# High-Dimensional Granger Causality Tests with an Application to VIX and News[*]

Andrii Babii[†]    Eric Ghysels[‡]    Jonas Striaukas[§]

March 25, 2021

## Abstract

We study Granger causality testing for high-dimensional time series using regularized regressions. To perform proper inference, we rely on heteroskedasticity and autocorrelation consistent (HAC) estimation of the asymptotic variance and develop the inferential theory in the high-dimensional setting. To recognize the time series data structures we focus on the sparse-group LASSO estimator, which includes the LASSO and the group LASSO as special cases. 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. This leads to valid time series inference for individual regression coefficients as well as groups, including Granger causality tests. The treatment relies on a new Fuk-Nagaev inequality for a class of $\tau$-mixing processes with heavier than Gaussian tails, which is of independent interest. In an empirical application, we study the Granger causal relationship between the VIX and financial news.

*Keywords:* Granger causality, high-dimensional time series, heavy tails, inference, tau-mixing, HAC estimator, sparse-group LASSO, Fuk-Nagaev inequality.

[†]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 using high-dimensional datasets, typically available at different frequencies. Conventional time series are often supplemented with 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 by Babii, Ghysels, and Striaukas (2021) to be a useful supplement to more traditional macroeconomic and financial datasets for nowcasting US GDP growth.

In his seminal paper, Clive Granger defined causality in terms of high-dimensional time series data. His formal definition, see (Granger, 1969, Definition 1), considered all the information accumulated in the universe up to time $t-1$ (a process he called $U_t$) and examined predictability using $U_t$ with and without a specific series of interest $Y_t$. It is still an open question how to implement Granger's test in a high-dimensional time series setting. It is the purpose of this paper to do this via regularized regressions using HAC-based inference. In a sense, we are trying to implement Granger's original idea of causality.[1]

It is worth relating our to the existing literature on Granger causality with high-dimensional data. Various dimensionality reduction schemes have been considered. For example, Box and Tiao (1977) used canonical correlation analysis, Peña and Box (1987) and Stock and Watson (2002) proposed factor models and principle component analysis. Koop (2013) analyzed large dimensional Bayesian VAR models. More closely related to our paper are Yuan and Lin (2006), Skripnikov and Michailidis (2019), Simon, Friedman, Hastie, and Tibshirani (2013), Nicholson, Wilms, Bien, and Matteson (2020), and Babii, Ghysels, and Striaukas (2021) who look at structured sparsity approaches without doing inference. Granger causality with sparsity and inference also appeared in a number of papers. Wilms, Gelper, and Croux (2016) use bootstrap but ignore post-selection issues, while Hecq, Margaritella, and Smeekes (2019) extend post-double selection approach of

---

[1]There exists an extensive literature on causal inference with machine learning methods within the *static* Neyman-Rubin's potential outcomes framework; see Athey and Imbens (2019) for the excellent review and further references.

Belloni, Chernozhukov, and Kato (2015) to Granger causality testing in linear sparse high-dimensional VAR.

Following Babii, Ghysels, and Striaukas (2021), we focus on the structured sparsity approach based on the sparse-group LASSO (sg-LASSO) regularization for the high-dimensional time series analysis. The sg-LASSO allows capturing the group structures present in high-dimensional time series regressions where a single covariate with its lags constitutes a group. Alternatively, we can also combine covariates of similar nature in groups. An attractive feature of this estimator is that it encompasses the LASSO and the group LASSO as special cases, hence, it allows 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 model selection problem that dates back to Fisher (1937).[2] In particular, the group structure can also accommodate data sampled at different frequencies as discussed in detail by Babii, Ghysels, and Striaukas (2021).

The proper inference for time-series data relies on the heteroskedasticity and autocorrelation consistent (HAC) estimation of the long-run variance; see Eicker (1963), Huber (1967), White (1980), Gallant (1987), Newey and West (1987), Andrews (1991), among others.[3] Despite the increasing popularity of the LASSO in finance and more generally in time series empirical research, to the best of our knowledge, the validity of HAC-based inference for LASSO has not been established in the relevant literature.[4] The HAC-based inference is robust to the model misspecification and leads to the valid Granger causal-

---

[2]The distributed lag literature can be traced back to Fisher (1925); see also Almon (1965), Sims (1971), and Shiller (1973), 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).

[3]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).

[4]See Chernozhukov, Härdle, Huang, and Wang (2020) for LASSO inference with dependent data and Feng, Giglio, and Xiu (2019) for an asset pricing application; see also Belloni, Chernozhukov, and Hansen (2014) and van de Geer, Bühlmann, Ritov, and Dezeure (2014) for i.i.d. data; and Chiang and Sasaki (2019) for exchangeable arrays.

ity tests even when the fitted regression function has only projection interpretation which is the case for the projection-based definition of the Granger causality. Developing the asymptotic theory for the linear projection model with autoregressive lags and covariates, however, is challenging because the underlying processes are typically *not* $\beta$-mixing.[5]

In this paper, we obtain the debiased central limit theorem with explicit bias correction for the sg-LASSO estimator and time series data, which extends van de Geer, Bühlmann, Ritov, and Dezeure (2014) and to the best of our knowledge is new. Next, we establish the formal statistical properties of the HAC estimator based on the sg-LASSO residuals in the high-dimensional environment when the number of covariates can increase faster than the sample size. The convergence rate of the HAC estimator can be affected by the tails and the persistence of the data, which is a new phenomenon compared to low-dimensional regressions. For the practical implementation, this implies that the optimal choice of the bandwidth parameter for the HAC estimator should scale appropriately with the number of covariates, the tails, and the persistence of the data. These results allow us to perform inference for groups of coefficients, including the (mixed-frequency) Granger causality tests.

Our asymptotic theory applies to the heavy-tailed time series data, which is often observed in financial and economic applications. To that end, we establish a new Fuk-Nagaev inequality, see Fuk and Nagaev (1971), for *$\tau$-mixing* processes with *polynomial tails*. The class of $\tau$-mixing processes is flexible enough for developing the asymptotic theory for the linear projection model and, at the same time, it contains the class of $\alpha$-mixing processes as a special case.

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 $\tau$-mixing 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 groups of regression coefficients. In section 3, we establish a suitable version of the Fuk-Nagaev inequality for $\tau$-mixing processes. We report on a Monte Carlo study in section 4

---

[5]More generally, it is known that the linear transformations based on infinitely many lags do not preserve the $\alpha$- or $\beta$-mixing property.

which provides further insights about the validity of our theoretical analysis in finite sample settings typically encountered in empirical applications. Section 5 covers an empirical application examining the Granger causal relations between the VIX and financial news. Conclusions appear in section 6. Proofs and supplementary results appear in the appendix and the supplementary material.

**Notation:** For a random variable $X \in \mathbf{R}$ and $q \geq 1$, let $\|X\|_q = (\mathbb{E}|X|^q)^{1/q}$ be its $L_q$ norm. For $p \in \mathbf{N}$, put $[p] = \{1, 2, \ldots, p\}$. For a vector $\Delta \in \mathbf{R}^p$ and a subset $J \subset [p]$, let $\Delta_J$ be a vector in $\mathbf{R}^p$ with the same coordinates as $\Delta$ 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 $\beta$ and $\mathcal{G}_0 = \{G \in \mathcal{G} : \ \beta_G \neq 0\}$ is its group support. For $b \in \mathbf{R}^p$ and $q \geq 1$, its $\ell_q$ norm is denoted $|b|_q = \left(\sum_{j \in [p]} |b_j|^q\right)^{1/q}$ if $q < \infty$ and $|b|_\infty = \max_{j \in [p]} |b_j|$ if $q = \infty$. 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 $\mathrm{vech}(A) \in \mathbf{R}^{p(p+1)/2}$ 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 $A$, let $\|A\|_\infty = \max_{j \in [p]} |A_j|_1$ be its matrix norm. 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-based inference for sg-LASSO

In this section, we cover the large sample approximation to the distribution of the sg-LASSO (LASSO and group LASSO) estimator for $\tau$-mixing processes. Next, we consider the HAC estimator of the asymptotic long-run variance based on the sg-LASSO residuals and consider the Granger causality tests. In the first subsection, we cover the debiased central limit theorem. The next subsection covers the HAC estimator and the final subsection

4

pertains to the Granger causality tests.

## 2.1 Debiased central limit theorem

Consider a generic linear projection model

$$y_t = \sum_{j=1}^{\infty} \beta_j x_{t,j} + u_t, \qquad \mathbb{E}[u_t x_{t,j}] = 0, \quad \forall j \geq 1, \qquad t \in \mathbf{Z},$$

where $(y_t)_{t \in \mathbf{Z}}$ is a real-valued stochastic process and predictors may include the intercept, some covariates, (mixed-frequency) 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, we write

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

where $\mathbf{y} = (y_1, \ldots, y_T)^\top$, $\mathbf{m} = (m_1, \ldots, m_T)^\top$ with $m_t = \sum_{j=1}^{\infty} \beta_j x_{t,j}$, and $\mathbf{u} = (u_1, \ldots, u_T)^\top$. We approximate $m_t$ with $x_t^\top \beta = \sum_{j=1}^{p} \beta_j x_{t,j}$ and put $\mathbf{X}\beta$, where $\mathbf{X}$ is $T \times p$ design matrix and $\beta \in \mathbf{R}^p$ is the unknown projection parameter. This approximation 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) described as

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

where $\phi(L) = I - \rho_1 L - \rho_2 L^2 - \cdots - \rho_J L^J$ is a low frequency lag polynomial and the MIDAS part $\psi(L^{1/m}; \beta_k)x_{t,k} = \frac{1}{m}\sum_{j=1}^{m} \beta_{k,j} x_{t-(j-1)/m,k}$ is a high-frequency lag polynomial; see Andreou, Ghysels, and Kourtellos (2013) and Babii, Ghysels, and Striaukas (2021). Note that when $m = 1$ we have all data sampled at the same frequency and recover the 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 (2021), we focus on the sparse-group LASSO (sg-LASSO) regularized estimator.[6] The leading example here is the MIDAS regression involving the projection of future low frequency series onto its own lags and lags of high

---

[6]The sg-LASSO estimator allows selecting groups and important group members at the same time.

frequency data aggregated via some dictionary, e.g., the set of Legendre polynomials. The setup also covers what is sometimes called the reverse MIDAS, see Foroni, Guérin, and Marcellino (2018) and mixed frequency VAR, see Ghysels (2016), involving the projection of high frequency data onto its own (high frequency) lags and low frequency data. Such regressions, which appear in the empirical application of the paper, simply amount to a different group structure.

The sg-LASSO, denoted $\hat{\beta}$, solves the regularized least-squares problem

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

with the regularization functional

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

where $|b|_1 = \sum_{j=1}^p |b_j|$ is the $\ell_1$ norm corresponding to the LASSO penalty, $\|b\|_{2,1} = \sum_{G \in \mathcal{G}} |b_G|_2$ is the group LASSO penalty, and the group structure $\mathcal{G}$ is a partition of $[p] = \{1, 2, \ldots, p\}$ specified by the econometrician.

We measure the persistence of the series with $\tau$-mixing coefficients. For a $\sigma$-algebra $\mathcal{M}$ and a random vector $\xi \in \mathbf{R}^l$, put

$$\tau(\mathcal{M}, \xi) = \left\| \sup_{f \in \mathrm{Lip}_1} |\mathbb{E}(f(\xi)|\mathcal{M}) - \mathbb{E}(f(\xi))| \right\|_1,$$

where $\mathrm{Lip}_1 = \{f : \mathbf{R}^l \to \mathbf{R} : |f(x) - f(y)| \le |x-y|_1\}$ is a set of 1-Lipschitz functions. Let $(\xi_t)_{t \in \mathbf{Z}}$ be a stochastic process and let $\mathcal{M}_t = \sigma(\xi_t, \xi_{t-1}, \ldots)$ be its natural filtration. The $\tau$-mixing coefficient is defined as

$$\tau_k = \sup_{j \ge 1} \frac{1}{j} \sup_{t+k \le t_1 < \cdots < t_j} \tau(\mathcal{M}_t, (\xi_{t_1}, \ldots, \xi_{t_j})), \qquad k \ge 0,$$

where the supremum is taken over $t$ and $(t_1, \ldots, t_j)$. The process is called $\tau$-mixing if $\tau_k \downarrow 0$ as $k \uparrow \infty$; see Lemma A.1.1 for the comparison of this coefficient to the mixingale and the $\alpha$-mixing coefficients. The following assumptions impose tail and moment conditions on the series of interest.

6

**Assumption 2.1** (Data). *The processes $(u_t, x_t)_{t \in \mathbf{Z}}$ is stationary for every $p \geq 1$ and such that (i) $\|u_t\|_q < \infty$ and $\max_{j \in [p]} \|x_{t,j}\|_r = O(1)$ for some $q > 2r/(r-2)$ and $r > 4$; (ii) for every $j, l \in [p]$, the $\tau$-mixing coefficients of $(u_t x_{t,j})_{t \in \mathbf{Z}}$ and $(x_{t,j} x_{t,l})$ are $\tau_k \leq ck^{-a}$ and $\tilde{\tau}_k \leq ck^{-b}$ for all $k \geq 0$ and some universal constants $c > 0$, $a > (\varsigma - 1)/(\varsigma - 2)$, $b > (r-2)/(r-4)$, and $\varsigma = qr/(q+r)$.*

Assumption 2.1 can be relaxed to non-stationary data with stable variances of partial sums at the cost of heavier notation. It allows for heavy-tailed and persistent data. For instance, it requires that either both covariates and the error process have at least $4 + \epsilon$ finite moments, or that the error process has at least $2 + \epsilon$ finite moments, whenever covariates are sufficiently integrable. It is also known that the $\tau$-mixing coefficients decline exponentially fast for geometrically ergodic Markov chains, including the stationary AR(1) process, so condition (ii) allows for relatively persistent data; see also Babii, Ghysels, and Striaukas (2021) for verification of these conditions in a toy heavy-tailed autoregressive model with covariates. Next, we require that the covariance matrix of covariates is invertible.

**Assumption 2.2** (Covariance). *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 $\gamma$.*

Assumption 2.2 ensures that the precision matrix $\Theta = \Sigma^{-1}$ exists and rules out perfect multicollinearity. It also requires that the smallest eigenvalue of $\Sigma$ is bounded away from zero by $\gamma$ independently of the dimension $p$ which is the case, e.g., for the spiked identity and the Toeplitz covariance structures. Strictly, speaking this condition can be relaxed to $\gamma \downarrow 0$ as $p \uparrow \infty$ at the cost of slower convergence rates and more involved conditions on rates, in which case $\gamma$ can be interpreted as a measure of ill-posedness; see Carrasco, Florens, and Renault (2007). The next assumption describes the rate of the regularization parameter, which is governed by the Fuk-Nagaev inequality; see Theorem 3.1 and Eq. 4.

**Assumption 2.3** (Regularization). *For some $\delta \in (0,1)$*

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

*where $\kappa = ((a+1)\varsigma - 1)/(a + \varsigma - 1)$, where $a, \varsigma$ are as in Assumption 2.1.*

In practice we recommend to select the tuning parameter in a data-driven way. It is beyond the scope of the present paper to study properties of estimators with data-driven tuning parameters; see Chetverikov, Liao, and Chernozhukov (2020) for this type of analysis with i.i.d. data. Lastly, we impose the following condition on the misspecification error, the number of covariates $p$, the sparsity constant $s_\alpha$, and the sample size $T$.

**Assumption 2.4.** *(i) $\|\mathbf{m} - \mathbf{X}\beta\|_T^2 = O_P\left(s_\alpha \lambda^2\right)$; (ii) $s_\alpha^\mu p^2 T^{1-\mu} \to 0$ and $p^2 \exp(-cT/s_\alpha^2) \to 0$ as $T \to \infty$, where $s_\alpha$ is the effective sparsity of $\beta$ (defined below) and $\mu = ((b+1)r - 2)/(r + 2(b-1))$.*

The effective sparsity constant $\sqrt{s_\alpha} = \alpha\sqrt{|S_0|} + (1-\alpha)\sqrt{|\mathcal{G}_0|}$ is a linear combination of the sparsity $|S_0|$ (number of non-zero coefficients) and the group sparsity $|\mathcal{G}_0|$ (number of active groups). It reflects the finite sample advantages of imposing the sparse-group structure as $|\mathcal{G}_0|$ can be significantly smaller than $|S_0|$ that appears in the theory of the standard LASSO estimator. Throughout the paper we assume that the groups have fixed size, which is well-justified in time-series applications of interest.

The four assumptions listed above are needed for the prediction and estimation consistency of the sg-LASSO estimator; see Theorem A.1 in the supplementary material. Next, let $v_{t,j}$ be the regression error in $j^{th}$ nodewise LASSO regression; see the following subsection for more details. Put also $s = s_\alpha \vee S$, $S = \max_{j \in G} S_j$, where $S_j$ is the number of non-zero coefficients in the $j^{th}$ row of $\Theta$. The following assumption describes an additional set of sufficient conditions for the debiased central limit theorem.

**Assumption 2.5.** *(i) $\sup_x \mathbb{E}[u_t^2|x_t = x] = O(1)$; (ii) $\|\Theta_G\|_\infty = O(1)$ for some $G \subset [p]$ of fixed size; (iii) the long run variance of $(u_t^2)_{t \in \mathbf{Z}}$ and $(v_{t,j}^2)_{t \in \mathbf{Z}}$ exists for every $j \in G$; (iv) $s^2 \log^2 p/T \to 0$ and $p/\sqrt{T^{\kappa-2} \log^\kappa p} \to 0$; (v) $\|\mathbf{m} - \mathbf{X}\beta\|_T = o_P(T^{-1/2})$; (vi) for every $j, l \in [p]$ and $k \geq 0$, the $\tau$-mixing coefficients of $(u_t u_{t+k} x_{t,j} x_{t+k,l})_{t \in \mathbf{Z}}$ are $\check{\tau}_t \leq ct^{-d}$ for some universal constants $c > 0$ and $d > 1$.*

Assumption (i) requires that the conditional variance of the regression error is bounded. Condition (ii) requires that the rows of the precision matrix have bounded $\ell_1$ norm and is a plausible assumption in the high-dimensional setting, where the inverse covariance matrix

8

is often sparse, e.g., in the Gaussian graphical model. Condition (iii) is a mild restriction needed for the consistency of the sample variance of regression errors. The rate imposed on the sparsity constant, $s^2 \log^2 p/T \to 0$, is also used in van de Geer, Bühlmann, Ritov, and Dezeure (2014) who assume that the regression errors are Gaussian, see their Corollary 2.1. On the other hand, the rate condition on the dimension $p/\sqrt{T^{\kappa-2} \log^\kappa p} \to 0$, is additional condition needed in our setting when regression errors are not Gaussian and may only have a certain number of finite moments. Lastly, condition (v) is trivially satisfied when the projection coefficients are sparse and, more generally, it requires that the misspecification error vanishes asymptotically sufficiently fast. Conditions of this type are standard in nonparametric literature.

Let $B = \hat{\Theta} \mathbf{X}^\top (\mathbf{y} - \mathbf{X}\hat{\beta})/T$ denote the bias-correction for the sg-LASSO estimator, where $\hat{\Theta}$ is the nodewise LASSO estimator of the precision matrix $\Theta$; see the following subsection for more details. The following result describes a large-sample approximation to the distribution of the debiased sg-LASSO estimator with serially correlated non-Gaussian 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 the long-run variance[7] $\Xi_G = \lim_{T \to \infty} \mathrm{Var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T u_t \Theta_G X_t\right)$.*

It is worth mentioning that since the group $G$ has fixed size and the rows of $\Theta$ have finite $\ell_1$ norm, the long-run variance $\Xi_G$ exists under the maintained assumptions; see Proposition A.1.1 in the Appendix for a precise statement of this result.

Theorem 2.1 extends van de Geer, Bühlmann, Ritov, and Dezeure (2014) to non-Gaussian, heavy-tailed and persistent time series data 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,

---

[7]With slight abuse of notation we use $\beta_G \in \mathbf{R}^{|G|}$ to denote the subvector of elements of $\beta \in \mathbf{R}^p$ indexed by $G$.

which requires an appropriate high-dimensional Gaussian approximation result for $\tau$-mixing processes and is left for future research; see Chernozhukov, Chetverikov, and Kato (2013) for a comprehensive review of related coupling results in the i.i.d. case.

**Remark 2.1.** *It is worth mentioning that the debiasing with explicit bias correction addresses the post-model selection issues, see Leeb and Pötscher (2005), and it is fairly straightforward to show that the convergence in Theorem 2.1 holds uniformly over the set of sparse vectors; see also van de Geer, Bühlmann, Ritov, and Dezeure (2014), Corollary 2.1. and the remark following that corollary.*

## 2.2   Nodewise LASSO

The bias-correction term $B$ and the expression of the long-run variance in Theorem 2.1 depend on the appropriate estimator of the precision matrix $\Theta = \Sigma^{-1}$. Following Meinshausen and Bühlmann (2006) and van de Geer, Bühlmann, Ritov, and Dezeure (2014), we focus on the nodewise LASSO estimator of $\Theta$. 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 similarly defined. By the partitioned inverse formula, the $1^{st}$ row of the precision matrix $\Theta = \Sigma^{-1}$ is

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

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 projection error.[8] This suggests estimating the $1^{st}$ row of the precision matrix as $\hat{\Theta}_j = \hat{\sigma}_j^{-2}\left(1 \quad -\hat{\gamma}_j^\top\right)$ 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 $x_j \in \mathbf{R}$ and $\mathbf{X}_{-j}$ is the $T \times (p-1)$ matrix of observations of $x_{-j} \in \mathbf{R}^{p-1}$. In the matrix notation, the nodewise LASSO estimator of $\Theta$ can be written then as $\hat{\Theta} = \hat{B}^{-1}\hat{C}$ with

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

## 2.3 HAC estimator

Next, we focus on the HAC estimator based on sg-LASSO residuals, 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 $\hat{\Gamma}_k = \hat{\Theta}_G \left(\frac{1}{T}\sum_{t=1}^{T-k} \hat{u}_t\hat{u}_{t+k}x_t x_{t+k}^\top\right)\hat{\Theta}_G^\top$, $\hat{u}_t$ is the sg-LASSO residual, and $\hat{\Gamma}_{-k} = \hat{\Gamma}_k^\top$. The kernel function $K : \mathbf{R} \to [-1,1]$ with $K(0) = 1$ is puts less weight on more distant noisy covariances, while $M_T \uparrow \infty$ is a bandwidth (or lag truncation) parameter, see Parzen (1957) and Andrews (1991). Several choices of the kernel function are possible, for example,

---

[8]To ensure that the projection coefficient is well defined and does not change with the dimension of the model $p$, we can consider the limiting linear projection model and take into account the approximation error.

the Parzen kernel is

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

It is worth recalling that the Parzen and the Quadratic spectral kernels are high-order kernels that superior to the Bartlett kernel used in Newey and West (1987); see appendix for more details on the choice of the kernel.

Note that under stationarity, the long-run variance in Theorem 2.1 simplifies to

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

where $\Gamma_k = \Theta_G \mathbb{E}[u_t x_t u_{t+k} x_{t+k}^\top] \Theta_G^\top$ and $\Gamma_{-k} = \Gamma_k^\top$. The following result characterizes the convergence rate of the HAC estimator pertaining to a group of regression coefficients $G \subset [p]$ based on the sg-LASSO residuals.

**Theorem 2.2.** *Suppose that Assumptions 2.1, 2.2, 2.3, 2.4, 2.5 are satisfied for the sg-LASSO regression and for each nodewise LASSO regression $j \in G$. Suppose also that Assumptions A.1.1, and A.1.2 in the Appendix are satisfied for $V_t = (u_t v_{t,j}/\sigma_j^2)_{j \in G}$, $\kappa \ge \tilde{q}$ and that $s^\kappa p T^{1-4\kappa/5} \to 0$ as $M_T \to \infty$ and $T \to \infty$. Then*

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

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. Theorem 2.2 suggests that the optimal choice of the bandwidth parameter should scale appropriately with the number of covariates $p$, the sparsity constant $s$, and the dependence-tails exponent $\kappa$.[9] This contrasts sharply with the HAC theory for regressions without regularization developed in Andrews (1991), see also Li and Liao (2019), and allows for faster convergence rates of the HAC estimator.

---

[9]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 future research, see, e.g., Lazarus, Lewis, Stock, and Watson (2018) for the recent literature review and practical recommendations in the low-dimensional case.

## 2.4 High-dimensional Granger causality tests

Consider a linear projection model

$$y_{t+h} = \sum_{j \in G} \beta_j x_{t,j} + \sum_{j \in G^c} \beta_j x_{t,j} + u_t, \qquad \mathbb{E}[u_t x_{t,j}] = 0, \qquad \forall j \geq 1,$$

where $h \geq 0$ is the horizon, $G \subset [p]$ is a group of regression coefficients of interest, $x_t = \{x_{t,j} : j \in G\}$ represents the series for which we wish to test the Granger causality, and $\{x_{t,j} : j \in G^c\}$ represents all the remaining information available at time $t$. For instance, $x_t$ may contain $L$ low-frequency lags of some series $(z_t)_{t \in \mathbf{Z}}$, in which case $x_t = (z_t, z_{t-1}, z_{t-2}, \ldots, z_{t-L})^\top$. Alternatively, it may contain low and/or high-frequency lags of $(z_t)_{t \in \mathbf{Z}}$ aggregated with dictionaries, e.g., Legendre polynomials as in Babii, Ghysels, and Striaukas (2021). In both cases the dimensionality of $x_t$ small. On the other hand, the set of controls representing all the information available at time $t$ is high-dimensional. The Granger causality test corresponds to the following hypotheses

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

where $\beta_G = \{\beta_j : j \in G\}$.

It is worth mentioning our framework is based on the weakest notion of the Granger causality corresponding to the marginal improvement in time series projections due to the information contained in $x_t$. A stronger notion of Granger non-causality corresponds to the independence on the conditional mean of $y_t$ of $x_t$. Yet, even stronger version of Granger non-causality requires that $y_t | x_t, \sigma(x_{t,j} : j \in G^c) =_d y_t | \sigma(x_{t,j} : j \in G^c)$, so that the entire conditional distribution of $y_t$ does not depend on $x_t$; see also Florens and Mouchart (1982).

Let $R$ be $r \times |G|$ matrix of linear restrictions imposed on $\beta_G$. For the Granger causality test, we set $R = I_{|G|}$, but more generally, we might be interested in testing other linear restrictions implied by the economic theory. Assuming that $R$ is a full row rank matrix, consider the debiased Wald statistics

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

where $A^+$ is the generalized inverse of $A$. It follows from Theorems 2.1 and 2.2 that under $H_0$, $W_T \xrightarrow{d} \chi_r^2$. The Wald test rejects when $W_T > q_{1-\alpha}$, where $q_{1-\alpha}$ is the quantile of

order $1 - \alpha$ of $\chi_r^2$. More generally, the linear restrictions can be extended to the nonlinear restrictions by the usual Delta method argument.

For testing hypotheses on the increasing set of regression coefficients, it might be preferable to use the non-pivotal sup-norm based statistics, see Ghysels, Hill, and Motegi (2020), due to the remarkable performance in the high-dimensional setting; see Chernozhukov, Chetverikov, and Kato (2013) for high-dimensional Gaussian approximations with i.i.d. data.

# 3  Fuk-Nagaev inequality

In this section, we describe 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 $\tau$-mixing coefficients decreasing at a polynomial rate. The following result does not require that the 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 $j \in [p]$, $\tau$-mixing coefficients of $\xi_{t,j}$ satisfy $\tau_k^{(j)} \leq ck^{-a}$ for some universal constants $a, c > 0$. Then there exist $c_1, c_2 > 0$ such that for every $u > 0$*

$$\Pr\left( \left\| \sum_{t=1}^T \xi_t \right\|_\infty > u \right) \leq c_1 p T u^{-\kappa} + 4p \exp\left( -\frac{c_2 u^2}{B_T^2} \right),$$

*where[10] $\kappa = ((a+1)q - 1)/(a + q - 1)$, $B_T^2 = \max_{j \in [p]} \sum_{t=1}^T \sum_{k=1}^T |\mathrm{Cov}(\xi_{t,j}, \xi_{k,j})|$.*

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 $\kappa$ 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, the inequality in Theorem 3.1 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

---

[10]It is worth mentioning that the notation in this section is specific to generic stochastic processes and is independent from the rest of the paper. Thus $B_T$ here denotes the variance of partial sums and not the bias correction term of the LASSO estimator.

with Rosenthal's moment inequality, cf. Nagaev (1998). The proof relies on the blocking technique, see Bosq (1993), and the coupling inequality for $\tau$-mixing sequences, see Dedecker and Prieur (2004), Lemma 5. In contrast to previous results, e.g., Dedecker and Prieur (2004), Theorem 2, the inequality reflects the mixture of the polynomial and the exponential tails.

For stationary processes, by Lemma A.1.2 in the appendix, $B_T^2 = O(T)$ as long as $a > (q-1)/(q-2)$, 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-based inferences for the low-dimensional parameter in a high dimensional data setting. To that end, we draw covariates $\{x_{t,j}, j \in [p]\}$ independently from the AR(1) process

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

The regression error follows the AR(1) process

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

where errors are $\epsilon, \nu \sim_{i.i.d.} N(0,1)$ or $\epsilon, \nu \sim_{i.i.d.}$ student-$t(5)$. The vector of population regression coefficients $\beta$ has the first five non-zero entries which are drawn from Uniform$(0,4)$ and all remaining entries are zero. The sample size is $T \in \{100, 1000\}$ and the number of covariates is $p \in \{10, 200\}$. We set the persistence parameter $\rho = 0.6$ and focus on the LASSO estimator to estimate coefficients $\hat{\beta}$. Throughout the experiment, we choose the LASSO tuning parameters using the 10-fold cross-validation, defining folds as adjacent over time blocks.

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

15

$M_T \in \{5, 10, \dots, 40\}$, using the Parzen kernel. We estimate the long run covariance matrix $\hat{\Xi}$ using the LASSO residuals, denoted $\hat{u}_t$. We also use the nodewise LASSO regressions to estimate the precision matrix $\Theta$. 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 $B_j^{(i)} = \hat{\Theta}_j^{(i)} \mathbf{X}^{\top(i)} \hat{\mathbf{u}}^{(i)}/T$, and $\hat{\mathbf{u}}^{(i)} = \mathbf{y}^{(i)} - \mathbf{X}^{(i)} \hat{\beta}^{(i)}$. 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 is set to $N = 5000$.

We report average results over the active and inactive sets of the vector of coefficients. Table 1 shows results for the small sample size ($T = 100$) and the large sample size ($T = 1000$) for Gaussian data, while in Table 2 we report results for student-$t(5)$ data. We find that the value of the bandwidth parameter $M_T$ should be smaller when the number of regressors $p$ is larger. For $T = 100$ (small sample size), for the active set of coefficients, the best coverage is achieved when the bandwidth parameter is set at 10 when $p = 10$ and at 5 when $p = 200$. Results for the inactive set and $T = 1000$ are similar. We also see that the increase in $p$ relative to $T$ leads to worse performance. Furthermore, the coverage improves when the bandwidth increases with the sample size. The performance for heavy-tailed data decreases relative to Gaussian data – notably, the decrease is larger when $p$ increases. Lastly, as the sample size increases, the average coverage approaches the nominal level of 0.95 and the confidence interval shrinks in size. Overall, the simulation results confirm our theoretical findings.

|  | Average coverage (av. cov) | | | | Confidence interval length | | | |
|---|---|---|---|---|---|---|---|---|
|  | Active set of $\beta$ | | Inactive set of $\beta$ | | Active set of $\beta$ | | Inactive set of $\beta$ | |
| $M_T \backslash p$ | 10 | 200 | 10 | 200 | 10 | 200 | 10 | 200 |
| | | | | T=100 | | | | |
| 5 | 0.830 | 0.755 | 0.827 | 0.747 | 0.286 | 0.382 | 0.277 | 0.368 |
| 10 | 0.834 | 0.750 | 0.835 | 0.746 | 0.305 | 0.401 | 0.291 | 0.376 |
| 15 | 0.824 | 0.753 | 0.832 | 0.745 | 0.314 | 0.411 | 0.294 | 0.376 |
| 20 | 0.821 | 0.735 | 0.826 | 0.743 | 0.320 | 0.420 | 0.296 | 0.374 |
| 25 | 0.816 | 0.739 | 0.820 | 0.741 | 0.325 | 0.427 | 0.297 | 0.372 |
| 30 | 0.807 | 0.738 | 0.814 | 0.740 | 0.329 | 0.434 | 0.297 | 0.370 |
| 35 | 0.806 | 0.733 | 0.807 | 0.738 | 0.333 | 0.441 | 0.297 | 0.368 |
| 40 | 0.801 | 0.737 | 0.802 | 0.735 | 0.337 | 0.447 | 0.297 | 0.366 |
| | | | | T=1000 | | | | |
| 5 | 0.913 | 0.848 | 0.913 | 0.879 | 0.081 | 0.067 | 0.081 | 0.067 |
| 10 | 0.932 | 0.865 | 0.932 | 0.894 | 0.087 | 0.070 | 0.087 | 0.070 |
| 15 | 0.935 | 0.866 | 0.936 | 0.894 | 0.088 | 0.070 | 0.088 | 0.070 |
| 20 | 0.936 | 0.868 | 0.936 | 0.897 | 0.089 | 0.071 | 0.088 | 0.071 |
| 25 | 0.936 | 0.867 | 0.936 | 0.895 | 0.089 | 0.071 | 0.089 | 0.071 |
| 30 | 0.937 | 0.866 | 0.937 | 0.895 | 0.089 | 0.071 | 0.089 | 0.070 |
| 35 | 0.936 | 0.866 | 0.936 | 0.895 | 0.089 | 0.071 | 0.089 | 0.070 |
| 40 | 0.934 | 0.865 | 0.934 | 0.894 | 0.089 | 0.071 | 0.088 | 0.070 |

Table 1: HAC-based 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 $\beta$ and for $T = 100$ and $T = 1000$. We report results for a set of bandwidth parameter $M_T$ values. The data is generated from Gaussian distribution.

|  | Average coverage (av. cov) | | | | Confidence interval length | | | |
|---|---|---|---|---|---|---|---|---|
|  | Active set of $\beta$ | | Inactive set of $\beta$ | | Active set of $\beta$ | | Inactive set of $\beta$ | |
| $M_T\backslash p$ | 10 | 200 | 10 | 200 | 10 | 200 | 10 | 200 |
| | | | | $T{=}100$ | | | | |
| 5 | 0.827 | 0.620 | 0.828 | 0.624 | 0.217 | 0.204 | 0.217 | 0.201 |
| 10 | 0.830 | 0.615 | 0.830 | 0.612 | 0.222 | 0.199 | 0.222 | 0.194 |
| 15 | 0.822 | 0.612 | 0.826 | 0.607 | 0.220 | 0.196 | 0.220 | 0.190 |
| 20 | 0.811 | 0.610 | 0.815 | 0.603 | 0.217 | 0.195 | 0.216 | 0.187 |
| 25 | 0.801 | 0.607 | 0.805 | 0.600 | 0.213 | 0.194 | 0.212 | 0.185 |
| 30 | 0.790 | 0.605 | 0.794 | 0.597 | 0.210 | 0.194 | 0.208 | 0.183 |
| 35 | 0.780 | 0.603 | 0.784 | 0.595 | 0.206 | 0.193 | 0.205 | 0.181 |
| 40 | 0.771 | 0.601 | 0.774 | 0.593 | 0.203 | 0.193 | 0.201 | 0.179 |
| | | | | $T{=}1000$ | | | | |
| 5 | 0.912 | 0.853 | 0.912 | 0.887 | 0.081 | 0.066 | 0.081 | 0.066 |
| 10 | 0.930 | 0.860 | 0.931 | 0.881 | 0.086 | 0.069 | 0.086 | 0.069 |
| 15 | 0.934 | 0.863 | 0.935 | 0.884 | 0.088 | 0.070 | 0.088 | 0.070 |
| 20 | 0.935 | 0.860 | 0.936 | 0.884 | 0.088 | 0.070 | 0.088 | 0.070 |
| 25 | 0.935 | 0.859 | 0.935 | 0.884 | 0.088 | 0.070 | 0.088 | 0.070 |
| 30 | 0.934 | 0.857 | 0.934 | 0.883 | 0.088 | 0.070 | 0.088 | 0.070 |
| 35 | 0.934 | 0.854 | 0.934 | 0.883 | 0.088 | 0.070 | 0.088 | 0.070 |
| 40 | 0.933 | 0.853 | 0.933 | 0.882 | 0.088 | 0.070 | 0.088 | 0.070 |

Table 2: HAC-based 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 $\beta$ and for $T = 100$ and $T = 1000$. We report results for a set of bandwidth parameter $M_T$ values. The data is generated from student-$t(5)$ distribution.

# 5  Testing Granger causality for VIX and financial news

The CBOE Volatility Index, known as the VIX, is a popular measure of market-based expectation of future volatility and is often referred to as the "fear index". The VIX index quotes the expected annualized change in the S&P 500 index over the following 30 days, as computed from options-based theory and current options-market data.

There is a large literature studying the theoretical and empirical properties of the VIX and it is impossible to cite only a few papers to do justice to all the outstanding research output on the topic. Focusing on Granger causal patterns, there are several studies pertaining to causality between the VIX and VIX futures. For example, Bollen, O'Neill, and Whaley (2017) suggest that the VIX futures lagged the VIX in the first few years after its introduction, and show an increasing dominance of VIX futures over time. Along similar lines, Shu and Zhang (2012) study pricediscovery between VIX futures the spot VIX index and find evidence of a bi-directional causal pattern.

We study the causal relationship between financial news and the VIX. There is also substantial literature on the impact of news releases on financial markets (e.g., Andersen, Bollerslev, Diebold, and Vega (2003)). Traditionally, such analysis looks at news releases and studies the behavior of asset prices pre- and post-release. News is usually quantified numerically via the surprise component measured as the difference between an expectation prior to the release and the announcement. In the age of machine learning, the characterization of news has been expanded into the textual analysis of news coverage. To paraphrase the title of Gentzkow, Kelly, and Taddy (2019), the text is treated as data. It is in this spirit that we conduct our high-dimensional Granger causality analysis between the VIX and news.

We use a data set from Bybee, Kelly, Manela, and Xiu (2020) which contains 180 news attention monthly series, all of which potentially Granger cause future US equity market volatility.[11] We estimate the following time series regression model

---

[11]We downloaded daily VIX data from St. Louis Fed FRED database and took the end-of-month values. The FRED mnemonic for the VIX is VIXCLS. Table with the full list of series appears in Appendix A.1.

$$y_{t+1} = \psi(L^{1/m}; \beta)y_t + \sum_{k=1}^{K} \rho_k x_{t,k} + u_t, \qquad t \in [T],$$

where $y_{t+1}$ is the value of the VIX at the end of month $t+1$, $\psi(L^{1/m}; \beta)y_t$ is a MIDAS polynomial of 22 daily VIX lags where the first lag is the last day of the month $t$, and $x_{t,k}$ is the $k$-th news attention series. Note that we only take one lag for the news attention series to simplify the model (and also an empirically justified simplification). The MIDAS polynomial of daily lags of the VIX involves Legendre polynomials of degree 3. Note that the specification is what is sometimes called a reverse MIDAS regression as mentioned earlier in the paper. Prior to estimating the regression model, we time demean the response and covariates such that the intercept is zero. We further standardize all covariates to have a unit standard deviation. The daily VIX lags are standardized before the aggregation.

We apply the sg-LASSO estimator to estimate the slope coefficients and nodewise LASSO regressions to estimate the precision matrix. To fully exploit the group sparsity of sg-LASSO, we group all high-frequency lags of daily VIX, see Babii, Ghysels, and Striaukas (2021) for further details on such grouping. The news attention series are monthly and we are interested in whether the most recent news Granger causes the VIX, hence we don't apply the group structure along the time dimension. Instead, we group news attention series based on a broader theme that is available for each series, see Bybee, Kelly, Manela, and Xiu (2020) for further details. Namely, the data set contains 24 broader topics which group each of the 180 news attention series.

## 5.1   Main results

We report the p-values for a range of $M_T$ values for series that appear to be significant at the 1% or 5% significance level for all $M_T \in \{20, 40, 60\}$ values and for two kernel functions, namely Parzen and Quadratic Spectral. The sample starts January 1990 January and ends June 2017, determined by the availability of the textual analysis data. Both sg-LASSO and LASSO tuning parameters are selected via 10-fold cross-validation, defining folds as adjacent blocks over the time dimension to take into account the time series nature of the data. Similarly, we tune nodewise LASSO regressions for the precision matrix estimation.

### 5.1.1 Granger causality of news topics

The results appear in Table 3 which contains two main row blocks reporting results for the structured sg-LASSO and unstructured LASSO estimators, and two-column blocks, reporting results for two kernel functions. Irrespective of the initial estimator and kernel function, the lagged daily VIX and the Financial crisis news series are highly significant at 1% significance level. Comparing results for the initial estimator, in the case of LASSO we see more significant predictors than for the sg-LASSO case, while the subset of significant covariates using sg-LASSO is a subset of the LASSO significant predictors. Many more series are selected by the initial LASSO estimator compared to the sg-LASSO, see Table A.1. This suggests that relevant group structures are important, and may help in recovering salient relationships in the data.

### 5.1.2 Bi-directional Granger causality

We also test whether the daily VIX Granger causes Financial crisis news series. For this we run the following MIDAS regression model

$$x_{t+1,j} = \psi(L^{1/m}; \beta)y_t + \sum_{k=1}^{K} \rho_k x_{t,k} + u_t, \qquad t \in [T],$$

where $x_{t+1,j}$ is the Financial crisis news series. We test whether daily VIX Granger causes future values of Financial crisis news series. Note that we only need to estimate the initial initial coefficient vector, since the precision matrix remains the same. The results appear in Table 4. They show a rather weak predictability of future news series by daily VIX suggesting a unidirectional Granger causality pattern.

### 5.1.3 Granger causal clusters of news topics

The news attention series are classified into 24 broader meta topicsthat group the individual news series according to a common theme. We test which group of individual news series Granger causes future VIX values. The results are reported in Table 5.They show that the group *Banks* is significant at 5% significance level. This group consists of news series

| Variable $\backslash M_T$ | 20 | 40 | 60 | 20 | 40 | 60 |
|---|---|---|---|---|---|---|
| | | sg-LASSO | | | | |
| | | Parzen | | | Quadratic Spectral | |
| | | 1% significance | | | | |
| Daily VIX lags | 0.000 | 0.000 | 0.001 | 0.000 | 0.001 | 0.002 |
| Financial crisis | 0.005 | 0.002 | 0.001 | 0.002 | 0.001 | 0.000 |
| | | 5% significance | | | | |
| Aerospace/defense | 0.014 | 0.014 | 0.017 | 0.012 | 0.018 | 0.027 |
| Recession | 0.011 | 0.008 | 0.009 | 0.008 | 0.008 | 0.013 |
| | | LASSO | | | | |
| | | Parzen | | | Quadratic Spectral | |
| | | 1% significance | | | | |
| Daily VIX lags | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Financial crisis | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Recession | 0.003 | 0.002 | 0.002 | 0.002 | 0.002 | 0.004 |
| Marketing | 0.001 | 0.001 | 0.000 | 0.001 | 0.000 | 0.000 |
| | | 5% significance | | | | |
| Aerospace/defense | 0.007 | 0.006 | 0.004 | 0.006 | 0.004 | 0.002 |
| NY politics | 0.012 | 0.015 | 0.013 | 0.016 | 0.012 | 0.008 |
| Acquired investment banks | 0.018 | 0.006 | 0.002 | 0.007 | 0.001 | 0.000 |

Table 3: VIX Granger causality results. We report p-values of series that are significant at 1% and 5% significance level for a range of $M_T$ values and both kernel functions.

| Variable $\backslash M_T$ | 20 | 40 | 60 | 20 | 40 | 60 |
|---|---|---|---|---|---|---|
| | | sg-LASSO | | | | |
| | | Parzen | | | Quadratic Spectral | |
| Daily VIX lags | 0.050 | 0.071 | 0.091 | 0.060 | 0.086 | 0.129 |

Table 4: Bi-directional Granger causality results. We report p-values for a range of $M_T$ values and both kernel functions.

| Variable $\backslash M_T$ | 20 | 40 | 60 | 20 | 40 | 60 |
|---|---|---|---|---|---|---|
| | | | sg-LASSO | | | |
| | | Parzen | | | Quadratic Spectral | |
| Banks | 0.032 | 0.024 | 0.008 | 0.023 | 0.001 | 0.000 |

Table 5: Group Granger causality results. We report p-values for a range of $M_T$ values and both kernel functions.

pertaining to news about Mortgages, Bank loans, Credit ratings, Nonperforming loans, Savings & loans, and the Financial crisis.

# 6 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 regression 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, such as Granger causality tests. Our treatment relies on a new suitable variation of the Fuk-Nagaev inequality for $\tau$-mixing 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 based on higher-order asymptotic expansions, see, e.g., Sun, Phillips, and Jin (2008) for steps in this direction in low dimensional settings.

In an empirical application we use a high-dimensional news attention series to study causal patterns between the VIX, sometimes called the fear index, and financial news. We find that almost exclusively the topic of financial crisis exhibits unidirectional Granger causality for the VIX.

# References

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

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND C. VEGA (2003): "Micro effects of macro announcements: Real-time price discovery in foreign exchange," *American Economic Review*, 93(1), 38–62.

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.

ATHEY, S., AND G. W. IMBENS (2019): "Machine learning methods that economists should know about," *Annual Review of Economics*, 11, 685–725.

BABII, A., E. GHYSELS, AND J. STRIAUKAS (2021): "Machine learning time series regressions with an application to nowcasting," *Journal of Business and Economic Statistics* (forthcoming).

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

BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2014): "Inference on treatment effects after selection among high-dimensional controls," *Review of Economic Studies*, 81(2), 608–650.

BELLONI, A., V. CHERNOZHUKOV, AND K. KATO (2015): "Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems," *Biometrika*, 102(1), 77–94.

BOLLEN, N. P., M. J. O'NEILL, AND R. E. WHALEY (2017): "Tail wags dog: Intraday price discovery in VIX markets," *Journal of Futures Markets*, 37(5), 431–451.

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

BOX, G. E., AND G. C. TIAO (1977): "A canonical analysis of multiple time series," *Biometrika*, 64(2), 355–365.

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.

CARRASCO, M., J.-P. FLORENS, AND E. RENAULT (2007): "Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization," *Handbook of Econometrics*, 6, 5633–5751.

CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2013): "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors," *Annals of Statistics*, 41(6), 2786–2819.

CHERNOZHUKOV, V., W. K. HÄRDLE, C. HUANG, AND W. WANG (2020): "Lasso-driven inference in time and space," *Annals of Statistics* (forthcoming).

CHETVERIKOV, D., Z. LIAO, AND V. CHERNOZHUKOV (2020): "On cross-validated lasso in high dimensions," *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 $\tau$-dependent sequences and applications," *Journal of Theoretical Probability*, 17(4), 861–885.

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

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.

FLORENS, J.-P., AND M. MOUCHART (1982): "A note on noncausality," *Econometrica*, 50(3), 583–591.

FORONI, C., P. GUÉRIN, AND M. MARCELLINO (2018): "Using low frequency information for predicting high frequency variables," *International Journal of Forecasting*, 34(4), 774–787.

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.

GENTZKOW, M., B. KELLY, AND M. TADDY (2019): "Text as data," *Journal of Economic Literature*, 57(3), 535–74.

GHYSELS, E. (2016): "Macroeconomics and the reality of mixed frequency data," *Journal of Econometrics*, 193(2), 294–314.

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*, 218, 633–654.

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.

GRANGER, C. W. (1969): "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, 37, 424–438.

HECQ, A., L. MARGARITELLA, AND S. SMEEKES (2019): "Granger causality testing in high-dimensional VARs: a post-double-selection procedure," arXiv preprint arXiv:1902.10991.

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.

KOOP, G. M. (2013): "Forecasting with medium and large Bayesian VARs," *Journal of Applied Econometrics*, 28(2), 177–203.

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.

LEEB, H., AND B. M. PÖTSCHER (2005): "Model selection and inference: Facts and fiction," *Econometric Theory*, pp. 21–59.

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.

NICHOLSON, W. B., I. WILMS, J. BIEN, AND D. S. MATTESON (2020): "High dimensional forecasting via interpretable vector autoregression," *Journal of Machine Learning Research*, 21(166), 1–52.

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

PEÑA, D., AND G. E. BOX (1987): "Identifying a simplifying structure in time series," *Journal of the American statistical Association*, 82(399), 836–843.

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.

SHU, J., AND J. E. ZHANG (2012): "Causality in the VIX futures market," *Journal of Futures Markets*, 32(1), 24–46.

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

SKRIPNIKOV, A., AND G. MICHAILIDIS (2019): "Joint estimation of multiple network Granger causal models," *Econometrics and Statistics*, 10, 120–133.

STOCK, J. H., AND M. W. WATSON (2002): "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97(460), 1167–1179.

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.

WILMS, I., S. GELPER, AND C. CROUX (2016): "The predictive power of the business and bank sentiment of firms: A high-dimensional Granger Causality approach," *European Journal of Operational Research*, 254(1), 138–147.

YUAN, M., AND Y. LIN (2006): "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1), 49–67.

# 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 $\hat{\Theta}$

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

where we use $\mathbf{y} = \mathbf{m} + \mathbf{u}$. Plugging in $\lambda z^*$ and multiplying by $\sqrt{T}$

$$\sqrt{T}(\hat{\beta} - \beta + B) = \hat{\Theta}\mathbf{X}^\top\mathbf{u}/\sqrt{T} + \sqrt{T}(I - \hat{\Theta}\hat{\Sigma})(\hat{\beta} - \beta) + \hat{\Theta}\mathbf{X}^\top(\mathbf{m} - \mathbf{X}\beta)/\sqrt{T}$$

$$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t\Theta x_t + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t(\hat{\Theta} - \Theta)X_t + \sqrt{T}(I - \hat{\Theta}\hat{\Sigma})(\hat{\beta} - \beta)$$

$$+ \hat{\Theta}\mathbf{X}^\top(\mathbf{m} - \mathbf{X}\beta)/\sqrt{T}.$$

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

$$\sqrt{T}(\hat{\beta}_G - \beta_G + B_G) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t\Theta_G x_t + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t(\hat{\Theta}_G - \Theta_G)x_t + \sqrt{T}(I - \hat{\Theta}\hat{\Sigma})_G(\hat{\beta} - \beta)$$

$$+ \hat{\Theta}_G\mathbf{X}^\top(\mathbf{m} - \mathbf{X}\beta)/\sqrt{T}$$

$$\triangleq I_T + II_T + III_T + IV_T.$$

We will show that $I_T \xrightarrow{d} N(0, \Xi_G)$ by the triangular array CLT, see 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)

$$\sum_{t=1}^{T} \mathbb{E}\left|\frac{z^\top \xi_t}{\sqrt{T}}\right|^2 = \mathbb{E}|u_t z^\top \Theta_G x_t|^2$$

$$\leq C z^\top \Theta_G \Sigma \Theta_G^\top z$$

$$= O(1).$$

Therefore, since $q > 2r/(r-2)$, we have $\varsigma > 2$, and for every $\epsilon > 0$

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

Next, under Assumptions 2.1 and 2.5 (i)-(ii), the long run variance

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

exists by Proposition A.1.1.

Next, put $\mathcal{M} = \sigma(\xi_0, \xi_{-1}, \xi_{-2}, \dots)$, $Y = g(z^\top \xi_{t_1 - t_h}/\sqrt{T}, \dots, z^\top \xi_0/\sqrt{T})z^\top \xi_0$, and $X = z^\top \xi_r$ for some $t_h \geq 0$. Note that $X$ and $|XY|$ are integrable and that $Y$ is $\mathcal{M}$-measurable. Therefore, 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 $1 \leq t_1 < t_2 < \cdots < t_h < t_h + r \leq t_h + s \leq T$

$$\left|\mathrm{Cov}\left(g(z^\top \xi_{t_1}/\sqrt{T}, \dots, z^\top \xi_{t_h}/\sqrt{T})z^\top \xi_{t_h}/\sqrt{T}, z^\top \xi_{t_h+r}/\sqrt{T}\right)\right|$$
$$= \frac{1}{T}\left|\mathrm{Cov}(Y, X)\right|$$
$$\leq \frac{1}{T}\int_0^{\gamma(\mathcal{M}, z^\top \xi_r)} Q_Y \circ G_{z^\top \xi_r}(u)\mathrm{d}u$$
$$\leq \frac{1}{T}\int_0^{\gamma(\mathcal{M}, z^\top \xi_r)} Q_{z^\top \xi_0} \circ G_{z^\top \xi_r}(u)\mathrm{d}u$$
$$\leq \frac{1}{T}\|\mathbb{E}(z^\top \xi_r|\mathcal{M}) - \mathbb{E}(z^\top \xi_r)\|_1^{\frac{\varsigma-2}{\varsigma-1}}\|z^\top \xi_0\|_\varsigma^{\varsigma/(\varsigma-1)}$$
$$\leq \frac{1}{T}|\Theta_G^\top z|_1^{\frac{\varsigma-2}{\varsigma-1}}\tau_r^{\frac{\varsigma-2}{\varsigma-1}}\|z^\top \xi_0\|_\varsigma^{\varsigma/(\varsigma-1)} \lesssim r^{-a\frac{\varsigma-2}{\varsigma-1}}$$

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

$$\int Q_{z^\top \xi_0}^{\varsigma-1} \circ G_{z^\top \xi_r}(u)\mathrm{d}u = \int_0^1 Q_{z^\top \xi_0}^\varsigma(u)\mathrm{d}u = \|z^\top \xi_0\|_\varsigma^\varsigma,$$

Appendix - 2

and the last by Lemma A.1.1 and Assumptions 2.1 (ii) and 2.5 (ii). Similarly,

$$\left| \mathrm{Cov} \left( g(z^\top \xi_{t_1}/\sqrt{T}, \ldots, z^\top \xi_{t_h}/\sqrt{T}), z^\top \xi_{t_h+r}/\sqrt{T} z^\top \xi_{t_h+s}/\sqrt{T} \right) \right|$$

$$= \frac{1}{T} \left| \mathrm{Cov} \left( g(z^\top \xi_{t_1-t_h}/\sqrt{T}, \ldots, z^\top \xi_0/\sqrt{T}), z^\top \xi_r z^\top \xi_s \right) \right|$$

$$\leq \frac{1}{T} \int_0^{\gamma(\mathcal{M}, z^\top \xi_r z^\top \xi_s)} Q_g \circ G_{z^\top \xi_r z^\top \xi_s}(u) \mathrm{d}u$$

$$\leq \frac{1}{T} \left\| \mathbb{E}(z^\top \xi_r z^\top \xi_s | \mathcal{M}) - \mathbb{E}(z^\top \xi_r z^\top \xi_s) \right\|_1$$

$$\leq \frac{1}{T} |\Theta_G^\top z|_1^2 \breve{\tau}_r \lesssim r^{-d}.$$

Since the sequence $(r^{-a(\varsigma-2)/(\varsigma-1)\wedge d})_{r\in\mathbf{N}}$ is summable under Assumption 2.1 (ii), all conditions of Neumann (2013), Theorem 2.1, are verified, whence $z^\top I_T \xrightarrow{d} z^\top N(0, \Xi_G)$ for every $z \in \mathbf{R}^{|G|}$.

Next,

$$|II_T|_\infty = \left| (\hat{\Theta} - \Theta)_G \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T u_t x_t \right) \right|_\infty$$

$$\leq \|\hat{\Theta}_G - \Theta_G\|_\infty \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T u_t x_t \right|_\infty$$

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

$$= o_P(1),$$

where the second line follows by $|Ax|_\infty \leq \|A\|_\infty |x|_\infty$, the third line by Proposition A.1.3 and the inequality in Eq. 4 under Assumption 2.1, and the last under Assumption 2.5 (iv). Likewise, using $|Ax|_\infty \leq \max_{j,k} |A_{j,k}|_\infty |x|_1$, by Proposition A.1.3 and Theorem A.1

$$|III_T|_\infty = \sqrt{T} |(I - \hat{\Theta}\hat{\Sigma})_G(\hat{\beta} - \beta)|_\infty$$

$$\leq \sqrt{T} \max_{j\in G} |(I - \hat{\Theta}\hat{\Sigma})_j|_\infty |\hat{\beta} - \beta|_1$$

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

$$= o_P(1)$$

under Assumption 2.5 (iv). Lastly, by the Cauchy-Schwartz inequality, under Assump-

tion 2.5 (v)

$$|IV_T|_\infty \leq \max_{j \in G} |\mathbf{X}\hat{\Theta}_j^\top|_2 \|\mathbf{m} - \mathbf{X}\beta\|_T$$

$$= \max_{j \in G} \sqrt{\hat{\Theta}_j \hat{\Sigma} \hat{\Theta}_j^\top} \, o_P(1)$$

$$= o_P(1),$$

where the last line follows since $\hat{\Theta}_j$ are consistent for $\Theta_j$ in the $\ell_1$ norm while $\hat{\Sigma}$ is consistent for $\Sigma$ in the entrywise maximum norm under the maintained assumptions. $\qquad\square$

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 $\Xi_G$ 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.1.4 with $V_t = (u_t v_{t,j}/\sigma_j^2)_{j \in G}$

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

Next,

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

$$\le |G| \sum_{|k|<T} \left| K\left(\frac{k}{M_T}\right) \right| \max_{j,h\in G} \left| \frac{1}{\hat{\sigma}_j^2 \hat{\sigma}_h^2 T} \sum_{t=1}^{T-k} \hat{u}_t \hat{u}_{t+k} \hat{v}_{t,j} \hat{v}_{t+k,h} - \frac{1}{\sigma_j^2 \sigma_h^2 T} \sum_{t=1}^{T-k} u_t u_{t+k} v_{t,j} v_{t+k,h} \right|$$

$$\le |G| \sum_{|k|<T} \left| K\left(\frac{k}{M_T}\right) \right| \max_{j,h\in G} \frac{1}{\hat{\sigma}_j^2 \hat{\sigma}_h^2} \left| \frac{1}{T} \sum_{t=1}^{T-k} \hat{u}_t \hat{u}_{t+k} \hat{v}_{t,j} \hat{v}_{t+k,h} - \frac{1}{T} \sum_{t=1}^{T-k} u_t u_{t+k} v_{t,j} v_{t+k,h} \right|$$

$$+ |G| \max_{j,h\in G} \left| \frac{1}{\hat{\sigma}_j^2 \hat{\sigma}_h^2} - \frac{1}{\sigma_j^2 \sigma_h^2} \right| \sum_{|k|<T} \left| K\left(\frac{k}{M_T}\right) \right| \left| \frac{1}{T} \sum_{t=1}^{T-k} u_t u_{t+k} v_{t,j} v_{t+k,h} \right|$$

$$\triangleq S_T^a + S_T^b.$$

By Proposition A.1.2, since $s_\alpha^2 \log p / T \to 0$ and $s_\alpha^\kappa p / T^{4\kappa/5-1} \to 0$, 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.1.2

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

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

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

$$\le O(M_T)|G|^2 \sup_{k\in\mathbf{Z}} \max_{j,h\in G} \mathbb{E}|u_t u_{t+k} v_{t,j} v_{t+k,h}|$$

$$= O(M_T),$$

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 $|k| < T$

$$\left| \frac{1}{T} \sum_{t=1}^{T-k} \hat{u}_t \hat{u}_{t+k} \hat{v}_{t,j} \hat{v}_{t+k,h} - \frac{1}{T} \sum_{t=1}^{T-k} u_t u_{t+k} v_{t,j} v_{t+k,h} \right|$$

$$\le \left| \frac{1}{T} \sum_{t=1}^{T-k} (\hat{u}_t \hat{v}_{t,j} - u_t v_{t,j}) u_{t+k} v_{t+k,h} \right| + \left| \frac{1}{T} \sum_{t=1}^{T-k} u_t v_{t,j} (\hat{u}_{t+k} \hat{v}_{t+k,h} - u_{t+k} v_{t+k,h}) \right|$$

$$+ \left| \frac{1}{T} \sum_{t=1}^{T-k} (\hat{u}_t \hat{v}_{t,j} - u_t v_{t,j})(\hat{u}_{t+k} \hat{v}_{t+k,h} - u_{t+k} v_{t+k,h}) \right| \triangleq I_T + II_T + III_T.$$

We bound the first term as

$$I_T \leq \left| \frac{1}{T} \sum_{t=1}^{T-k} (\hat{u}_t - u_t) v_{t,j} u_{t+k} v_{t+k,h} \right| + \left| \frac{1}{T} \sum_{t=1}^{T-k} u_t (\hat{v}_{t,j} - v_{t,j}) u_{t+k} v_{t+k,h} \right|$$

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

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

$$I_T^a = \left| \frac{1}{T} \sum_{t=1}^{T-k} \left( x_t^\top (\beta - \hat{\beta}) + m_t - x_t^\top \beta \right) v_{t,j} u_{t+k} v_{t+k,h} \right|$$

$$\leq (\|\mathbf{X}(\hat{\beta} - \beta)\|_T + \|\mathbf{m} - \mathbf{X}\beta\|_T) \sqrt{\frac{1}{T} \sum_{t=1}^{T-k} v_{t,j}^2 u_{t+k}^2 v_{t+k,h}^2}$$

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

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

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

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

$$I_T^c \leq (\|\mathbf{X}(\hat{\beta} - \beta)\|_T + o_P(T^{-1/2}))(\|\mathbf{X}_{-j}(\hat{\gamma}_j - \gamma_j)\|_T + o_P(T^{-1/2})) \max_{t \in [T]} |u_t v_{t,h}|$$

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

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

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

Appendix - 6

and by symmetry

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

Lastly, by the Cauchy-Schwartz inequality

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

$$\leq \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t\hat{v}_{t,j} - u_t v_{t,j})^2 \frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t\hat{v}_{t,h} - u_t v_{t,h})^2}.$$

For each $j \in G$

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{u}_t\hat{v}_{t,j} - u_t v_{t,j})^2 \leq \frac{3}{T}\sum_{t=1}^{T}|\hat{u}_t - u_t|^2 v_{t,j}^2 + \frac{3}{T}\sum_{t=1}^{T}|\hat{v}_{t,j} - v_{t,j}|^2 u_t^2$$

$$+ \frac{3}{T}\sum_{t=1}^{T}|\hat{u}_t - u_t|^2|\hat{v}_{t,j} - v_{t,j}|^2$$

$$\triangleq III_T^a + III_T^b + III_T^c.$$

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

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

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

$$III_T^c \leq 3(\|\mathbf{X}(\hat{\beta} - \beta)\|_T^2 + o_P(T^{-1/2}))\max_{t\in[T]}|X_{t,-j}^\top(\hat{\gamma}_j - \gamma_j) + m_t - X_t^\top\beta|^2$$

$$\leq O_P\left(\frac{s_\alpha p^{2/\kappa}}{T^{2-2/\kappa}} \vee \frac{s_\alpha \log p}{T}\right)\left(2\max_{t\in[T]}|X_t|_\infty^2|\hat{\gamma}_j - \gamma_j|_1^2 + 2T\|\mathbf{m} - \mathbf{X}^\top\beta\|_T^2\right)$$

$$= O_P\left(\left(\frac{s_\alpha p^{2/\kappa}}{T^{2-2/\kappa}} \vee \frac{s_\alpha \log p}{T}\right)\left(\frac{S^2 p^{2/\kappa}}{T^{2-2/\kappa}} \vee S^2\frac{\log p}{T}\right)(pT)^{1/\kappa}\right)$$

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

$$= O_P\left(\frac{s^3 p^{5/\kappa}}{T^{4-5/\kappa}} + \frac{s^3 p^{3/\kappa}\log p}{T^{3-3/\kappa}}\right),$$

where we use the fact that $\kappa > 2$, $s = s_\alpha \vee S$, $s^\kappa p = o(T^{4\kappa/5-1})$, and $s^2 \log p / T \to 0$ as $T \to \infty$. Then for every $j \in G$

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

and whence

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

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

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

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

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

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

The result follows from combining this estimate with the estimate in equation (A.1). $\qquad\square$

*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

$$\left|\sum_{t=1}^{T}\xi_t\right| \leq \left|\sum_{t\geq 1}V_{2t}^*\right| + \left|\sum_{t\geq 1}V_{2t-1}^*\right| + \sum_{t=3}^{[T/J]+1}|V_t - V_t^*|$$

$$\triangleq I_T + II_T + III_T.$$

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

$$\Pr(I_T \geq x) \leq \frac{c_q^{(1)}}{x^q}\sum_{t\geq 1}\mathbb{E}|V_{2t}^*|^q + 2\exp\left(-\frac{c_q^{(2)}x^2}{\sum_{t\geq 1}\mathrm{Var}(V_{2t}^*)}\right)$$

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

where the second inequality follows since $\sum_{t\geq 1}\mathrm{Var}(V_{2t}^*) = \sum_{t\geq 1}\mathrm{Var}(V_{2t}) \leq B_T^2$. 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, by Markov's inequality and property 3.

$$\Pr\left(III_T \geq x\right) \leq \frac{1}{x}\sum_{t=3}^{[T/J]+1}\tau(\mathcal{M}_t, V_t)$$

$$\leq \frac{1}{x}\sum_{t=3}^{[T/J]+1}\tau(\mathcal{M}_t, (\xi_{(t-1)J+1}, \ldots, \xi_{tJ}))$$

$$\leq \frac{1}{x}[T/J]\sup_{t+J+1\leq t_1<\cdots<t_J}\tau(\mathcal{M}_t, (\xi_{t_1}, \ldots, \xi_{t_J}))$$

$$\leq \frac{T}{x}\tau_{J+1},$$

where the second inequality follows since the sum is a 1-Lipschitz function with respect to $|.|_1$-norm and the third since $\mathcal{M}_t$ and $(\xi_{(t-1)J+1}, \ldots, \xi_{tJ})$ are separated by $J+1$ lags of $(\xi_t)_{t\in\mathbf{Z}}$.

Combining all the estimates together

$$\Pr\left(\left|\sum_{t=1}^{T}\xi_t\right| \geq 3x\right) \leq \Pr(I_T \geq x) + \Pr(II_T \geq x) + \Pr(III_T \geq x)$$

$$\leq \frac{c_q^{(1)}}{x^q}\sum_{t=1}^{[T/J]+1}\mathbb{E}|V_t|^q + 4\exp\left(-\frac{c_q^{(2)}x^2}{B_T^2}\right) + \frac{T}{x}\tau_{J+1}$$

$$\leq \frac{c_q^{(1)}}{x^q}J^{q-1}\sum_{t=1}^{T}\|\xi_t\|_q^q + \frac{T}{x}c(J+1)^{-a} + 4\exp\left(-\frac{c_q^{(2)}x^2}{B_T^2}\right).$$

Appendix - 9

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. □

For a stationary process $(\xi_t)_{t \in \mathbf{Z}}$, let

$$\gamma_k = \|\mathbb{E}(\xi_k | \mathcal{M}_0) - \mathbb{E}(\xi_k)\|_1$$

be its $L_1$ mixingale coefficient with respect to the canonical filtration $\mathcal{M}_0 = \sigma(\xi_0, \xi_{-1}, \xi_{-2}, \dots)$. Let $\alpha_k$ be the $\alpha$-mixing coefficient and let $Q$ be the quantile function of $|\xi_0|$. The following covariance inequality allows us controlling the autocovariances in terms of the $\tau$-mixing coefficient as well as comparing the latter to the mixingale and the $\alpha$-mixing coefficients.

**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*

$$\gamma_t \leq \tau_t \leq 2 \int_0^{2\alpha_t} Q(u) \mathrm{d}u.$$

*Proof.* Let $G$ be the generalized 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. □

The following result shows that the variance of partial sums can be controlled provided that the $\tau$-mixing coefficients decline sufficiently fast.

**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. $\square$

Lastly, we show that in the linear regression setting, the long-run variance for a group of projection coefficients $G \subset [p]$ of fixed size exists under mild conditions. Let $\xi_t = u_t \Theta_G x_t$, where $\Theta_G$ are rows of the precision matrix $\Theta = \Sigma^{-1}$ corresponding to indices in $G$.

**Proposition A.1.1.** *Suppose that (i) $(u_t x_t)_{t \in \mathbf{Z}}$ is stationary for every $p \geq 1$; (ii) $\|u_t\|_q < \infty$ and $\max_{j \in [p]} \|x_{t,j}\|_r = O(1)$ for some $q > 2r/(r-2)$ and $r > 4$; (iii) for every $j \in [p]$, the $\tau$-mixing coefficients of $(u_t x_{t,j})_{t \in \mathbf{Z}}$ are $\tau_k \leq ck^{-a}$ for all $k \geq 0$, where $c > 0$ and $a > (\varsigma - 1)/(\varsigma - 2)$, and $\varsigma = qr/(q + r)$ are some universal constants; (iv) $\|\Theta_G\|_\infty = O(1)$ and $\sup_x \mathbb{E}[|u_t|^2 | x_t = x] = O(1)$. Then for every $z \in \mathbf{R}^{|G|}$, the limit*

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

*exists.*

*Proof.* Under assumption (i), by Hölder's inequality, $\max_{j \in [p]} \|u_t x_{t,j}\|_\varsigma = O(1)$ with $\varsigma = qr/(q + r)$, whence by the Minkowski inequality and assumption (iv)

$$\|z^\top \xi_t\|_\varsigma \leq \sum_{k \in G} \sum_{j \in [p]} |\Theta_{k,j}| \|u_t x_{t,j}\|_\varsigma \leq |G| \|\Theta_G\|_\infty = O(1). \tag{A.2}$$

Since $\varsigma > 2$, this shows that $\text{Var}(z^\top \xi_0)$ exists. Moreover,

$$\text{Var}(z^\top \xi_0) = z^\top \Theta_G \text{Var}(u_0 x_0) \Theta_G^\top z$$
$$= \sum_{j,k \in [p]} (z^\top \Theta_G)_j (z^\top \Theta_G)_k \mathbb{E}[u_0^2 x_{0,j} x_{0,k}],$$

where the sum converges as $p \to \infty$ by the comparison test under assumption (iv) implying that $\lim_{T \to \infty} \text{Var}(z^\top \xi_0)$ exists. Next, under assumption (i), for every $z \in \mathbf{R}^{|G|}$

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

By Lemma A.1.1, we bound covariances by the mixingale coefficient for every $k \geq 1$

$$|\text{Cov}(z^\top \xi_0, z^\top \xi_k)| \lesssim \|z^\top \xi_0\|_\varsigma^{\varsigma/(\varsigma-1)} \|\mathbb{E}(z^\top \xi_k | \mathcal{M}_0) - \mathbb{E}(z^\top \xi_k)\|_1^{\frac{\varsigma-2}{\varsigma-1}}$$
$$\lesssim |z^\top \Theta_G|_1^{\frac{\varsigma-2}{\varsigma-1}} \max_{j \in [p]} \|\mathbb{E}(u_k x_{k,j} | \mathcal{M}_0) - \mathbb{E}(u_k x_{k,j})\|_1^{\frac{\varsigma-2}{\varsigma-1}}$$
$$\lesssim \tau_k^{\frac{\varsigma-2}{\varsigma-1}}$$

where the first inequality follows by Lemma A.1.1, the second by equation (A.2), and the third by Lemma A.1.1. Under assumption (iii), $\sum_{k=1}^\infty \tau_k^{(\varsigma-2)/(\varsigma-1)}$ converges, which implies that

$$\sum_{k=1}^\infty |\text{Cov}(z^\top \xi_0, z^\top \xi_k)| < \infty$$

by the comparison test. Therefore, by Lebesgue's dominated convergence, this shows that the long run variance

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

exists. $\square$

# 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 (2021), 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.1.2.** *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

$$|\hat{\sigma}^2 - \sigma^2| = \left|\|\mathbf{u}\|_T^2 + 2\langle\mathbf{u}, \mathbf{m} - \mathbf{X}\hat{\beta}\rangle_T - \|\mathbf{m} - \mathbf{X}\hat{\beta}\|_T^2 + \lambda\Omega(\hat{\beta}) - \sigma^2\right|$$

$$\leq |\sigma^2 - \|\mathbf{u}\|_T^2| + 2\|\mathbf{u}\|_T\|\mathbf{m} - \mathbf{X}\hat{\beta}\|_T + 2\|\mathbf{X}(\hat{\beta} - \beta)\|_T^2 + 2\|\mathbf{m} - \mathbf{X}\beta\|_T^2 + \lambda\Omega(\hat{\beta})$$

$$\triangleq I_T + II_T + III_T + IV_T + V_T.$$

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

$$II_T = O_P(1)\|\mathbf{m} - \mathbf{X}\hat{\beta}\|_T$$

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

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

$$V_T = \lambda\Omega(\hat{\beta} - \beta) + \lambda\Omega(\beta)$$

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

The result follows from combining all estimates together. $\qquad\square$

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 $\gamma_j$.

**Proposition A.1.3.** *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 $S^\kappa p T^{1-\kappa} \to 0$ and $S^2 \log p/T \to 0$*

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

*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.1.2 with $\alpha = 1$ (corresponding to the LASSO estimator of $\gamma_j$ and $\sigma_j^2$)

$$\|\hat{\Theta}_G - \Theta_G\|_\infty = \max_{j \in G} |\hat{\Theta}_j - \Theta_j|_1$$

$$\leq \max_{j \in G} \left\{ |\hat{\gamma}_j|_1 \left| \hat{\sigma}_j^{-2} - \sigma_j^{-2} \right| + |\hat{\gamma}_j - \gamma_j|_1 |\sigma_j^{-2}| \right\}$$

$$= O_P \left( \frac{S p^{1/\kappa}}{T^{1-1/\kappa}} \vee S \sqrt{\frac{\log p}{T}} \right),$$

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

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

$$\mathbf{X}_j^\top (\mathbf{X}_j - \mathbf{X}_{-j} \hat{\gamma}_j)/T = \|\mathbf{X}_j - \mathbf{X}_{-j} \hat{\gamma}_j\|_T^2 + \hat{\gamma}_j^\top \mathbf{X}_{-j}^\top (\mathbf{X}_j - \mathbf{X}_{-j} \hat{\gamma}_j)/T$$

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

and whence

$$|(I - \hat{\Theta}\hat{\Sigma})_j|_\infty = |I_j - (\mathbf{X}_j - \mathbf{X}_{-j} \hat{\gamma}_j)^\top \mathbf{X}/(T\hat{\sigma}_j^2)|_\infty$$

$$= \max \left\{ |1 - \mathbf{X}_j^\top (\mathbf{X}_j - \mathbf{X}_{-j} \hat{\gamma}_j)/(T\hat{\sigma}_j^2)|, |\mathbf{X}_{-j}^\top (\mathbf{X}_j - \mathbf{X}_{-j} \hat{\gamma}_j)/(T\hat{\sigma}_j^2)|_\infty \right\}$$

$$= \lambda_j |z^*|_\infty / \hat{\sigma}_j^2 = O_P \left( \frac{p^{1/\kappa}}{T^{1-1/\kappa}} \vee \sqrt{\frac{\log p}{T}} \right),$$

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. $\qquad\square$

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) $\Gamma \in \mathcal{G}(\varsigma, D_1, D_2)$, where*

$$\mathcal{G}(\varsigma, D_1, D_2) = \left\{ \sum_{k \in \mathbf{Z}} |k|^\varsigma \|\Gamma_k\| \leq D_1, \quad \sup_{k \in \mathbf{Z}} \sum_{l \in \mathbf{Z}} \sum_{t \in \mathbf{Z}} \sum_{j,h \in [p]} |\mathrm{Cov}(V_{0,j} V_{k,h}, V_{t,j} V_{t+l,h})| \leq D_2 \right\}$$

*for some $D_1, D_2 > 0$.*

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/$\tau$-mixing 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.1.4.** *Suppose that Assumption A.1.2 is satisfied. Then*

$$\sup_{\Gamma \in \mathcal{G}(\varsigma, D_1, D_2)} \mathbb{E}\|\tilde{\Xi} - \Xi\|^2 \leq C_1 \frac{M_T}{T} + C_2 M_T^{-2\varsigma} + C_3 T^{-2(\varsigma \wedge 1)},$$

*where $C_1 = D_2 \left( \int |K(u)| \mathrm{d}u + o(1) \right)$, $C_2 = 2 \left( D_1 L + \frac{2D_1}{\varepsilon^\varsigma} \right)^2$, and $C_3 = 2D_1^2$.*

*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

$$
\begin{aligned}
I_T &= \sum_{|k|<\varepsilon M_T} \left| K(0) - K\left(\frac{k}{M_T}\right) \right| \|\Gamma_k\| + \sum_{\varepsilon M_T \leq |k|<T} \left| K\left(\frac{k}{M_T}\right) - K(0) \right| \|\Gamma_k\| \\
&\leq L M_T^{-\varsigma} \sum_{|k|<\varepsilon M_T} |k|^\varsigma \|\Gamma_k\| + 2 \sum_{\varepsilon M_T \leq |k|<T} \|\Gamma_k\| \\
&\leq \frac{D_1 L}{M_T^\varsigma} + \frac{2}{\varepsilon^\varsigma M_T^\varsigma} \sum_{\varepsilon M_T \leq |k|<T} |k|^\varsigma \|\Gamma_k\| \\
&\leq \frac{D_1 L}{M_T^\varsigma} + \frac{2 D_1}{\varepsilon^\varsigma M_T^\varsigma},
\end{aligned}
$$

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)

$$
\begin{aligned}
II_T + III_T &\leq \begin{cases} \frac{D_1}{T} & \varsigma \geq 1 \\[2mm] \frac{D_1}{T^\varsigma} & \varsigma \in (0,1) \end{cases} \\
&= \frac{D_1}{T^{\varsigma \wedge 1}}.
\end{aligned}
$$

This shows that

$$
\|\mathbb{E}[\tilde{\Xi}] - \Xi\| \leq \frac{D_1 L}{M_T^\varsigma} + \frac{2 D_1}{\varepsilon^\varsigma M_T^\varsigma} + \frac{D_1}{T^{\varsigma \wedge 1}}. \tag{A.3}
$$

Next, under Assumption A.1.2 (i)

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

$$\leq \sum_{|k|<T} \left| K\left(\frac{k}{M_T}\right)\right| \sup_{|k|<T} \sum_{|l|<T} \left| \mathbb{E}\left\langle \tilde{\Gamma}_k - \mathbb{E}\tilde{\Gamma}_k, \tilde{\Gamma}_l - \mathbb{E}\tilde{\Gamma}_l \right\rangle \right|,$$

where under Assumptions A.1.2 (iii)

$$T\left| \mathbb{E}\left\langle \tilde{\Gamma}_k - \mathbb{E}\tilde{\Gamma}_k, \tilde{\Gamma}_l - \mathbb{E}\tilde{\Gamma}_l \right\rangle \right| \leq \frac{1}{T}\sum_{t=1}^{T-k}\sum_{r=1}^{T-l}\sum_{j,h\in[p]} |\text{Cov}(V_{t,j}V_{t+k,h}, V_{r,j}V_{r+l,h})|$$

$$\leq \sum_{t\in\mathbf{Z}}\sum_{j,h\in[p]} |\text{Cov}(V_{0,j}V_{k,h}, V_{t,j}V_{t+l,h})|.$$

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

$$\mathbb{E}\|\tilde{\Xi} - \mathbb{E}[\tilde{\Xi}]\|^2 \leq M_T\left(\int |K(u)|\mathrm{d}u + o(1)\right)\frac{D_2}{T}. \tag{A.4}$$

The result follows from combining estimates in equations (A.3) and (A.4). $\qquad\square$

| | News topic | Meta topic | LASSO | sg-LASSO |
|---|---|---|---|---|
| 1 | Accounting | Asset Managers & I-Banks | | |
| 2 | Acquired investment banks | Asset Managers & I-Banks | ✓ | ✓ |
| 3 | Activists | Activism/Language | | |
| 4 | Aerospace/defense | Trans/Retail/Local Politics | ✓ | ✓ |
| 5 | Agreement reached | Negotiations | | |
| 6 | Agriculture | Oil & Mining | | |
| 7 | Airlines | Trans/Retail/Local Politics | | |
| 8 | Announce plan | Activism/Language | | |
| 9 | Arts | Social/Cultural | | |
| 10 | Automotive | Trans/Retail/Local Politics | | |
| 11 | Bank loans | Banks | | |
| 12 | Bankruptcy | Buyouts & Bankruptcy | | |
| 13 | Bear/bull market | Financial Markets | ✓ | ✓ |
| 14 | Biology/chemistry/physics | Science/Language | | |
| 15 | Bond yields | Financial Markets | | |
| 16 | Broadcasting | Entertainment | ✓ | ✓ |
| 17 | Buffett | Activism/Language | | |
| 18 | Bush/Obama/Trump | Leaders | | |
| 19 | C-suite | Management | | |
| 20 | Cable | Industry | | |
| 21 | California | Trans/Retail/Local Politics | | |
| 22 | Canada/South Africa | International Affairs | | |
| 23 | Casinos | Industry | ✓ | |
| 24 | Challenges | Challenges | | |
| 25 | Changes | Challenges | | |
| 26 | Chemicals/paper | Industry | | |
| 27 | China | International Affairs | | |
| 28 | Clintons | Leaders | | |
| 29 | Committees | Negotiations | | |
| 30 | Commodities | Financial Markets | | |
| 31 | Company spokesperson | Negotiations | | |
| 32 | Competition | Industry | | |
| 33 | Computers | Technology | | |
| 34 | Connecticut | Management | | |
| 35 | Control stakes | Buyouts & Bankruptcy | ✓ | |
| 36 | Convertible/preferred | Buyouts & Bankruptcy | | |
| 37 | Corporate governance | Buyouts & Bankruptcy | ✓ | ✓ |
| 38 | Corrections/amplifications | Activism/Language | | |
| 39 | Couriers | Industry | | |
| 40 | Courts | Courts | ✓ | |
| 41 | Credit cards | Industry | | |
| 42 | Credit ratings | Banks | ✓ | ✓ |
| 43 | Cultural life | Social/Cultural | | |

| 44 | Currencies/metals | Financial Markets | | |
| 45 | Disease | Trans/Retail/Local Politics | ✓ | ✓ |
| 46 | Drexel | Buyouts & Bankruptcy | ✓ | |
| 47 | Earnings | Corporate Earnings | | |
| 48 | Earnings forecasts | Corporate Earnings | | |
| 49 | Earnings losses | Corporate Earnings | ✓ | |
| 50 | Economic growth | Economic Growth | | |
| 51 | Economic ideology | Social/Cultural | | |
| 52 | Elections | Leaders | ✓ | ✓ |
| 53 | Electronics | Technology | ✓ | ✓ |
| 54 | Environment | Government | ✓ | |
| 55 | European politics | Leaders | ✓ | |
| 56 | European sovereign debt | Economic Growth | ✓ | ✓ |
| 57 | Exchanges/composites | Financial Markets | | |
| 58 | Executive pay | Labor/income | | |
| 59 | Fast food | Industry | ✓ | |
| 60 | Federal Reserve | Economic Growth | ✓ | ✓ |
| 61 | Fees | Labor/income | | |
| 62 | Financial crisis | Banks | ✓ | ✓ |
| 63 | Financial reports | Corporate Earnings | | |
| 64 | Foods/consumer goods | Industry | | |
| 65 | France/Italy | International Affairs | | |
| 66 | Futures/indices | Activism/Language | | |
| 67 | Gender issues | Social/Cultural | | |
| 68 | Germany | International Affairs | ✓ | ✓ |
| 69 | Government budgets | Labor/income | | |
| 70 | Health insurance | Labor/income | | |
| 71 | Humor/language | Social/Cultural | | |
| 72 | Immigration | Social/Cultural | | |
| 73 | Indictments | Courts | | |
| 74 | Insurance | Industry | | |
| 75 | International exchanges | Financial Markets | | |
| 76 | Internet | Technology | | |
| 77 | Investment banking | Asset Managers & I-Banks | | |
| 78 | IPOs | Financial Markets | | |
| 79 | Iraq | Terrorism/Mideast | ✓ | |
| 80 | Japan | International Affairs | ✓ | |
| 81 | Job cuts | Labor/income | | |
| 82 | Justice Department | Courts | | |
| 83 | Key role | Challenges | | |
| 84 | Latin America | International Affairs | ✓ | ✓ |
| 85 | Lawsuits | Courts | | |
| 86 | Long/short term | Challenges | | |
| 87 | Luxury/beverages | Industry | | |
| 88 | M&A | Buyouts & Bankruptcy | ✓ | |

| | | | | |
|---|---|---|---|---|
| 89 | Machinery | Oil & Mining | | |
| 90 | Macroeconomic data | Economic Growth | | |
| 91 | Major concerns | Activism/Language | | |
| 92 | Management changes | Management | ✓ | |
| 93 | Marketing | Entertainment | ✓ | |
| 94 | Mexico | Activism/Language | ✓ | |
| 95 | Microchips | Technology | | |
| 96 | Mid-level executives | Management | | |
| 97 | Mid-size cities | Trans/Retail/Local Politics | | |
| 98 | Middle east | Terrorism/Mideast | ✓ | |
| 99 | Mining | Oil & Mining | | |
| 100 | Mobile devices | Technology | | |
| 101 | Mortgages | Banks | ✓ | |
| 102 | Movie industry | Entertainment | ✓ | |
| 103 | Music industry | Entertainment | | |
| 104 | Mutual funds | Asset Managers & I-Banks | | |
| 105 | NASD | Asset Managers & I-Banks | ✓ | ✓ |
| 106 | National security | Government | ✓ | |
| 107 | Natural disasters | Trans/Retail/Local Politics | ✓ | ✓ |
| 108 | Negotiations | Negotiations | | |
| 109 | News conference | Negotiations | | |
| 110 | Nonperforming loans | Banks | ✓ | |
| 111 | Nuclear/North Korea | Terrorism/Mideast | ✓ | |
| 112 | NY politics | Trans/Retail/Local Politics | ✓ | |
| 113 | Oil drilling | Oil & Mining | | |
| 114 | Oil market | Oil & Mining | ✓ | |
| 115 | Optimism | Economic Growth | | |
| 116 | Options/VIX | Financial Markets | ✓ | ✓ |
| 117 | Pensions | Labor/income | | |
| 118 | People familiar | Negotiations | | |
| 119 | Pharma | Trans/Retail/Local Politics | | |
| 120 | Phone companies | Technology | | |
| 121 | Police/crime | Trans/Retail/Local Politics | | |
| 122 | Political contributions | Government | | |
| 123 | Positive sentiment | Social/Cultural | | |
| 124 | Private equity/hedge funds | Asset Managers & I-Banks | | |
| 125 | Private/public sector | Government | | |
| 126 | Problems | Challenges | | |
| 127 | Product prices | Economic Growth | | |
| 128 | Profits | Corporate Earnings | ✓ | ✓ |
| 129 | Programs/initiatives | Science/Language | | |
| 130 | Publishing | Entertainment | ✓ | |
| 131 | Rail/trucking/shipping | Trans/Retail/Local Politics | | |
| 132 | Reagan | Leaders | | |
| 133 | Real estate | Buyouts & Bankruptcy | | |

| # | Term | Category | | |
|---|---|---|---|---|
| 134 | Recession | Economic Growth | ✓ | ✓ |
| 135 | Record high | Economic Growth | | |
| 136 | Regulation | Government | | |
| 137 | Rental properties | Trans/Retail/Local Politics | ✓ | |
| 138 | Research | Science/Language | | |
| 139 | Restraint | Negotiations | | |
| 140 | Retail | Trans/Retail/Local Politics | | |
| 141 | Revenue growth | Industry | | |
| 142 | Revised estimate | Corporate Earnings | | |
| 143 | Russia | International Affairs | ✓ | |
| 144 | Safety administrations | Government | | |
| 145 | Sales call | Social/Cultural | | |
| 146 | Savings & loans | Banks | ✓ | ✓ |
| 147 | Scenario analysis | Science/Language | | |
| 148 | Schools | Social/Cultural | | |
| 149 | SEC | Buyouts & Bankruptcy | | |
| 150 | Share payouts | Financial Markets | | |
| 151 | Short sales | Financial Markets | ✓ | |
| 152 | Size | Science/Language | | |
| 153 | Small business | Industry | | |
| 154 | Small caps | Financial Markets | ✓ | ✓ |
| 155 | Small changes | Corporate Earnings | | |
| 156 | Small possibility | Challenges | | |
| 157 | Soft drinks | Industry | | |
| 158 | Software | Technology | ✓ | |
| 159 | Southeast Asia | International Affairs | | |
| 160 | Space program | Science/Language | | |
| 161 | Spring/summer | Challenges | | |
| 162 | State politics | Government | | |
| 163 | Steel | Oil & Mining | | |
| 164 | Subsidiaries | Industry | | |
| 165 | Systems | Science/Language | | |
| 166 | Takeovers | Buyouts & Bankruptcy | ✓ | |
| 167 | Taxes | Labor/income | ✓ | |
| 168 | Terrorism | Terrorism/Mideast | | |
| 169 | Tobacco | Industry | | |
| 170 | Trade agreements | International Affairs | ✓ | ✓ |
| 171 | Trading activity | Financial Markets | | |
| 172 | Treasury bonds | Financial Markets | | |
| 173 | UK | International Affairs | | |
| 174 | Unions | Labor/income | ✓ | ✓ |
| 175 | US defense | Trans/Retail/Local Politics | ✓ | |
| 176 | US Senate | Leaders | | |
| 177 | Utilities | Government | ✓ | |
| 178 | Venture capital | Industry | | |

Appendix - 22

| 179 | Watchdogs | Government |
| 180 | Wide range | Science/Language |

Table A.1: News series – The column *News topic* are the news series topics, column *Meta topic* are meta topics/groups of news series. Columns *LASSO* and *sg-LASSO* reports whether the series was selected (✓) or not by the respective initial estimator.