. The following assumption describes the relevant class of distributions and kernel functions.

Assumption A.1.2. *Suppose that (i) $K : \mathbf{R} \to [-1, 1]$ is a Riemann integrable function such that $K(0) = 1$; (ii) there exists some $\varepsilon, \varsigma > 0$ such that $|K(0) - K(x)| \leq L|x|^{\varsigma}$ for all $|x| < \varepsilon$; (iii) $(V_t)_{t\in\mathbf{Z}}$ is fourth-order stationary; (iv) $\Gamma \in \mathcal{G}(\varsigma, B_1, B_2)$, where*

$$\mathcal{G}(\varsigma, B_1, B_2) = \left\{ \sum_{k\in\mathbf{Z}} |k|^{\varsigma} \|\Gamma_k\| \leq B_1, \quad \sup_{k\in\mathbf{Z}} \sum_{l\in\mathbf{Z}} \sum_{t\in\mathbf{Z}} \sum_{j,h\in[p]} |\mathrm{Cov}(V_{0,j}V_{k,h}, V_{t,j}V_{t+l,h})| \leq B_2 \right\}$$

for some $B_1, B_2 > 0$.

Condition (ii) describes the smoothness (or order) of the kernel in the neighborhood of zero. $\varsigma = 1$ for the Bartlett kernel and $\varsigma = 2$ for the Parzen, Tukey-Hanning, and Quadratic spectral kernels, see Andrews (1991). Since the bias of the HAC estimator is limited by the order of the kernel, it is typically not recommended to use the Bartlett kernel in practice. Higher-order kernels with $\varsigma > 2$ do not ensure the positive definiteness of the HAC estimator and require additional spectral regularization, see Politis (2011). Condition (iv) describes the class of autocovariances that vanish rapidly enough. Note that if (iv) holds for some $\bar{\varsigma}$, then it also holds for every $\varsigma < \bar{\varsigma}$ and that if (ii) holds for some $\tilde{\varsigma} > \varsigma$, then it also holds for $\tilde{\varsigma} = \varsigma$. The covariance condition in (iv) can be justified under more primitive moment and summability conditions imposed on L_1-mixingale/τ-dependence coefficients, see Proposition A.1.1 and Andrews (1991), Lemma 1. The following result gives a nonasymptotic risk bound uniformly over the class $\mathcal{G}$ and corresponds to the asymptotic convergence rates for the spectral density evaluated at zero derived in Parzen (1957).

Proposition A.1.3. *Suppose that Assumption A.1.2 is satisfied. Then*

$$\sup_{\Gamma\in\mathcal{G}(\varsigma,B_1,B_2)} \mathbb{E}\|\tilde{\Xi} - \Xi\|^2 \leq C_1 \frac{M_T}{T} + C_2 M_T^{-2\varsigma} + C_3 T^{-2(\varsigma\wedge 1)},$$

where $C_1 = B_2 \left(\int |K(u)| \mathrm{d}u + o(1) \right)$, $C_2 = 2\left(B_1 L + \frac{2B_1}{\varepsilon^{\varsigma}}\right)^2$, and $C_3 = 2B_1^2$.

Proof. By the triangle inequality, under Assumption A.1.2 (i)

$$\begin{aligned}
\|\mathbb{E}[\tilde{\Xi}] - \Xi\| &= \left\| \sum_{|k|<T} K\left(\frac{k}{M_T}\right) \frac{T-k}{T} \Gamma_k - \sum_{k\in\mathbf{Z}} \Gamma_k \right\| \\
&\leq \sum_{|k|<T} \left| K\left(\frac{k}{M_T}\right) - K(0) \right| \|\Gamma_k\| + \frac{1}{T} \sum_{|k|<T} |k| \|\Gamma_k\| + \sum_{|k|\geq T} \|\Gamma_k\| \\
&\triangleq I_T + II_T + III_T.
\end{aligned}$$

For the first term, we obtain

$$
\begin{aligned}
I_T &= \sum_{|k|<\varepsilon M_T} \left| K(0) - K\left(\frac{k}{M_T}\right)\right| \|\Gamma_k\| + \sum_{\varepsilon M_T \leq |k| < T} \left| K\left(\frac{k}{M_T}\right) - K(0)\right| \|\Gamma_k\| \\
&\leq L M_T^{-\varsigma} \sum_{|k|<\varepsilon M_T} |k|^{\varsigma} \|\Gamma_k\| + 2 \sum_{\varepsilon M_T \leq |k| < T} \|\Gamma_k\| \\
&\leq \frac{B_1 L}{M_T^{\varsigma}} + \frac{2}{\varepsilon^{\varsigma} M_T^{\varsigma}} \sum_{\varepsilon M_T \leq |k| < T} |k|^{\varsigma} \|\Gamma_k\| \\
&\leq \frac{B_1 L}{M_T^{\varsigma}} + \frac{2 B_1}{\varepsilon^{\varsigma} M_T^{\varsigma}},
\end{aligned}
$$

where the second sum is defined to be zero if $T \leq \varepsilon M_T$, the second line follows under Assumption A.1.2 (i)-(ii) and the last two under Assumption A.1.2 (iii). Next, if $\varsigma \geq 1$,

$$
\sum_{|k|<T} |k| \|\Gamma_k\| \leq \sum_{|k|<T} |k|^{\varsigma} \|\Gamma_k\|,
$$

while if $\varsigma \in (0,1)$

$$
\sum_{|k|<T} |k| \|\Gamma_k\| \leq T^{1-\varsigma} \sum_{|k|<T} |k|^{\varsigma} \|\Gamma_k\|.
$$

Therefore, since $\sum_{|k|\geq T} \|\Gamma_k\| \leq T^{-\varsigma} \sum_{|k|\geq T} |k|^{\varsigma} \|\Gamma_k\|$, under Assumption A.1.2 (iv)

$$
\begin{aligned}
II_T + III_T &\leq \begin{cases} \frac{B_1}{T} & \varsigma \geq 1 \\ \frac{B_1}{T^{\varsigma}} & \varsigma \in (0,1) \end{cases} \\
&= \frac{B_1}{T^{\varsigma \wedge 1}}.
\end{aligned}
$$

This shows that

$$
\|\mathbb{E}[\tilde{\Xi}] - \Xi\| \leq \frac{B_1 L}{M_T^{\varsigma}} + \frac{2 B_1}{\varepsilon^{\varsigma} M_T^{\varsigma}} + \frac{B_1}{T^{\varsigma \wedge 1}}. \tag{A.3}
$$

Next, under Assumption A.1.2 (i)

$$
\begin{aligned}
\mathbb{E}\|\tilde{\Xi} - \mathbb{E}[\tilde{\Xi}]\|^2 &= \sum_{|k|<T} \sum_{|l|<T} K\left(\frac{k}{M_T}\right) K\left(\frac{l}{M_T}\right) \mathbb{E}\left\langle \tilde{\Gamma}_k - \mathbb{E}\tilde{\Gamma}_k, \tilde{\Gamma}_l - \mathbb{E}\tilde{\Gamma}_l \right\rangle \\
&\leq \sum_{|k|<T} \left| K\left(\frac{k}{M_T}\right)\right| \sup_{|k|<T} \sum_{|l|<T} \left| \mathbb{E}\left\langle \tilde{\Gamma}_k - \mathbb{E}\tilde{\Gamma}_k, \tilde{\Gamma}_l - \mathbb{E}\tilde{\Gamma}_l \right\rangle \right|,
\end{aligned}
$$

where under Assumptions A.1.2 (iii)

$$
\begin{aligned}
T \left| \mathbb{E}\left\langle \tilde{\Gamma}_k - \mathbb{E}\tilde{\Gamma}_k, \tilde{\Gamma}_l - \mathbb{E}\tilde{\Gamma}_l \right\rangle \right| &\leq \frac{1}{T} \sum_{t=1}^{T-k} \sum_{r=1}^{T-l} \sum_{j,h\in[p]} |\mathrm{Cov}(V_{t,j} V_{t+k,h}, V_{r,j} V_{r+l,h})| \\
&\leq \sum_{t\in\mathbf{Z}} \sum_{j,h\in[p]} |\mathrm{Cov}(V_{0,j} V_{k,h}, V_{t,j} V_{t+l,h})|.
\end{aligned}
$$

Therefore, under Assumptions A.1.2 (i), (iv)

$$\mathbb{E}\|\tilde{\Xi} - \mathbb{E}[\tilde{\Xi}]\|^2 \leq M_T \left(\int |K(u)| \mathrm{d}u + o(1) \right) \frac{B_2}{T}. \tag{A.4}$$

The result follows from combining estimates in equations (A.3) and (A.4). □