High-dimensional Gaussian Graphical Model Testing for Long-Memory Time Series

Percy S. Zhai
The University of Chicago
Chicago, IL 60637
percy.zhai@chicagobooth.edu

Ping-Shou Zhong
University of Illinois Chicago
Chicago, IL 60607
pszhong@uic.edu

Wei Biao Wu
The University of Chicago
Chicago, IL 60637
wbwu@uchicago.edu

Abstract

Many real-world high-dimensional time series exhibit long-memory, but Gaussian graphical model testing in this regime remains understudied. We develop a direct, data-adaptive test statistic for assessing conditional independence in the graph structure of stationary Gaussian time series. We establish a finite-sample, Berry–Esseen type Gaussian approximation bound for the statistic, which applies to both short-memory and long-memory time series. The testing procedure uses a block bootstrap method, on which we provide a finite-sample validity result including in the ultra-high-dimensional scenario, and can be extended to comparing graphical structures in two-sample tests. We also develop a consistency-empowered correction to the statistic and show that such tests attain asymptotic consistency in both size and power. Our proposed method is applied to real-world fMRI data to understand functional connectivities within brain in different periods.

1 Introduction

Testing on Gaussian graphical models for high-dimensional time series remains relatively understudied as opposed to its i.i.d. counterparts. Existing work has largely been confined to the short-memory regime, where the autocorrelation decays fast enough. The main difficulty, though, is that real-world time series data usually exhibits long-range temporal dependence (i.e., long-memory). This paper proposes a novel method for testing Gaussian graphical models of stationary high-dimensional time series that allows for long-range dependence.

Gaussian graphical models encode conditional independence through the zero pattern of the precision matrix (Lauritzen, 1996). In the i.i.d. setting, a large body of work has studied precision matrix estimation, including neighborhood selection (Meinshausen and Bühlmann, 2006), penalized likelihood (Friedman et al., 2008), and constrained ℓ_1 methods (Cai et al., 2011). The corresponding inference literature is more recent. One direction focuses on edgewise or low-dimensional inference for entries of Ω, where multiple-testing procedures (Liu, 2013) or debiased precision estimators (Ren et al., 2015; Janková and van de Geer, 2015) are used to test individual conditional dependencies. Another direction develops global or structural tests for precision matrices, including neighborhood tests (Verzelen and Villers, 2009; Zhao et al., 2024), tests for special structures such as bandedness (Cheng et al., 2017), and two-sample differential-network tests (Xia et al., 2015; Zhao et al., 2022). Le et al. (forthcoming) proposed direct plug-in and consistency-empowered tests for the adequacy of a prespecified Gaussian graph. More recently, Lin et al. (2025) proposed exact finite-sample

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

goodness-of-fit tests for high-dimensional Gaussian graphical models via exchangeable sampling. These methods provide important tools for i.i.d. Gaussian observations, but require independence or exchangeability.

For time series, graph-related work has mostly focused on estimation rather than graph-specification testing. Chen et al. (2013) studied covariance and precision matrix estimation for high-dimensional time series using functional dependence measure. Qiu et al. (2016) considered joint estimation of multiple graphical models from high-dimensional time series with dependent observations. Shu and Nan (2019) developed covariance and precision estimation for processes with slowly decaying temporal dependence, and Baek et al. (2023) studied high-dimensional long-run variance and long-run precision estimation under general dependence, including long-range dependence. Apart from these valuable estimation methods, a data-driven goodness-of-fit test for long-memory high-dimensional time series is largely unexplored.

Meanwhile, a separate line of work develops high-dimensional inference for short-memory time series. Zhang and Wu (2017) developed Gaussian approximation for high-dimensional stationary time series, and Zhang and Cheng (2014) studied bootstrap inference for high-dimensional time series. These results are designed for short-range dependence and do not easily extend into the long-memory regime. Only recently did Zhai et al. (2026) establish Gaussian approximation and block bootstrap validity for sample covariance errors in high-dimensional long-memory time series. The same work also developed precision matrix inference, but only in the low-dimensional regime where the sample covariance matrix is invertible.

This paper fills the gap between these two literatures. We use a graph-specific statistic adapted from Le et al. (forthcoming) based on off-support local inverse, and demonstrate that the first-order behavior of the statistic is governed by the sample covariance error, which further enables us to import the machinery in Zhai et al. (2026), without the need of estimating the full precision matrix. The new graph-specific steps are the linearization and plug-in error control in Propositions 1 and 2, together with verification of the imported assumptions for the transformed process $Y_t = HX_t$. Our Theorem 1 provides a finite-sample Gaussian approximation result on the proposed statistic. We then propose a testing method based on block-bootstrap, and show its finite-sample validity in Theorem 2. We further add a consistency-empowered correction for the case where the hypothesized graph is larger than the true graph. The correction screens precision estimates near zero on the null support and artificially amplifies them. It preserves the null calibration of the direct test, while restoring power against fixed over-specified alternatives; see Theorem 4.

For application, we consider a two-sample extension. Given two samples, we estimate a graph from the first group and use it as the null support for testing the second group. Corollary 3 shows that the Gaussian approximation and block bootstrap validity remains as long as the graph estimator is support consistent. Our proposed method is illustrated on the Sherlock fMRI dataset that compares the brain connectivity pattern of subjects when watching a movie and when verbally recalling the contents.

We would list the following three major contributions.

1. We propose a novel approach for testing edges in Gaussian graphical models for stationary time series that allows **long-memory**. We apply a local-inverse-based test statistic related to sample covariance, and provide a Berry-Esseen type bound of Gaussian approximation and a block-bootstrap validity result. These **finite-sample bounds** demonstrate that our test applies to the **ultra-high-dimensional** regime under explicit sparsity and dependence conditions, and guarantees the test size.
2. We also propose a **consistency-empowered** corrected test statistic for time series, which is additionally capable of detecting edges in the hypothesized graph that are not significant in the true graph. We develop theoretical results that the test size is not affected, while the test power is guaranteed.
3. We impose a **two-sample extension** of the theory that tests whether the graphical structure is identical across the two samples by combining two directional support-containment tests. This setup is more convenient in application, and we demonstrate this two-sample methodology on the Sherlock fMRI dataset.

The structure of the paper is as follows. Section 2 introduces the problem setup, propose the statistic, describe the block bootstrap algorithm, and propose the consistency-empowered correction on the

statistic. Section 3 provides theoretical guarantee of our proposed methods. Quantitative study that illustrates the theory can be found in Section 4. The movie study on Sherlock fMRI dataset is given in Section 5.

2 Methodology

We observe a p-dimensional stationary Gaussian linear process $\{X_t\}_{t=1}^n$, where $X_t = \sum_{l=0}^{\infty} A_l \epsilon_{t-l}$ and $\epsilon_t \sim N(0, I_d)$. The dimension p is allowed to be much larger than n. The temporal dependence is captured by the decay rate of the coefficient matrices, as specified in the following condition.

Condition 1. *There exists a constant $\beta > 3/4$ and $C_0 > 0$ such that*

$$\max_{1 \le j \le p} |(A_l)_{j\cdot}| \le C_0 (1 \vee l)^{-\beta}, \qquad \beta > \frac{3}{4}. \tag{1}$$

A larger β corresponds to weaker temporal dependence of the process. $\beta > 1$ corresponds to short-range dependence (i.e., short memory). The more challenging regime is long-range dependence (i.e., long memory) for $\beta \le 1$, where the autocovariance matrices $\{\Gamma_k\}_{k=0}^{\infty}$ are no longer summable, thus the usual Central Limit Theorem for time series is not applicable.

Denote the covariance and precision matrices of the process by $\Sigma = \mathbb{E}[X_t X_t^\top]$ and $\Omega = \Sigma^{-1}$. For Gaussian processes, the support of the precision matrix $E = \mathrm{supp}(\Omega)$ induces a graph. We want to test whether it is identical to a hypothesized graph E_0. That is,

$$H_0 : E = E_0, \qquad H_1 : E \neq E_0.$$

2.1 Direct Statistic

We build our test statistic based on the sample covariance matrix, $\hat{\Sigma}_n = n^{-1} \sum_{t=1}^n X_t X_t^\top$, following the construction as in Le et al. (forthcoming). For each $1 \le j \le p$, let $S_{0,j} = \{i : i = j \text{ or } (i,j) \in E_0\}$ be the column-wise support set, with degree $s_{0,j} = |S_{0,j}|$. The sparsity level of the whole graph is measured by $s_0 = \max_{1 \le j \le p} s_{0,j}$. We first assume that the graph is reasonably sparse, in that $s_0 < n$. Let $B_{0,j} \in \mathbb{R}^{p \times s_{0,j}}$ be the corresponding selection matrix, such that $B_{0,j}^\top x$ extracts the coordinates of a vector $x \in \mathbb{R}^p$ indexed by $S_{0,j}$.

We write the j-th column of the precision matrix Ω_0 as $\boldsymbol{\omega}_{0,j}$. Under H_0, $\Sigma \boldsymbol{\omega}_{0,j} = e_j$ and $\boldsymbol{\omega}_{0,j}$ is zero outside $S_{0,j}$, we thus have $\boldsymbol{\omega}_{0,j} = B_{0,j}(B_{0,j}^\top \Sigma B_{0,j})^{-1} B_{0,j}^\top e_j$. Replacing Σ by $\hat{\Sigma}_n$ gives a local-inverse estimator

$$\hat{\boldsymbol{\omega}}_{0,j} = B_{0,j}(B_{0,j}^\top \hat{\Sigma}_n B_{0,j})^{-1} B_{0,j}^\top e_j. \tag{2}$$

Throughout this paper, whenever $B_{0,j}^\top \hat{\Sigma}_n B_{0,j}$ is singular, its inverse is understood as the zero matrix. Under the conditions stated later, this exceptional event has vanishing probability. Our proposed direct statistic is based on the fact that $\Sigma \Omega_0 = I$ on H_0. In particular, define the residual $\hat{r}_{ij} = e_i^\top \hat{\Sigma}_n \hat{\boldsymbol{\omega}}_{0,j} - \mathbf{1}\{i = j\}$. Since the graph test only concerns entries outside the null support, we focus on the off-support index set $\mathcal{J}_0 = \{(i,j) : i \notin S_{0,j}\}$. The direct statistic is defined as the root-scale maximum of the residuals over $\mathcal{J}_0$. Since $i \neq j$ on this set, it can be written as

$$\hat{T}_n = \sqrt{n} \max_{(i,j) \in \mathcal{J}_0} |e_i^\top \hat{\Sigma}_n \hat{\boldsymbol{\omega}}_{0,j}|. \tag{3}$$

When the true underlying graph E contains edges that are missing in the hypothesized graph E_0, we will obtain a larger $\hat{T}_n$, where the rejection region is situated. There are two caveats though. First, we need a data-adaptive method to determine the rejection region for $\hat{T}_n$ — the block bootstrap method in Section 2.2 addresses this concern. Second, while it is easier for $\hat{T}_n$ to reject E_0 when $E_0 \subsetneq E$, it is less capable of rejecting a hypothesized graph larger than the true graph — we propose a consistency-empowered corrected statistic in Section 2.3 that solves this issue.

2.2 Block Bootstrap

Despite the Gaussian approximation results on $\hat{T}_n$, its covariance structure is usually unknown in practice. We use the following block-bootstrap method is proposed to imitate the behavior of

Algorithm 1 Test procedure for the direct graph statistic

Require: $\{X_t\}_{t=1}^n$, null graph E_0, block length l, and level α

1: Compute sample covariance matrix $\hat{\Sigma}_n$ and the local inverse estimators $\hat{\boldsymbol{\omega}}_{0,j}$ from (2)
2: Compute the direct statistic $\hat{T}_n$ from (3)
3: Compute the plug-in score vectors $\hat{\mathcal{X}}_t$ from (4) for $t = 1, \ldots, n$
4: Form the overlapping block sums $\hat{B}_{r,l}$ for $r = l, \ldots, n$
5: Construct the empirical CDF $\hat{F}_{n,l}$ from (5) and the quantile $\hat{q}_{1-\alpha}$ from (6)
6: Reject H_0 if $\hat{T}_n > \hat{q}_{1-\alpha}$

the Gaussian approximation in a data-adaptive manner with a carefully chosen block length. For $(i,j) \in \mathcal{J}_0$, define the plug-in vector $\hat{\boldsymbol{u}}_{ij} = e_i - B_{0,j}(B_{0,j}^\top \hat{\Sigma}_n B_{0,j})^{-1} B_{0,j}^\top \hat{\Sigma}_n e_i$. For each $1 \le t \le n$ and $(i,j) \in \mathcal{J}_0$, define the plug-in score

$$\hat{\mathcal{X}}_{t,ij} = \hat{\boldsymbol{u}}_{ij}^\top (X_t X_t^\top - \hat{\Sigma}_n)\hat{\boldsymbol{\omega}}_{0,j}, \tag{4}$$

and collect these coordinates into the vector $\hat{\mathcal{X}}_t = (\hat{\mathcal{X}}_{t,ij})_{(i,j)\in\mathcal{J}_0}$. Because the map in (4) is linear in $X_t X_t^\top - \hat{\Sigma}_n$, we have $n^{-1}\sum_{t=1}^n \hat{\mathcal{X}}_t = 0$, hence no additional centering is needed. Let $l = l_n$ be the block length. For $r = l, \ldots, n$, define the overlapping block sums $\hat{B}_{r,l} = \sum_{t=r-l+1}^r \hat{\mathcal{X}}_t$. The empirical distribution function of the block statistics is

$$\hat{F}_{n,l}(u) = \frac{1}{n-l+1}\sum_{r=l}^n \mathbf{1}\left\{ l^{-1/2}|\hat{B}_{r,l}|_\infty \le u \right\}, \qquad u \in \mathbb{R}. \tag{5}$$

Define the bootstrap critical value by

$$\hat{q}_{1-\alpha} = \inf\{u \in \mathbb{R} : \hat{F}_{n,l}(u) \ge 1-\alpha\}. \tag{6}$$

The direct test rejects H_0 whenever $\hat{T}_n > \hat{q}_{1-\alpha}$; see Algorithm 1.

2.3 Consistency-empowered Correction

While $\hat{T}_n$ is sensitive to missing edges in E that are not included in E_0, it is not sensitive to false edges in E_0 that are not in E. To recover power against included alternatives, $E \subsetneq E_0$, we propose a corrected statistic $\tilde{T}_n$ similar to the consistency-empowered mechanism in Le et al. (forthcoming), but modified to adapt to the time series setup.

The detection of false edges in E_0 relies on the estimator in (2). For $i \in S_{0,j}$, if the estimated precision entry $\hat{\omega}_{0,ij}$ is too close to zero, we add an artificial pump C_n to it such that the test statistic is inflated. Define $\hat{\Delta}_{ij} = C_n \mathbf{1}\{|\hat{\omega}_{0,ij}|/\hat{\tau}_{ij} \le \delta_n\}$, where C_n and δ_n are tuning parameters, and set $\hat{\Delta}_{ij} = 0$ for $i \notin S_{0,j}$. Here, $\hat{\tau}_{ij}$ is the estimated long-run standard error of $\hat{\omega}_{0,ij}$ in the sampling space; see Appendix B.2 for detailed buildup. We replace the closed-form i.i.d. standard error of Le et al. (forthcoming) with an overlapping-block estimator of the long-run standard error and establish its uniform consistency in Lemma 1. Let $\hat{\boldsymbol{\Delta}}_j = (\hat{\Delta}_{1j}, \ldots, \hat{\Delta}_{pj})^\top$, and $\tilde{\boldsymbol{\omega}}_{0,j} = \hat{\boldsymbol{\omega}}_{0,j} + \hat{\boldsymbol{\Delta}}_j$. The consistency-empowered statistic is defined as

$$\tilde{T}_n = \sqrt{n}\max_{1\le i,j\le p}\left| e_i^\top \hat{\Sigma}_n \tilde{\boldsymbol{\omega}}_{0,j} - \mathbf{1}\{i=j\}\right|. \tag{7}$$

We shall see that for carefully calibrated C_n and δ_n, $\tilde{T}_n$ does an equally good job in detecting missing edges as $\hat{T}_n$ while ensuring the power of the test. The consistency-empowered test also relies on the same block-bootstrap critical value $\hat{q}_{1-\alpha}$ in (6). H_0 is rejected whenever $\tilde{T}_n > \hat{q}_{1-\alpha}$; see Algorithm 2 in Appendix A.2.

3 Theoretical Guarantee for the Proposed Statistics

3.1 Gaussian Approximation

For $(i,j) \in \mathcal{J}_0$, define $\boldsymbol{u}_{ij} = e_i - B_{0,j}(B_{0,j}^\top \Sigma B_{0,j})^{-1} B_{0,j}^\top \Sigma e_i$, and the linearized score process $\mathcal{X}_{t,ij} = \boldsymbol{u}_{ij}^\top (X_t X_t^\top - \Sigma)\boldsymbol{\omega}_{0,j}$. Then $\mathcal{X}_t = (\mathcal{X}_{t,ij})_{(i,j)\in\mathcal{J}_0}$ is the oracle version of $\hat{\mathcal{X}}_t$. The major theo-

retical finding of this work is that the distribution of the proposed statistic $\hat{T}_n$ is close to that of $|G|_\infty$, where G is a centered Gaussian vector in $\mathbb{R}^{|\mathcal{J}_0|}$ with covariance matrix $\Xi = \sum_{k=-\infty}^{\infty} \mathrm{Cov}(\mathcal{X}_0, \mathcal{X}_k)$.

Our proposed direct statistic $\hat{T}_n$ is based on local inverse of sample covariance matrix. The following proposition bridges the gap between two inference tasks, one on the statistic $\hat{T}_n$, the other on sample covariance error $n^{1/2}|\hat{\Sigma}_n - \Sigma|_\infty$ by showing that $\hat{T}_n$ is dominated by a linear term with respect to the covariance error. This result is a key prelude to the subsequent Gaussian approximation result on $\hat{T}_n$, essentially enabling high-dimensional Gaussian graphical model goodness-of-fit testing without the need of inference theory on the precision matrix. To see this, define the associated linearized oracle statistic as $T_n^{\mathrm{L}} = \left|n^{-1/2}\sum_{t=1}^{n} \mathcal{X}_t\right|_\infty$. The following technical conditions are needed.

Condition 2. *There exist positive constants C_Σ and C_Ω such that the true underlying covariance and precision matrices satisfy $|\Sigma|_1 \le C_\Sigma$ and $|\Omega|_1 \le C_\Omega$, respectively.*

Condition 3. *There exists a positive constant C_A such that $\max_{1\le j\le p} |(B_{0,j}^\top \Sigma B_{0,j})^{-1}|_1 \le C_A$.*

Examples satisfying Conditions 2–3 are given at the beginning of Appendix D. A first-order expansion of the local inverse gives the linearized statistic T_n^{L} plus a controllable residual.

Proposition 1. *Assume Conditions 2 and 3. On the event $2C_A s_0|\hat{\Sigma}_n - \Sigma|_\infty \le 1$, we have under H_0*

$$|\hat{T}_n - T_n^{\mathrm{L}}| \le C_1\sqrt{n}s_0^2|\hat{\Sigma}_n - \Sigma|_\infty^2,$$

where $C_1 > 0$ depends only on C_Σ, C_Ω, and C_A, and thus

$$\hat{T}_n \le C_1\sqrt{n}|\hat{\Sigma}_n - \Sigma|_\infty + C_1\sqrt{n}s_0^2|\hat{\Sigma}_n - \Sigma|_\infty^2.$$

The proof is deferred to Appendix D.1. This reduction bridges the gap between our proposed statistic and covariance error. The validity of the goodness-of-fit test is thus associated to the Gaussian approximation of $n^{1/2}|\hat{\Sigma}_n - \Sigma|_\infty$. The latter can be solved by an application of Theorem 1 by Zhai et al. (2026) for high-dimensional, long-memory time series. The following condition on the autocovariance matrices $\Gamma_k = \mathrm{Cov}(X_t, X_{t-k})$ is imposed to prevent the trivial scenario where the distribution of $|\hat{\Sigma}_n - \Sigma|_\infty$ is degenerate.

Condition 4. *There exists $c_0 > 0$ such that for all $1 \le s, t \le p$, $\sum_{k=-\infty}^{\infty}((\Gamma_k)_{ss}(\Gamma_k)_{tt} + (\Gamma_k)_{st}(\Gamma_k)_{ts}) \ge c_0$.*

Note that the quantity on the left-hand side is the long-run variance of $n^{1/2}\hat{\Sigma}_n$. In a similar spirit to Condition 4, we impose the following condition to prevent the case where the distribution of the score $\hat{T}_n$ is degenerate. This time, the long-run variance of the aforementioned score process $\mathcal{X}_t$, which is directly associated with the test statistic $\hat{T}_n$, shall be nondegenerate.

Condition 5. *There exists $c_1 > 0$ such that for all $(i, j) \in \mathcal{J}_0$, $\sum_{k=-\infty}^{\infty} \mathrm{Cov}(\mathcal{X}_{0,ij}, \mathcal{X}_{k,ij}) \ge c_1$.*

Under Conditions 1 and 4, Theorem 1 of Zhai et al. (2026) established a bound on the Kolmogorov distance for Gaussian approximation of the sample covariance error,

$$\sup_{u\in\mathbb{R}} \left|\mathbb{P}(n^{1/2}|\hat{\Sigma}_n - \Sigma|_\infty \ge u) - \mathbb{P}(|Z|_\infty \ge u)\right| \le C\Psi(p, n) := C\frac{\log^{(5\tilde{\beta}+12)/(4\tilde{\beta}+8)}(pn)}{n^{\tilde{\beta}/(4\tilde{\beta}+8)}},$$

where $\tilde{\beta} = (4\beta - 3) \wedge (2\beta - 1)$, and Z is a centered Gaussian vector with the same covariance structure as $n^{1/2}\,\mathrm{vec}(\hat{\Sigma}_n - \Sigma)$. Combining this result with Proposition 1, we obtain the following theorem for the finite-sample Gaussian approximation bound for our proposed direct statistic.

Theorem 1. *Assume Conditions 1-5. Then there exist positive constant C and c, depending only on C_0, c_0, c_1, C_Σ, C_Ω, and C_A, such that under H_0,*

$$\sup_{u\in\mathbb{R}} |\mathbb{P}(\hat{T}_n \le u) - \mathbb{P}(|G|_\infty \le u)| \le C\Psi(p, n) + Cp^2\exp\left(-\frac{cn}{s_0^2}\right) + C\frac{s_0^2\log^{3/2}(pn)}{\sqrt{n}}.$$

In particular, when the right-hand side converges to zero, then the Gaussian approximation is asymptotically valid. Specifically, the bound vanishes if $\log(pn) = o(n^{\tilde{\beta}/(5\tilde{\beta}+12)})$ and $s_0^2\log^{3/2}(pn)/\sqrt{n} \to 0$. Here, the second term is to control the probability of the event that Proposition 1 is not covered, and the third term accounts for the linerarization residual between T_n^{L} and $\hat{T}_n$. The proof of Theorem 1 is deferred to Appendix D.2.

3.2 Block Bootstrap

The Gaussian approximation theorem identifies the target null law, but that law is not directly usable because the covariance of G is unknown. Now we provide a fully data-driven calibration rule based on overlapping block bootstrap. The theorem below is what justifies the actual rejection rule used in practice.

The next proposition controls the gap between the plug-in score process and its oracle analogue. This is the extra approximation step that does not appear in the covariance bootstrap problem of Zhai et al. (2026). Let $\bar{\mathcal{X}}_n = \frac{1}{n}\sum_{t=1}^{n} \mathcal{X}_t$ and $M_n = \max_{1\le t\le n} |X_t|_\infty$.

Proposition 2. *Assume Conditions 2 and 3. On the event $2C_A s_0|\hat{\Sigma}_n - \Sigma|_\infty \le 1$, we have $\max_{(i,j)\in\mathcal{J}_0} |\hat{\boldsymbol{u}}_{ij} - \boldsymbol{u}_{ij}|_1 \le C_1' s_0 |\hat{\Sigma}_n - \Sigma|_\infty$. Moreover,*

$$\max_{1\le t\le n} \max_{(i,j)\in\mathcal{J}_0} |\hat{\mathcal{X}}_{t,ij} - (\mathcal{X}_{t,ij} - \bar{\mathcal{X}}_{n,ij})| \le C_1' s_0 |\hat{\Sigma}_n - \Sigma|_\infty (M_n^2 + 1). \tag{8}$$

Here $C_1' > 0$ depends only on C_A and C_Σ.

Proposition 2 identifies the second approximation error in the bootstrap theorem. Once this plug-in error is controlled, the bootstrap problem reduces to the oracle block-bootstrap problem for the linearized score process. The proof is deferred to Appendix D.3.

Theorem 2. *Assume Conditions 1-5. Consider any fixed $0 < \epsilon < 1$. Set $\phi = \frac{2\tilde{\beta}+4}{(3-\epsilon)\tilde{\beta}+4}$, $\psi = \frac{(5\tilde{\beta}+12)(1-\epsilon)}{(3-\epsilon)\tilde{\beta}+4}$, and $l \asymp n^{\phi}\log^{\psi}(p)$. Further define the aggregate bootstrap error rate*

$$\mathfrak{R}_{n,l}(\epsilon) = \frac{\log^{\frac{5\tilde{\beta}+12}{(6-2\epsilon)\tilde{\beta}+8}}(pn)}{n^{\frac{\tilde{\beta}}{(6-2\epsilon)\tilde{\beta}+8}}} + p^2\exp\left(-\frac{cn}{s_0^2}\right) + \frac{s_0^2\log^{3/2}(pn)}{\sqrt{n}} + s_0\sqrt{\frac{l}{n}}\log^2(pn).$$

Then under the null hypothesis H_0, there exist constants $C > 0$ depending only on C_0, c_0, c_1, C_Σ, C_Ω, and C_A, as well as $r_\epsilon \in (0,1]$ depending only on C and ϵ, such that whenever $\mathfrak{R}_{n,l}(\epsilon) \le r_\epsilon$, with probability no less than $1 - 3\mathfrak{R}_{n,l}(\epsilon)^\epsilon$,

$$\sup_{u\in\mathbb{R}} |\hat{F}_{n,l}(u) - \mathbb{P}(\hat{T}_n \le u)| \le C\mathfrak{R}_{n,l}(\epsilon)^{1-\epsilon}.$$

The proof of Theorem 2 is deferred to Appendix D.4. Theorem 2 provides a finite-sample bound. When $\mathfrak{R}_{n,l}(\epsilon) \to 0$, the block bootstrap method is then asymptotically valid. Under the null hypothesis, as shown in Appendix D.4,

$$\mathbb{P}_{H_0}(\hat{T}_n > \hat{q}_{1-\alpha}) \to \alpha.$$

In addition to the conditions stated after Theorem 1, the bound vanishes if $s_0\sqrt{l/n}\log^2(pn) \to 0$.

3.3 Theoretical Extensions

A natural extension of Theorems 1 and 2 is for two-sample testing. Consider that we observe two stationary Gaussian linear processes $\{X_t^{(1)}\}_{t=1}^{n_1}$ and $\{X_t^{(2)}\}_{t=1}^{n_2}$, and want to test whether the two underlying precision graphs, $E^{(1)}$ and $E^{(2)}$, are identical. We may, for instance, estimate $E^{(1)}$ by $\hat{E}^{(1)}$, and use it as the hypothesized graph for the test of $E^{(2)}$.

Corollary 3 (directional two-sample test, informal)**.** *The Gaussian approximation result in Theorem 1 and the block bootstrap validity guarantee in Theorem 2 still apply to this two-sample test, but with an extra term related to $\pi_{n_1} = \mathbb{P}(\hat{E}^{(1)} \ne E^{(1)})$ in the upper bound, respectively.*

The detailed testing setup and the formal statement of the corollary is deferred to Appendix B.1. Thresholded CLIME provides a support-consistent graph estimator under the conditions given in Appendix B.3. A symmetric test of graph equality is obtained by defining $p_{\mathrm{sym}} = \min\{1, 2\min(p_{1\to2}, p_{2\to1})\}$, where $p_{1\to2}$ and $p_{2\to1}$ are the two directional p-values. With support-consistent graph selectors for both samples, the Bonferroni argument gives asymptotic level at most α without requiring independence between the directional p-values; see Appendix B.1.

On the other hand, we also provide a brief, informal theorem for the consistency-empowered estimator $\tilde{T}_n$ in (7). In theory, if we fix the two tuning parameters as $\delta_n = \sqrt{\log n}$ and $C_n = C_\Delta \sqrt{\log p}$ for some fixed constant $C_\Delta > 0$, then the artificial pump is strong enough for $\tilde{T}_n$ to be rejected when E_0 contains false edges in the true underlying graph E, thus ensuring the power of the test, without affecting the size of the test. Under the same assumptions, $\delta_n = \sqrt{\log n}$ may be replaced by any sequence satisfying $\delta_n \to \infty$ and $\delta_n + \sqrt{\log p} = o(\sqrt{n})$.

Theorem 4 (consistency-empowered estimator, informal)**.** *Assume that the conditions in Theorems 1 and 2 hold. Let $0 < \epsilon < 1$, choose the block length l as in Theorem 2, and assume $\mathfrak{R}_{n,l}(\epsilon) \to 0$. Then*

1. *Under $H_0 : E = E_0$, $\mathbb{P}_{H_0}(\tilde{T}_n = \hat{T}_n) \to 1$. Consequently, Theorems 1 and 2 remain valid with $\tilde{T}_n$ in place of $\hat{T}_n$, and $\mathbb{P}_{H_0}(\tilde{T}_n > \hat{q}_{1-\alpha}) \to \alpha$.*
2. *Under fixed alternative $H_1 : E \neq E_0$ subject to some conditions, $\mathbb{P}_{H_1}(\tilde{T}_n > \hat{q}_{1-\alpha}) \to 1$.*

The detailed technical discussion and the formal statement of the theorem (i.e. Theorem 6) is deferred to Appendix B.2. The proof combines long-run standard-error estimation with coordinate screening; see Appendix E.4.

4 Quantitative Study

4.1 Null Test Validity

The first quantitative study aims at demonstrating Gaussian approximation result in Theorem 1 and the bootstrap validity guarantee in Theorem 2. We use the a similar high-dimensional sample-size and dimension regimes as in Zhai et al. (2026), i.e., $n \in \{200, 400, 800\}$ and $p \in \{2000, 4000, 8000\}$. For the temporal dependence, we consider three levels, $\beta \in \{2, 0.9, 0.55\}$. The case $\beta = 2$ represents short memory, $\beta = 0.9$ represents long memory covered by the theory, and $\beta = 0.55$ is included as an out-of-theory ultra-long-memory case.

The null graph E_0 is taken to be a banded graph with fixed sparsity level, $E_0 = \{(i, j) : 1 \leq |i-j| \leq b\}$. In the original null-validity experiment, we set $b = 4$. We first construct a banded baseline precision matrix Ω_0 supported on E_0. For $i \neq j$, set $(\Omega_0)_{ij} = \theta_0 \rho^{|i-j|-1} \mathbf{1}\{1 \leq |i-j| \leq b\}$, where $\theta_0 = 0.25$ and $\rho = 0.5$. The diagonal entries are chosen by strict diagonal dominance: $(\Omega_0)_{ii} = 1 + \sum_{j \neq i} |(\Omega_0)_{ij}| + \delta_0$, where we set $\delta_0 = 0.10$. This guarantees that Ω_0 is positive definite and that its off-diagonal support is exactly E_0.

For each precision matrix, we generate 200 copies of time series (see Appendix C.1 for details). For each copy, we can compute a realization of $\hat{\Sigma}_n$ and the statistic $\hat{T}_n$. Meanwhile, we may compute the covariance matrix Ξ and generate copies of the Gaussian vector G. Based on each time series, we may also compute the bootstrap block sum $\hat{B}_{r,l}$. Here we choose $l = n^{2/3}$, same as in Zhai et al. (2026). This choice is a convenient practical substitute for the β-dependent block-length order in Theorem 2. To demonstrate the validity of the Gaussian approximation result in Theorem 1 or 2, we may compare the distribution of $\hat{T}_n$ against that of $|G|_\infty$ or $l^{-1/2}|\hat{B}_{r,l}|_\infty$, respectively.

The QQ plots for the three temporal dependence levels are given in Figures 1, 2, and 3 (Fig. 3 in Appendix C.2). As a supplement, the corresponding Kolmogorov and Wasserstein-1 distances are given in Tables 3 and 4 (both in Appendix C.2). Generally, the shorter-memory, the larger n and smaller p, the more accurate Gaussian approximation and block bootstrap are. This agrees with the trend of finite-sample bounds in Theorems 1 and 2. This simulation setup confirms that they have reasonable empirical validity even when p is a lot larger than n, and when the temporal dependence is long-range. The case where $\beta = 0.55$, which exceeds the range in Condition 1, fails expectedly.

We further examine the effect of sparsity by fixing $n = 800$ and $p = 2000$ and varying $b \in \{2, 4, 6, 8, 12, 16\}$ under $\beta = 2$ and 0.9. The corresponding results in Appendix C.2 show that approximation accuracy and null calibration generally deteriorate at larger $s_0 = 2b + 1$, especially under long memory.

To examine the behavior in the long memory regime, we also add $\beta = 0.85$ and 0.8. We fix $b = 4$ and $p = 2000$, using $n = 800$ and 1600. The results in Appendix C.2 show deteriorating bootstrap calibration near $\beta = 3/4$ and improvement in these near-boundary settings when n is doubled.

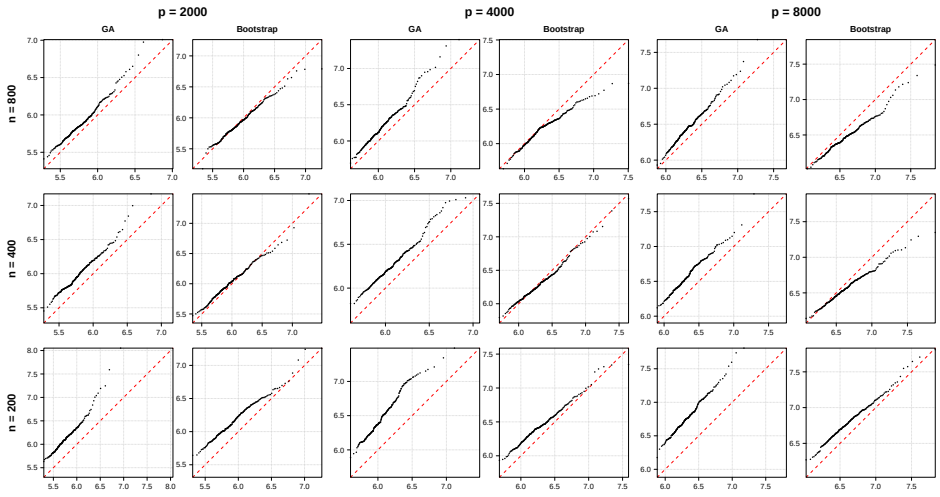

Figure 1: QQ-plots for null tests under short-memory, $\beta = 2$. The x-axis is $|G|_\infty$ for GA and $l^{-1/2}|\hat{B}_{r,l}|_\infty$ for bootstrap. The y-axis is $\hat{T}_n$ for both. The red dashed line is $y = x$. Closer to this line indicates better proximity.

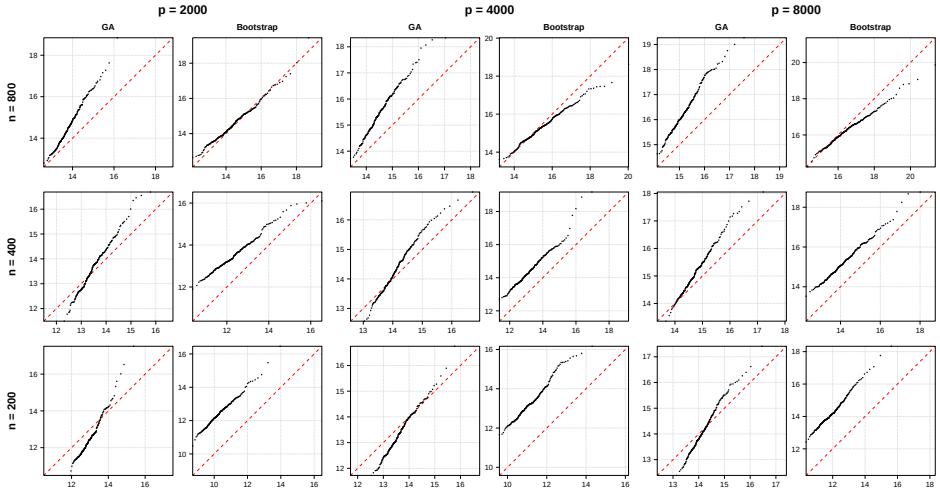

Figure 2: QQ-plots for null tests under long-memory, $\beta = 0.9$. The x-axis is $|G|_\infty$ for GA and $l^{-1/2}|\hat{B}_{r,l}|_\infty$ for bootstrap. The y-axis is $\hat{T}_n$ for both. The red dashed line is $y = x$. Closer to this line indicates better proximity.

4.2 Alternative Test Power

We next study the power of the proposed graph test against random off-band alternatives. For the null graph E_0, we apply the same banded structure and precision matrix setup as the null test experiment in Section 4.1. For the true precision matrix Ω^* used to generate time series, however, we add m off-support entries sampled uniformly without replacement from $\{(i,j) : i < j,\ |i-j| > b\}$. In this experiment, we consider the low-dimensional setup $n = 2000, p = 800$ and the high-dimensional setup $n = 800, p = 2000$. For each setup, we consider the short-memory regime $\beta = 2$ and the long-memory regime $\beta = 0.9$. The strength of the planted edges is calibrated by the effective strength, denoted by $\kappa = |\Omega^*_{ij}|/\sqrt{\Omega^*_{ii}\Omega^*_{jj}}$ for the added edges. In this experiment, we let $m \in \{80, 200, 400, 800, 1200\}$ for the low-dimensional setup and $m \in \{200, 500, 1000, 2000, 3000\}$ for the high-dimensional setup, and $\kappa \in \{0.03, 0.06, 0.09, 0.12, 0.15, 0.18\}$. We test the null graph E_0 on each of the 200 instances of FFT-generated time series at the level $\alpha = 0.1$, and report the rejection rate.

Table 1: Empirical rejection rates under random off-band alternatives.

Low-dimensional regime: $(n, p) = (2000, 800)$

$\beta = 2$							$\beta = 0.9$						
$m\backslash\kappa$	0.03	0.06	0.09	0.12	0.15	0.18	$m\backslash\kappa$	0.03	0.06	0.09	0.12	0.15	0.18
80	0.135	0.375	0.995	1.000	1.000	1.000	80	0.090	0.100	0.085	0.100	0.195	0.335
200	0.140	0.650	1.000	1.000	1.000	1.000	200	0.090	0.100	0.115	0.155	0.230	0.520
400	0.145	0.840	1.000	1.000	1.000	1.000	400	0.095	0.120	0.145	0.210	0.390	0.770
800	0.165	0.935	1.000	1.000	1.000	1.000	800	0.110	0.125	0.155	0.270	0.485	0.680
1200	0.175	0.955	1.000	1.000	1.000	1.000	1200	0.105	0.135	0.210	0.320	0.465	0.550

High-dimensional regime: $(n, p) = (800, 2000)$

$\beta = 2$							$\beta = 0.9$						
$m\backslash\kappa$	0.03	0.06	0.09	0.12	0.15	0.18	$m\backslash\kappa$	0.03	0.06	0.09	0.12	0.15	0.18
200	0.085	0.115	0.175	0.770	1.000	1.000	200	0.075	0.090	0.090	0.110	0.110	0.125
500	0.100	0.140	0.355	0.945	1.000	1.000	500	0.090	0.085	0.105	0.145	0.160	0.165
1000	0.085	0.105	0.475	0.970	1.000	1.000	1000	0.095	0.095	0.130	0.160	0.180	0.195
2000	0.050	0.120	0.585	0.990	1.000	1.000	2000	0.070	0.100	0.130	0.160	0.190	0.205
3000	0.070	0.175	0.665	0.965	0.990	1.000	3000	0.085	0.105	0.160	0.215	0.260	0.295

Table 1 summarizes the resulting rejection probabilities. The weak-signal columns stay close to the nominal level, while rejection rates generally increase as the effective edge strength κ grows, or as the number of extra edges m grows. A larger number of stronger extra edges are usually more easily detected. In the short-memory case, the power converges to one, as described in theory. This convergence is slower for longer memory or higher dimension, as explained by the slower rates in Theorem 2.

We show another simulation that demonstrates the power of consistency-empowered statistic $\tilde{T}_n$. We apply the same banded precision matrix structure Ω_0 with $b = 4$, but with $\theta_0 = 1$, and $\rho = 1$. This setup is to prevent the precision entries from being too small such that the assumption in Theorem 6 is violated. For the true precision matrix Ω^* used to generate time series, we randomly select η of all entries in Ω_0, and these entries are weakened by the ratio of χ. The experiment is on the same (n, p, β) combinations. The tuning parameters δ_n and C_Δ are altered in a grid on a heldout copy of time series, until the null test rejection rate approach the nominal size $\alpha = 0.10$. For our setup, the choice is $(\delta_n, C_\Delta) = (0.05, 0.5)$, with $C_n = C_\Delta\sqrt{\log p}$. As a side-by-side comparison, the tests are also run under the direct statistic $\hat{T}_n$ under each setup. The results are in Table 2.

The consistency-empowered tests are indeed recovering the edges in Ω_0 that are missing in Ω^*, compared to the direct tests. This phenomenon is more prominent when the temporal dependence is weaker and the dimension is lower. For the short-memory, low-dimensional case, in particular, we can see the rejection rate rapidly converging to one as η and χ increases. The rejection rate for the short-memory, high-dimensional case is also growing, but not as rapidly. In both cases, the consistency-empowered rejection rates are significantly higher than their direct test counterparts. This demonstrates the recovered power of $\tilde{T}_n$ as $E \subsetneq E_0$. In the long-memory case, however, the growth is significantly slower, and the advantage upon direct test is far less significant. This is unsurprising, since the rate in Theorem 6, which directly depends on $\mathfrak{R}_{n,l}(\epsilon)$, converges significantly slower as the temporal dependence becomes stronger.

5 Sherlock fMRI Data Analysis

We apply our proposed test on the Sherlock fMRI dataset (Chen et al., 2017). Each of the 17 subjects were observed in two phases. In the movie phase, the subject watched the same movie as a stimulus. In the recall phase, the subject verbally recalled the movie content. The fMRI scan in each phase was transformed to multivariate time series over 268 cortical regions of interest (ROI, Shen et al. (2013)).

Table 2: Empirical rejection rates under within-band weakening alternatives.

(η, χ)	$n = 2000,\ p = 800$				$n = 800,\ p = 2000$			
	$\beta = 2$		$\beta = 0.9$		$\beta = 2$		$\beta = 0.9$	
	$\tilde{T}_n$	$\hat{T}_n$	$\tilde{T}_n$	$\hat{T}_n$	$\tilde{T}_n$	$\hat{T}_n$	$\tilde{T}_n$	$\hat{T}_n$
$(0.0, 0.0)$	**0.140**	0.130	**0.080**	0.070	**0.070**	0.070	**0.110**	0.090
$(0.1, 0.5)$	**0.835**	0.135	**0.140**	0.125	**0.130**	0.125	**0.120**	0.090
$(0.3, 1.0)$	**1.000**	0.165	**0.100**	0.085	**0.270**	0.115	**0.110**	0.085
$(0.5, 1.0)$	**1.000**	0.150	**0.100**	0.100	**0.350**	0.110	**0.120**	0.100

The scientific question is whether the conditional independence graph structure during the movie phase is sufficient to explain the recall phase. This comparison is natural because movie watching and recall involve related cognitive content, but may entail different patterns of brain connectivity (Baldassano et al., 2017). Our methodology is useful for this type of data, since fMRI time series typically exhibit long memory (see Appendix C.4). For each subject, the movie phase contains 1976 time points, while the recall phase has subject-specific length ranging from 437 to 1824 time points.

The goal of the experiment is a directional two-sample test on the conditional-independence graph between the movie phase and the recall phase. We first estimate the movie-phase graph as follows. For each subject, compute $\hat{\Sigma}^{\mathrm{mov}}$, the movie-phase sample covariance matrix, and let $\hat{\Omega}^{\mathrm{mov}} = (\hat{\Sigma}^{\mathrm{mov}})^{-1}$. For threshold $\nu > 0$, which is a tuning parameter, let the movie graph be $E_\nu^{\mathrm{mov}} = \{(i, j) : i < j,\ |\hat{\Omega}_{ij}^{\mathrm{mov}}| > \nu\}$. We select the largest value of ν such that the graph passes the movie self-diagnostic on the individual's movie-phase data at level $\alpha = 0.05$. Equivalently, this selects the sparsest graph in the threshold path whose movie self-test p-value is at least 0.05.

We then test the recall phase against this movie-phase graph $E_{\hat{\nu}}^{\mathrm{mov}}$ using $\hat{T}_n$. A rejection indicates that the recall graph contains off-support conditional dependencies not contained in the selected movie graph, signaling extra brain connectivity during the recall phase. The recall p-values are computed by the block-bootstrap method, and the 17 subject-level recall p-values are adjusted using the Holm procedure to control family-wise error rate. Under the conditions of Corollary 5, support-consistent graph selection preserves asymptotic validity of the recall p-values; see Appendix B.1. For each selected graph, we report the average column-wise support size $\bar{s}$ and the quartiles of node degrees. The result is available in Table 7 in Appendix C.3.

All 17 selected movie graphs pass the self-diagnostic at level $\alpha = 0.05$. After Holm correction across subjects, all 17 recall tests reject at level $\alpha = 0.05$. Thus, the recall phase exhibits conditional-dependence structure that is not contained in the corresponding selected movie-phase graph for every subject in the cohort. This result provides evidence of a structural difference of brain connectivity between watching a movie (receiving information) and recalling it (utilizing memory and describing), relative to the selected movie graphs.

6 Concluding Remark

This paper develops a framework for testing Gaussian graphical model structure in high-dimensional long-memory time series. The key contribution is to reduce the graph specification testing problem to simultaneous inference of covariance through local-inverse based on the sample covariance matrix. This avoids full precision-matrix estimation and enables finite-sample Gaussian approximation and block-bootstrap calibration. The theory relies on temporal dependence $\beta > 3/4$. For ultra-long-memory time series with $1/2 < \beta < 3/4$, the asymptotic distribution can be Rosenblatt-type, and is beyond the scope of this paper. Another limitation is that the one-sample test requires a predetermined graphical structure to be known. Entrywise testing requires inference theory on the precision matrix. Doing so under high dimension is a very interesting yet challenging aspect of study, and we leave it for future work.

Acknowledgments and Disclosure of Funding

Wei Biao Wu's research was partially supported by NSF DMS-2311249. Ping-Shou Zhong's research was partially supported by NSF DMS-2152070.

References

Baek, C., Düker, M.-C., and Pipiras, V. (2023). Local whittle estimation of high-dimensional long-run variance and precision matrices. *The Annals of Statistics*, 51(6):2386–2414.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., and Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, 95(3):709–721.e5.

Cai, T. T., Liu, W., and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., and Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, 20(1):115–125.

Chen, X., Xu, M., and Wu, W. B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics*, 41(6):2994–3021.

Cheng, G., Zhang, Z., and Zhang, B. (2017). Test for bandedness of high-dimensional precision matrices. *Journal of Nonparametric Statistics*, 29(4):884–902.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields*, 162(1–2):47–70.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Janková, J. and van de Geer, S. (2015). Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics*, 9(1):1205–1229.

Lauritzen, S. L. (1996). *Graphical Models*. Number 17 in Oxford Statistical Science Series. Clarendon Press, Oxford.

Le, T.-M., Zhong, P.-S., and Leng, C. (forthcoming). Universally consistent tests for the graph of a gaussian graphical model. *Statistica Sinica*. Preprint SS-2023-0385, to appear.

Lin, X., Li, W., Tian, F., and Huang, D. (2025). Goodness-of-fit tests for high-dimensional gaussian graphical models via exchangeable sampling. *Journal of the Royal Statistical Society: Series B*, 87(5):1619–1635.

Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics*, 41(6):2948–2978.

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

Qiu, H., Han, F., Liu, H., and Caffo, B. (2016). Joint estimation of multiple graphical models from high dimensional time series. *Journal of the Royal Statistical Society: Series B*, 78(2):487–504.

Ren, Z., Sun, T., Zhang, C.-H., and Zhou, H. H. (2015). Asymptotic normality and optimalities in estimation of large gaussian graphical models. *The Annals of Statistics*, 43(3):991–1026.

Shen, X., Tokoglu, F., Papademetris, X., and Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage*, 82:403–415.

Shu, H. and Nan, B. (2019). Estimation of large covariance and precision matrices from temporally dependent observations. *The Annals of Statistics*, 47(3):1321–1350.

Verzelen, N. and Villers, F. (2009). Tests for gaussian graphical models. *Computational Statistics & Data Analysis*, 53(5):1894–1905.

Xia, Y., Cai, T., and Cai, T. T. (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika*, 102(2):247–266.

Zhai, P. S., Kolar, M., and Wu, W. B. (2026). Simultaneous inference for covariance and precision matrices of long-range dependent time series. *IEEE Transactions on Information Theory*, 72(6):4246–4296.

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

Zhang, X. and Cheng, G. (2014). Bootstrapping high dimensional time series. arXiv preprint arXiv:1406.1037.

Zhao, B., Wang, Y. S., and Kolar, M. (2022). Fudge: A method to estimate a functional differential graph in a high-dimensional setting. *Journal of Machine Learning Research*, 23(82):1–82.

Zhao, B., Zhai, P. S., Wang, Y. S., and Kolar, M. (2024). High-dimensional functional graphical model structure learning via neighborhood selection approach. *Electronic Journal of Statistics*, 18(1):1042 – 1129.

Algorithm 2 Test procedure for the consistency-empowered statistic

Require: $\{X_t\}_{t=1}^n$, null graph E_0, block length l, level α, and tuning parameters C_n, δ_n

1: Compute sample covariance matrix $\hat{\Sigma}_n$ and the local inverse estimators $\hat{\boldsymbol{\omega}}_{0,j}$ from (2)
2: For each j and $i \in S_{0,j}$, compute $\hat{\boldsymbol{v}}_{0,ij} = B_{0,j}(B_{0,j}^\top \hat{\Sigma}_n B_{0,j})^{-1} B_{0,j}^\top e_i$ and the plug-in estimated local inverse score process $\hat{\mathcal{X}}_{t,ij}^{\omega}$ from (9)
3: Compute its block-bootstrap estimated long-run variance, $\hat{L}_{ij}^{\omega}$, from (10), and set $\hat{\tau}_{ij} = \sqrt{\hat{L}_{ij}^{\omega}/n}$
4: Define $\hat{\Delta}_{ij} = C_n \mathbf{1}\{|\hat{\omega}_{0,ij}|/\hat{\tau}_{ij} \le \delta_n\}$ for $i \in S_{0,j}$, and set $\hat{\Delta}_{ij} = 0$ for $i \notin S_{0,j}$.
5: Let $\hat{\boldsymbol{\Delta}}_j = (\hat{\Delta}_{1j}, \ldots, \hat{\Delta}_{pj})^\top$ and compute $\tilde{\boldsymbol{\omega}}_{0,j} = \hat{\boldsymbol{\omega}}_{0,j} + \hat{\boldsymbol{\Delta}}_j$
6: Compute $\tilde{T}_n$ from (7)
7: Compute the plug-in score vectors $\hat{\mathcal{X}}_t$ from (4) for $t = 1, \ldots, n$
8: Form the overlapping block sums $\hat{B}_{r,l}$ for $r = l, \ldots, n$
9: Construct the empirical CDF $\hat{F}_{n,l}$ from (5) and the quantile $\hat{q}_{1-\alpha}$ from (6)
10: Reject H_0 if $\tilde{T}_n > \hat{q}_{1-\alpha}$

A Miscellaneous Contents

A.1 Notation

Throughout the paper, $|\cdot|_\infty$ denotes the maximum absolute entry of a vector or a matrix, and $|\cdot|_1$ denotes the matrix ℓ_1 norm. For a vector v, $|v|$ denotes its Euclidean norm. For a matrix A, $A_{j\cdot}$ denotes its j-th row, $A_{\cdot j}$ denotes its j-th column, and $A^\top$ denotes its transpose. We write $[p] = \{1, \ldots, p\}$, and let e_j be the j-th canonical basis vector in $\mathbb{R}^p$. For a matrix A, $\mathrm{supp}(A) = \{(i,j) : A_{ij} \neq 0\}$ denotes its entrywise support. For a random variable X, $\|X\|_q$ is defined as $(\mathbb{E}|X|^q)^{1/q}$. For a set S, $|S|$ denotes its cardinality. We use $\mathbf{1}\{\cdot\}$ for the indicator function. For two positive sequences a_n and b_n, $a_n \lesssim b_n$ means $a_n \le C b_n$ for some constant $C > 0$, and $a_n \asymp b_n$ means $a_n \lesssim b_n$ and $b_n \lesssim a_n$. We use $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. The notation $O_{\mathbb{P}}(\cdot)$ and $o_{\mathbb{P}}(\cdot)$ has its usual meaning. Unless otherwise stated, constants may change from line to line and do not depend on n or p.

A.2 Algorithms

This paper introduces two types of goodness-of-fit test statistics for Gaussian graphical models of time series. The first one is the direct statistic $\hat{T}_n$ as discussed in Section 2.1, while the second one is the consistency-empowered corrected statistic $\tilde{T}_n$ as discussed in Section 2.3.

Algorithm 1 describes the detailed algorithm that computes $\hat{T}_n$, and how to compute the rejection threshold $\hat{q}_{1-\alpha}$ using block bootstrap method. Meanwhile, Algorithm 2 describes the detailed algorithm that computes $\tilde{T}_n$, and uses the same rejection threshold $\hat{q}_{1-\alpha}$ computed by the same block bootstrap method.

B Supplemental Theory

B.1 Detailed Discussion on Two-sample Extension

We first record a direct two-sample consequence of Theorems 1 and 2. Let $\{X_t^{(1)}\}_{t=1}^{n_1}$ and $\{X_t^{(2)}\}_{t=1}^{n_2}$ be two possibly dependent processes, each marginally satisfying the stationary Gaussian linear-process setup of Section 2. For $k = 1, 2$, the true underlying covariance and precision matrices are $\Sigma^{(k)} = \mathbb{E}[X_t^{(k)} X_t^{(k)\top}]$ and $\Omega^{(k)} = (\Sigma^{(k)})^{-1}$, with the graph $E^{(k)} = \mathrm{supp}(\Omega^{(k)})$. We also have each sample covariance $\hat{\Sigma}^{(k)}$. Let $\hat{E}^{(1)}$ be a graph estimator based only on the first sample. We use $\hat{E}^{(1)}$ as the null graph for the second sample and test

$$H_0^{1\to 2} : E^{(2)} \subseteq E^{(1)}, \qquad H_1^{1\to 2} : E^{(2)} \nsubseteq E^{(1)}.$$

This construction is a directional support-containment test because the second sample is tested against the first-sample graph. Let $S_{0,j}^{(1)}$ and $B_{0,j}^{(1)}$ denote the support set and selection matrix constructed

from the fixed population graph $E^{(1)}$. Under $H_0^{1\to2}$, the jth column $\boldsymbol{\omega}_j^{(2)}$ of $\Omega^{(2)}$ is supported on $S_{0,j}^{(1)}$, so

$$B_{0,j}^{(1)}\big(B_{0,j}^{(1)\top}\Sigma^{(2)}B_{0,j}^{(1)}\big)^{-1}B_{0,j}^{(1)\top}e_j=\boldsymbol{\omega}_j^{(2)}.$$

Thus the fixed-reference population local inverses and graph scores use $\Sigma^{(2)}$ and the reference graph $E^{(1)}$. The direct-statistic proofs of Theorems 1 and 2 use only this support containment and the resulting off-support moment identities, so their bounds extend from graph equality to this containment null under the same matrix-norm, local-inverse stability, and selected-score nondegeneracy conditions.

For each $1\le j\le p$, define $\hat S_{0,j}^{(1)}=\{i:i=j \text{ or } (i,j)\in\hat E^{(1)}\}$, $\hat s_{0,j}^{(1)}=|\hat S_{0,j}^{(1)}|$ and $\hat s_0^{(1)}=\max_{1\le j\le p}\hat s_{0,j}^{(1)}$. Let $\hat B_{0,j}^{(1)}\in\mathbb{R}^{p\times\hat s_{0,j}^{(1)}}$ be the corresponding selection matrix. Define

$$\hat{\boldsymbol{\omega}}_{0,j}^{(1)}=\hat B_{0,j}^{(1)}(\hat B_{0,j}^{(1)\top}\hat\Sigma^{(2)}\hat B_{0,j}^{(1)})^{-1}\hat B_{0,j}^{(1)\top}e_j.$$

Let $\hat{\mathcal{J}}^{(1)}=\{(i,j):i\notin\hat S_{0,j}^{(1)}\}$. The directional two-sample statistic is

$$\hat T^{1\to2}=\sqrt{n_2}\max_{(i,j)\in\hat{\mathcal{J}}^{(1)}}|e_i^\top\hat\Sigma^{(2)}\hat{\boldsymbol{\omega}}_{0,j}^{(1)}|.$$

The bootstrap distribution is formed as in Section 2.2, but using the second sample and the random null graph $\hat E^{(1)}$. For $(i,j)\in\hat{\mathcal{J}}^{(1)}$, define $\hat{\boldsymbol{u}}_{ij}^{(1)}=e_i-\hat B_{0,j}^{(1)}(\hat B_{0,j}^{(1)\top}\hat\Sigma^{(2)}\hat B_{0,j}^{(1)})^{-1}\hat B_{0,j}^{(1)\top}\hat\Sigma^{(2)}e_i$. For $1\le t\le n_2$, define $\hat{\mathcal{X}}_{t,ij}^{(1)}=\hat{\boldsymbol{u}}_{ij}^{(1)\top}(X_t^{(2)}X_t^{(2)\top}-\hat\Sigma^{(2)})\hat{\boldsymbol{\omega}}_{0,j}^{(1)}$ and $\hat{\mathcal{X}}_t^{(1)}=(\hat{\mathcal{X}}_{t,ij}^{(1)})_{(i,j)\in\hat{\mathcal{J}}^{(1)}}$. For a block length $l=l_{n_2}$, let $\hat B_{r,l}^{(1)}=\sum_{t=r-l+1}^{r}\hat{\mathcal{X}}_t^{(1)}$ for $r=l,\dots,n_2$. The directional bootstrap CDF and critical value are

$$\hat F_l^{1\to2}(u)=\frac{1}{n_2-l+1}\sum_{r=l}^{n_2}\mathbf{1}\left\{l^{-1/2}|\hat B_{r,l}^{(1)}|_\infty\le u\right\},\quad \hat q_{1-\alpha}^{1\to2}=\inf\{u\in\mathbb{R}:\hat F_l^{1\to2}(u)\ge1-\alpha\}.$$

The directional two-sample test rejects $H_0^{1\to2}$ whenever $\hat T^{1\to2}>\hat q_{1-\alpha}^{1\to2}$. Its bootstrap p-value is $p_{1\to2}=1-\hat F_l^{1\to2}(\hat T^{1\to2}-)$, where $F(x-)$ denotes the left limit of F at x. If the selected off-support set is empty, its maximum is defined as zero and the p-value is one. Let $s_0^{(1)}=\max_{1\le j\le p}|\{i:i=j \text{ or } (i,j)\in E^{(1)}\}|$ be the sparsity level of $E^{(1)}$. Let $G^{1\to2}$ be the centered Gaussian vector obtained by applying Theorem 1 to the second sample with null graph $E^{(1)}$. The following corollary of Theorems 1 and 2 is the rigorous version of Corollary 3.

Corollary 5 (Directional two-sample extension)**.** *Let $\pi_{n_1}=\mathbb{P}(\hat E^{(1)}\ne E^{(1)})$ and assume $\pi_{n_1}\to0$. Under $H_0^{1\to2}$, assume that the second sample satisfies the assumptions of Theorems 1 and 2 with the fixed reference graph $E^{(1)}$, using the containment interpretation above and a nonempty fixed off-support set. Then*

1. *The following Gaussian approximation bound holds:*

$$\sup_{u\in\mathbb{R}}|\mathbb{P}_{H_0^{1\to2}}(\hat T^{1\to2}\le u)-\mathbb{P}(|G^{1\to2}|_\infty\le u)|$$
$$\le C\pi_{n_1}+C\Psi(p,n_2)+Cp^2\exp\{-cn_2/(s_0^{(1)})^2\}+C\frac{(s_0^{(1)})^2\log^{3/2}(pn_2)}{\sqrt{n_2}}.$$

2. *For any fixed $0<\epsilon<1$, take $\phi=\frac{2\tilde\beta+4}{(3-\epsilon)\tilde\beta+4}$, $\psi=\frac{(5\tilde\beta+12)(1-\epsilon)}{(3-\epsilon)\tilde\beta+4}$, and $l\asymp n_2^\phi\log^\psi(p)$. Whenever $\mathfrak{R}_{n,l}^{1\to2}(\epsilon)\le1$, we have*

$$\mathbb{P}(\sup_{u\in\mathbb{R}}|\hat F_l^{1\to2}(u)-\mathbb{P}_{H_0^{1\to2}}(\hat T^{1\to2}\le u)|>C\mathfrak{R}_{n,l}^{1\to2}(\epsilon)^{1-\epsilon})\le C\mathfrak{R}_{n,l}^{1\to2}(\epsilon)^{\epsilon},$$

where $\Psi_B(p,n,\epsilon)=\log^{(5\tilde\beta+12)/[(6-2\epsilon)\tilde\beta+8]}(pn)/n^{\tilde\beta/[(6-2\epsilon)\tilde\beta+8]}$ is the first term of the rate in Theorem 2, and

$$\mathfrak{R}_{n,l}^{1\to2}(\epsilon)=\pi_{n_1}+\Psi_B(p,n_2,\epsilon)+\Psi(p,n_2)+p^2\exp\{-cn_2/(s_0^{(1)})^2\}+(pn_2)^{-1}$$
$$+\frac{(s_0^{(1)})^2\log^{3/2}(pn_2)}{\sqrt{n_2}}+s_0^{(1)}\sqrt{\frac{l}{n_2}}\log^2(pn_2).$$

In particular, if $\mathfrak{R}_{n,l}^{1\to 2}(\epsilon) \to 0$*, and the distribution of* $|G^{1\to 2}|_\infty$ *is continuous at its* $(1-\alpha)$*-quantile, then we have*

$$\mathbb{P}_{H_0^{1\to 2}}(\hat{T}^{1\to 2} > \hat{q}_{1-\alpha}^{1\to 2}) \to \alpha.$$

The proof of Corollary 5 is given in Appendix E.1.

Construct $p_{2\to 1}$ by interchanging the two samples, and define

$$p_{\mathrm{sym}} = \min\{1, 2\min(p_{1\to 2}, p_{2\to 1})\}.$$

Since $E^{(1)} = E^{(2)}$ is equivalent to containment in both directions, this tests graph equality by rejecting whenever either directional test rejects at level $\alpha/2$. Suppose that $\pi_{n_k} = \mathbb{P}(\hat{E}^{(k)} \neq E^{(k)}) \to 0$ for $k = 1, 2$ and, under the equality null, each tested sample marginally satisfies the assumptions of Corollary 5 with the other population graph as reference, with the corresponding oracle error rates tending to zero. Then, for each fixed $0 < \alpha < 1$,

$$\mathbb{P}_{E^{(1)}=E^{(2)}}(p_{\mathrm{sym}} \le \alpha) \le \alpha + \pi_{n_1} + \pi_{n_2} + o(1) = \alpha + o(1).$$

The reference choice may affect finite-sample power and selection error; the symmetric test treats both directions equally.

Proof. For each direction $k \to h$, let $p_{k\to h}^{\circ}$ use the fixed reference graph $E^{(k)}$ and the same sample h. On $\{\hat{E}^{(k)} = E^{(k)}\}$ the selected-graph and oracle procedures coincide pathwise, so

$$\mathbb{P}(p_{k\to h} \le \alpha/2) \le \mathbb{P}(p_{k\to h}^{\circ} \le \alpha/2) + \pi_{n_k} \le \alpha/2 + \pi_{n_k} + o(1).$$

The union bound over the two directions proves the claim without requiring independence between the samples or between their directional p-values. □

B.2 Detailed Discussion on Consistency-empowered Estimator

This subsection gives the technical details behind the consistency-empowered correction in Sections 2.3 and 3.3. The new object is the standard error used to decide whether an inside-support coordinate $\hat{\omega}_{0,ij}$ is stochastically close to zero. Because the observations are temporally dependent unlike in Le et al. (forthcoming), this standard error must be defined through a long-run variance. Note that the local inverse precision estimation error can be written as

$$\hat{\omega}_{0,ij} - \omega_{0,ij} = -\frac{1}{n}\sum_{t=1}^{n} \mathcal{X}_{t,ij}^{\omega} + R_{n,ij}^{\omega}.$$

Here, the local-inverse score process is defined as

$$\mathcal{X}_{t,ij}^{\omega} = \boldsymbol{v}_{0,ij}^{\top}(X_t X_t^{\top} - \Sigma)\boldsymbol{\omega}_{0,j}, \qquad i \in S_{0,j},$$

where $\boldsymbol{v}_{0,ij} = B_{0,j}(B_{0,j}^{\top}\Sigma B_{0,j})^{-1}B_{0,j}^{\top}e_i$. Note that the residual $R_{n,ij}^{\omega}$ is of smaller order than the leading linear term on the same local stability event used in Proposition 1, using the same proof technique. Therefore, the stochastic scale of $\hat{\omega}_{0,ij}$ is governed by the long-run variance of $\mathcal{X}_{t,ij}^{\omega}$, which is denoted by

$$L_{ij}^{\omega} = \sum_{h=-\infty}^{\infty} \mathrm{Cov}(\mathcal{X}_{0,ij}^{\omega}, \mathcal{X}_{h,ij}^{\omega}).$$

We refer to $\tau_{ij} = \sqrt{L_{ij}^{\omega}/n}$ as the long-run standard error of $\hat{\omega}_{0,ij}$. The estimation of this quantity is directly useful in the implementation of the method; see Section 2.3. The long-run variance L_{ij}^{ω} is useful in the subsequent proof.

Condition 6. *There exist constants* $0 < c_\omega < C_\omega < \infty$ *such that* $c_\omega \le L_{ij}^{\omega} \le C_\omega$ *for all* $1 \le j \le p$ *and all* $i \in S_{0,j}$*.*

This condition plays the same role as the score nondegeneracy condition for the direct statistic. It prevents the standard error used in the consistency-empowered screening rule from degenerating.

We estimate L_{ij}^{ω} by the same overlapping-block principle used in the bootstrap construction. For $i \in S_{0,j}$, define $\hat{\boldsymbol{v}}_{0,ij} = B_{0,j}(B_{0,j}^{\top}\hat{\Sigma}_n B_{0,j})^{-1}B_{0,j}^{\top}e_i$. The plug-in local-inverse estimator is

$$\hat{\mathcal{X}}_{t,ij}^{\omega} = \hat{\boldsymbol{v}}_{0,ij}^{\top}(X_t X_t^{\top} - \hat{\Sigma}_n)\hat{\boldsymbol{\omega}}_{0,j}. \tag{9}$$

Using the same block length l as in the buildup of Theorem 2, define the block-bootstrap estimator of the long-run variance by

$$\hat{L}_{ij}^{\omega} = \frac{1}{n-l+1}\sum_{r=l}^{n}\left(l^{-1/2}\sum_{t=r-l+1}^{r}\hat{\mathcal{X}}_{t,ij}^{\omega}\right)^2, \tag{10}$$

and $\hat{\tau}_{ij} = \sqrt{\hat{L}_{ij}^{\omega}/n}$. This estimator is scalar, yet it is constructed from the same block-sampling idea similar to that of $\hat{F}_{n,l}$ in (5). The following lemma verifies that this estimator is uniformly consistent for our Gaussian linear process setup. The proof is deferred to Appendix E.2.

Lemma 1. *Assume Conditions 1-6. Let $0 < \epsilon < 1$, choose the block length l as in Theorem 2, and assume $\mathfrak{R}_{n,l}(\epsilon) \to 0$. Then*

$$\max_{1\le j\le p}\max_{i\in S_{0,j}}\left|\frac{\hat{\tau}_{ij}}{\tau_{ij}} - 1\right| = o_{\mathbb{P}}(1).$$

The following lemma records the screening behavior of the consistency-empowered correction. It is the time-series analogue of the zero-versus-nonzero separation used in Le et al. (forthcoming). The proof is deferred to Appendix E.3.

Lemma 2. *Assume Conditions 1-6. Assume also the block-length and rate conditions of Lemma 1. Let $\delta_n = \sqrt{\log n}$. Under $H_0 : E = E_0$, assume that the nonzero coordinates $\omega_{0,ij}$ with $i \in S_{0,j}$ are uniformly bounded away from zero. Then*

$$\mathbb{P}_{H_0}\left(\hat{\boldsymbol{\Delta}}_j = 0 \text{ for all } 1 \le j \le p\right) \to 1.$$

Under every fixed included alternative $E \subsetneq E_0$,

$$\mathbb{P}\left(\hat{\boldsymbol{\Delta}}_j \neq 0 \text{ for some } 1 \le j \le p\right) \to 1.$$

The screening conclusions also hold for any sequence satisfying $\delta_n \to \infty$ and $\delta_n + \sqrt{\log p} = o(\sqrt{n})$, with the block length and rate conditions of Lemma 1; see Appendix E.3.

We can now state the formal version of Theorem 4. The theorem uses the fixed consistency-empowerment levels $\delta_n = \sqrt{\log n}$ and $C_n = C_\Delta\sqrt{\log p}$ with any fixed $C_\Delta > 0$. It ensures that under certain conditions, the consistency-empowered statistic is equal to the direct test statistic in probability, yet unlike the latter, it also ensures that the test power converge to one under fixed alternatives. We have redirected the proof of this theorem to Appendix E.4.

Theorem 6. *Assume Conditions 1-6 under H_0. Let $0 < \epsilon < 1$, choose the block length l as in Theorem 2, and assume $\mathfrak{R}_{n,l}(\epsilon) \to 0$. Set $\delta_n = \sqrt{\log n}$ and $C_n = C_\Delta\sqrt{\log p}$ for any fixed constant $C_\Delta > 0$.*

1. *Under the null hypothesis $H_0 : E = E_0$, assume additionally that the nonzero coordinates $\omega_{0,ij}$ with $i \in S_{0,j}$ are uniformly bounded away from zero by a fixed constant. Then*

   $$\mathbb{P}_{H_0}(\tilde{T}_n = \hat{T}_n) \to 1.$$

   Consequently, Theorems 1 and 2 remain valid with $\tilde{T}_n$ in place of $\hat{T}_n$, and we have $\mathbb{P}_{H_0}(\tilde{T}_n > \hat{q}_{1-\alpha}) \to \alpha$.

2. *Furthermore, under any fixed alternative $H_1 : E \neq E_0$ for which Conditions 1-3 and Condition 6 hold with the alternative population covariance and precision matrices, and for which $\mathfrak{R}_{n,l}(\epsilon) \to 0$, we have the asymptotic guarantee on test power,*

   $$\mathbb{P}_{H_1}(\tilde{T}_n > \hat{q}_{1-\alpha}) \to 1.$$

Remark 1. *The signal condition is only needed to prevent true nonzero precision coordinates from being mistakenly pumped under the null. If, alternatively, the nonzero entries are allowed to shrink with n, then an explicit beta-min condition will be necessary. Here "fixed alternative" means that the population graph and covariance structure do not approach the null with n. The choice $\delta_n = \sqrt{\log n}$ is a convenient sufficient choice: the same conclusions hold when $\delta_n \to \infty$ and $\delta_n + \sqrt{\log p} = o(\sqrt{n})$, with $C_n = C_\Delta\sqrt{\log p}$ unchanged.*

B.3 Support Recovery by Thresholded CLIME

We give a thresholded CLIME selector whose error bound below verifies the consistency requirement in Corollary 5 whenever it vanishes. For each $1 \le j \le p$, define

$$\check{\boldsymbol{\omega}}_j^{(1)} \in \operatorname*{arg\,min}_{\boldsymbol{\omega}\in\mathbb{R}^p} |\boldsymbol{\omega}|_1 \quad \text{subject to} \quad |\hat{\Sigma}^{(1)}\boldsymbol{\omega} - e_j|_\infty \le \lambda_{n_1},$$

and let

$$\check{\Omega}^{(1)} = \big(\check{\boldsymbol{\omega}}_1^{(1)}, \dots, \check{\boldsymbol{\omega}}_p^{(1)}\big).$$

If a feasible set is empty, we set the corresponding column equal to zero. The proof below shows that all feasible sets are nonempty on the event of interest.

Following Cai et al. (2011), symmetrize $\check{\Omega}^{(1)}$ by retaining, for each $i < j$, whichever of $\check{\omega}_{ij}^{(1)}$ and $\check{\omega}_{ji}^{(1)}$ has smaller absolute value (breaking ties arbitrarily), and use that value for both symmetric entries. Keep the diagonal unchanged and denote the resulting matrix by $\hat{\Omega}^{\mathrm{CL},(1)}$. This is the CLIME estimator of $\Omega^{(1)}$. For a threshold $\tau_{n_1,p} > 0$, select

$$\hat{E}^{(1)} = \Big\{(i,j) : 1 \le i,j \le p, \quad i = j \text{ or } |\hat{\omega}_{ij}^{\mathrm{CL},(1)}| > \tau_{n_1,p}\Big\}.$$

Thus the diagonal is included by convention, as in the entrywise-support definition used in the main paper, while only the off-diagonal entries are thresholded.

For the theoretical guarantee, choose

$$\lambda_{n_1} = C_\Omega K_A \sqrt{\frac{\log(pn_1)}{n_1}}, \qquad \tau_{n_1,p} = 4C_\Omega\lambda_{n_1}, \qquad K_A^2 \ge 2v_\Sigma(C_0,\beta)(A+2),$$

where $A > 0$ is a fixed constant, and recall that C_Ω is the constant in Condition 2. Under Condition 1, the proof of Lemma 18 in Appendix A of Zhai et al. (2026) also gives the absolute-covariance bound, uniformly over $1 \le s,t \le p$,

$$\sum_{k=-\infty}^{\infty} |(\Gamma_k)_{ss}(\Gamma_k)_{tt} + (\Gamma_k)_{st}(\Gamma_k)_{ts}| \le v_\Sigma(C_0,\beta),$$

where $\Gamma_k = \mathrm{Cov}(X_0^{(1)}, X_{-k}^{(1)})$ and $v_\Sigma(C_0,\beta)$ is a finite constant depending only on C_0 and the fixed $\beta > 3/4$. The finite-sample coordinate variances in the covariance Gaussian approximation are Bartlett-weighted sums of these covariances and are therefore also bounded by $v_\Sigma(C_0,\beta)$. Hence both λ_{n_1} and $\tau_{n_1,p}$ have order $\{\log(pn_1)/n_1\}^{1/2}$. The quantity A is only a tail-probability exponent. Taking $A = 1$ is a simple default; increasing A reduces the tail bound but raises the threshold and the required signal strength.

Proposition 3 (Support recovery by thresholded CLIME)**.** *Suppose that the first sample satisfies Conditions 1, 2, and 4. For any fixed $A > 0$, choose K_A as above. If every nonzero off-diagonal entry satisfies*

$$\min_{1\le i\ne j\le p:\omega_{ij}^{(1)}\ne 0} |\omega_{ij}^{(1)}| > 2\tau_{n_1,p},$$

then

$$\pi_{n_1} = \mathbb{P}\left\{\hat{E}^{(1)} \ne E^{(1)}\right\} \le C\Psi(p,n_1) + 2(pn_1)^{-A},$$

where $C > 0$ depends only on C_0, c_0, and the fixed $\beta > 3/4$ in Conditions 1 and 4.

Proof. Define the event

$$\mathcal{E}_{n_1} = \left\{|\hat{\Sigma}^{(1)} - \Sigma^{(1)}|_\infty \le \frac{\lambda_{n_1}}{C_\Omega}\right\}.$$

Let $\boldsymbol{\omega}_j^{(1)}$ be the jth column of $\Omega^{(1)}$. On $\mathcal{E}_{n_1}$,

$$|\hat{\Sigma}^{(1)}\boldsymbol{\omega}_j^{(1)} - e_j|_\infty \le \frac{\lambda_{n_1}}{C_\Omega}|\boldsymbol{\omega}_j^{(1)}|_1 \le \lambda_{n_1}.$$

Thus the true columns are feasible, and optimality gives $|\check{\Omega}^{(1)}|_1 \le |\Omega^{(1)}|_1 \le C_\Omega$. Both matrices satisfy the CLIME constraint, so $|\hat{\Sigma}^{(1)}(\check{\Omega}^{(1)} - \Omega^{(1)})|_\infty \le 2\lambda_{n_1}$. Consequently,

$$\begin{aligned}|\Sigma^{(1)}(\check{\Omega}^{(1)} - \Omega^{(1)})|_\infty &\le 2\lambda_{n_1} + \frac{\lambda_{n_1}}{C_\Omega}\left(|\check{\Omega}^{(1)}|_1 + |\Omega^{(1)}|_1\right) \\ &\le 4\lambda_{n_1}.\end{aligned}$$

Since $\Omega^{(1)}$ is symmetric, its maximum absolute row sum is $|\Omega^{(1)}|_1 \le C_\Omega$. Using the identity $\Omega^{(1)}\Sigma^{(1)} = I$ gives

$$|\check{\Omega}^{(1)} - \Omega^{(1)}|_\infty \le 4C_\Omega\lambda_{n_1} = \tau_{n_1,p}.$$

Every entry of $\hat{\Omega}^{\mathrm{CL},(1)}$ is one of $\check{\omega}_{ij}^{(1)}$ and $\check{\omega}_{ji}^{(1)}$. Since $\omega_{ij}^{(1)} = \omega_{ji}^{(1)}$, the symmetrization preserves this bound:

$$|\hat{\Omega}^{\mathrm{CL},(1)} - \Omega^{(1)}|_\infty \le \tau_{n_1,p}.$$

Hence every zero off-diagonal entry is removed by thresholding, while every nonzero entry is retained because $|\hat{\omega}_{ij}^{\mathrm{CL},(1)}| \ge |\omega_{ij}^{(1)}| - \tau_{n_1,p} > \tau_{n_1,p}$. Because $\Omega^{(1)}$ is positive definite, its diagonal belongs to $E^{(1)}$; these pairs are included in $\hat{E}^{(1)}$ by definition. Thus $\hat{E}^{(1)} = E^{(1)}$ on $\mathcal{E}_{n_1}$.

It remains to bound $\mathbb{P}(\mathcal{E}_{n_1}^c)$. Let $Z_\Sigma^{(1)}$ be the Gaussian vector in the covariance approximation for the first sample. Under Conditions 1 and 4, Theorem 1 of Zhai et al. (2026) bounds the approximation error by $C\Psi(p, n_1)$, and under Condition 1, each coordinate variance is bounded by $v_\Sigma(C_0, \beta)$. The Gaussian union bound, the choice of K_A, and $p^2 \le (pn_1)^2$ therefore give

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{n_1}^c) &\le C\Psi(p, n_1) + 2p^2 \exp\left\{-\frac{K_A^2 \log(pn_1)}{2v_\Sigma(C_0, \beta)}\right\} \\ &\le C\Psi(p, n_1) + 2(pn_1)^{-A}.\end{aligned}$$

Since $\hat{E}^{(1)} = E^{(1)}$ on $\mathcal{E}_{n_1}$, $\pi_{n_1} \le \mathbb{P}(\mathcal{E}_{n_1}^c)$, which proves the result. □

C Additional Experiment Details

C.1 Generating Copies of Time Series using Fast Fourier Transformation

The simulation studies require many repeated realizations of the Gaussian linear process $X_i = \sum_{t=0}^{\infty} A_t \epsilon_{i-t}$. A direct implementation is computationally expensive because each observation involves an infinite summation. Following the FFT-based simulation scheme summarized in Zhai et al. (2026), we first replace the process by the truncated version

$$X_i^* = \sum_{t=0}^{N-1} A_t \epsilon_{i-t}.$$

The truncation length N needs to be large when the process has long memory, since the coefficient matrices A_t decay slowly. In all simulations, we take $N = n^2$ and write $K_{N,n} = \lfloor N/n \rfloor$. The key computational idea is to generate one long circular convolution and then divide it into $K_{N,n}$ consecutive length-n segments, which gives many approximate copies of $\boldsymbol{X}_n^* = [X_1^*, X_2^*, \ldots, X_n^*]^\top$.

We first describe the univariate case. Suppose $p = d = 1$ and $X_i^* = \sum_{t=0}^{N-1} a_t \epsilon_{i-t}$. Let

$$\boldsymbol{a}_N = [a_1, a_2, \ldots, a_{N-1}, a_0]^\top, \qquad \boldsymbol{e}_N = [\epsilon_n, \epsilon_{n-1}, \ldots, \epsilon_{n-N+1}]^\top.$$

The circulant embedding approximation is

$$\boldsymbol{W}_N = [W_1, W_2, \ldots, W_N]^\top = \mathrm{Circ}(\boldsymbol{a}_N)\boldsymbol{e}_N,$$

where $\mathrm{Circ}(\boldsymbol{a}_N)$ denotes the circulant matrix generated from $\boldsymbol{a}_N$. Since circulant matrices are diagonalized by the discrete Fourier transform, this multiplication can be computed by

$$\boldsymbol{b}_N = \mathrm{FFT}(\boldsymbol{a}_N), \qquad \boldsymbol{f}_N = \mathrm{FFT}(\boldsymbol{e}_N), \qquad \boldsymbol{W}_N = \mathrm{FFT}^{-1}(\boldsymbol{p}_N),$$

where $\boldsymbol{p}_N = \boldsymbol{b}_N \circ \boldsymbol{f}_N$ and $\circ$ denotes entrywise multiplication. The vector $\boldsymbol{W}_N$ is then partitioned into consecutive blocks

$$\boldsymbol{W}^{(k)} = [W_{(k-1)n+1}, W_{(k-1)n+2}, \ldots, W_{kn}]^\top, \qquad 1 \le k \le K_{N,n}.$$

Each $\boldsymbol{W}^{(k)}$ is treated as one simulated realization of $\boldsymbol{X}_n^*$. These blocks are not exactly independent because they are extracted from the same long circular convolution, but the resulting dependence is weak enough for the Monte Carlo approximation used in the simulation study.

The multidimensional case is handled coordinate-by-coordinate while using the same innovation sequence, which preserves the cross-sectional dependence induced by the coefficient matrices. For $X_i^* \in \mathbb{R}^p$, $A_t \in \mathbb{R}^{p\times d}$, and $\epsilon_t \in \mathbb{R}^d$, first generate

$$\boldsymbol{e}_N = [\epsilon_n^\top, \epsilon_{n-1}^\top, \ldots, \epsilon_{n-N+1}^\top]^\top \in \mathbb{R}^{Nd}, \qquad \boldsymbol{f}_N = \mathrm{FFT}(\boldsymbol{e}_N).$$

For the j-th coordinate, define

$$\boldsymbol{a}_{N,j} = [(A_1)_{j\cdot}, (A_2)_{j\cdot}, \ldots, (A_{N-1})_{j\cdot}, (A_0)_{j\cdot}]^\top \in \mathbb{R}^{Nd}, \qquad \boldsymbol{A}^j = \mathrm{Circ}(\boldsymbol{a}_{N,j}).$$

We compute

$$\boldsymbol{b}_{N,j} = \mathrm{FFT}(\boldsymbol{a}_{N,j}), \qquad \boldsymbol{p}_{N,j} = \boldsymbol{b}_{N,j} \circ \boldsymbol{f}_N, \qquad \boldsymbol{Y}^j = \mathrm{FFT}^{-1}(\boldsymbol{p}_{N,j}) \in \mathbb{R}^{Nd}.$$

The simulated value of the j-th coordinate at time i is then read from the first coordinate of the corresponding d-dimensional block, namely $W_{ij} = (\boldsymbol{Y}^j)_{(i-1)d+1}$. Repeating this operation for $j = 1, \ldots, p$ fills the full simulated array. Finally, for $1 \le k \le K_{N,n}$, we form

$$\boldsymbol{W}^{(k)} = (W_{ij})_{(k-1)n+1\le i\le kn,\ 1\le j\le p} \in \mathbb{R}^{n\times p}.$$

We treat $\boldsymbol{W}^{(1)}, \boldsymbol{W}^{(2)}, \ldots, \boldsymbol{W}^{(K_{N,n})}$ as $K_{N,n}$ approximate copies of $\boldsymbol{X}_n^*$ and use the first 200 available blocks in the experiments. This FFT construction is the reason that the high-dimensional simulation regimes in Section 4 remain computationally feasible.

C.2 Additional Null Test Results

Figure 3 gives the QQ plots for the out-of-theory case $\beta = 0.55$, while Tables 3 and 4 report the Kolmogorov and Wasserstein-1 distances for all three dependence levels in the original $b = 4$ experiment of Section 4.1.

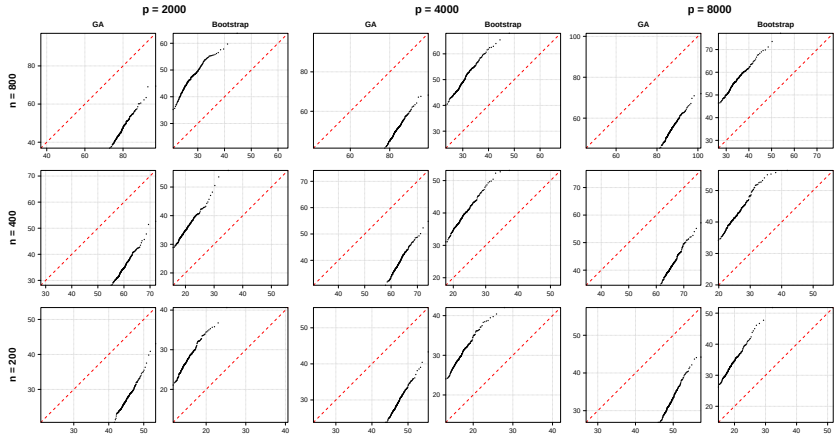

Figure 3: QQ-plots for null tests under ultra-long-memory, $\beta = 0.55$, which exceeds the range in Condition 1. The x-axis is $|G|_\infty$ for GA and $l^{-1/2}|\hat{B}_{r,l}|_\infty$ for bootstrap. The y-axis is $\hat{T}_n$ for both. The red dashed line is $y = x$. Closer to this line indicates better proximity.

Table 3: Kolmogorov distance for the statistic $\hat{T}_n$ under H_0. The GA section compares $\hat{T}_n$ against its finite-sample Gaussian approximation $|G|_\infty$, while the Bootstrap section compares $\hat{T}_n$ against its overlapping block bootstrap approximation $l^{-1/2}|\hat{B}_{r,l}|_\infty$. Range of the Kolmogorov distance is $[0, 1]$. A smaller value indicates a closer approximation.

		$\beta = 2$			$\beta = 0.9$			$\beta = 0.55$		
	n	$p = 2000$	$p = 4000$	$p = 8000$	$p = 2000$	$p = 4000$	$p = 8000$	$p = 2000$	$p = 4000$	$p = 8000$
GA	800	0.20	0.24	0.26	0.36	0.49	0.58	1.00	1.00	1.00
	400	0.36	0.36	0.43	0.15	0.18	0.25	1.00	1.00	1.00
	200	0.52	0.56	0.67	0.43	0.32	0.20	1.00	1.00	1.00
Bootstrap	800	0.06	0.14	0.22	0.10	0.10	0.14	0.96	0.97	0.98
	400	0.10	0.08	0.16	0.57	0.52	0.47	0.98	0.97	0.98
	200	0.35	0.26	0.30	0.84	0.83	0.85	0.98	0.98	0.98

Table 4: Wasserstein-1 distance for the statistic $\hat{T}_n$ under H_0. The GA section compares $\hat{T}_n$ against its finite-sample Gaussian approximation $|G|_\infty$, while the Bootstrap section compares $\hat{T}_n$ against its overlapping block bootstrap approximation $l^{-1/2}|\hat{B}_{r,l}|_\infty$. Range of the W1-distance is $[0, \infty)$. A smaller value indicates a closer approximation.

		$\beta = 2$			$\beta = 0.9$			$\beta = 0.55$		
	n	$p = 2000$	$p = 4000$	$p = 8000$	$p = 2000$	$p = 4000$	$p = 8000$	$p = 2000$	$p = 4000$	$p = 8000$
GA	800	0.11	0.14	0.13	0.76	0.92	1.10	32.68	32.96	32.45
	400	0.20	0.19	0.23	0.30	0.26	0.38	24.96	23.97	23.79
	200	0.36	0.41	0.44	0.62	0.34	0.29	17.76	17.84	16.78
Bootstrap	800	0.04	0.07	0.13	0.15	0.24	0.30	18.90	19.28	21.42
	400	0.05	0.04	0.09	1.16	1.19	1.04	15.08	15.86	16.54
	200	0.22	0.14	0.16	2.09	2.10	2.26	12.02	13.06	14.25

We extend the same null-validity setup and simulation scheme in Appendix C.1 in two directions. Each configuration uses 200 simulated time series, and the bootstrap rejection rates are reported at the nominal level $\alpha = 0.10$. The distances below compare $\hat{T}_n$ with the same Gaussian and block-bootstrap approximations as in Tables 3 and 4.

In Extension 1, we fix $n = 800$ and $p = 2000$ and vary the bandwidth $b \in \{2, 4, 6, 8, 12, 16\}$ under both $\beta = 2$ and 0.9. For this banded graph, $s_0 = 2b + 1$, giving $s_0 \in \{5, 9, 13, 17, 25, 33\}$. The precision-matrix construction is unchanged, and the $b = 4$ results are reused from the original experiment. Table 5 reports the results. Although the distances are not exactly monotone at small bandwidths, approximation accuracy and null calibration generally deteriorate once b, and hence s_0, becomes large, especially under long memory. This is consistent with the s_0-dependent remainder terms in Theorems 1 and 2.

Table 5: Extension 1: varying sparsity under short memory ($\beta = 2$, left) and long memory ($\beta = 0.9$, right), with $n = 800$ and $p = 2000$. Each configuration uses 200 repetitions. Rejection rates use the nominal level $\alpha = 0.10$.

		$\beta = 2$					$\beta = 0.9$				
		Gaussian		Bootstrap		Rejection rate	Gaussian		Bootstrap		Rejection rate
b	s_0	d_K	W_1	d_K	W_1		d_K	W_1	d_K	W_1	
2	5	0.11	0.05	0.18	0.11	0.05	0.67	1.58	0.06	0.11	0.11
4	9	0.20	0.11	0.06	0.04	0.10	0.36	0.76	0.10	0.15	0.08
6	13	0.22	0.14	0.04	0.02	0.11	0.22	0.31	0.17	0.31	0.12
8	17	0.24	0.13	0.08	0.04	0.11	0.48	0.66	0.35	0.62	0.21
12	25	0.28	0.17	0.15	0.08	0.15	0.83	1.61	0.59	1.05	0.41
16	33	0.40	0.23	0.31	0.16	0.20	0.94	2.32	0.73	1.38	0.66

In Extension 2, we retain $p = 2000$ and $b = 4$ and add $\beta = 0.85$ and 0.8 to examine behavior closer to the boundary $\beta = 3/4$. The left panel of Table 6 gives the results at $n = 800$, including the

original $\beta = 2$ and 0.9 results for comparison. At this fixed sample size and dimension, the bootstrap distances become larger and the null rejection rates become increasingly inflated in the near-boundary settings. We then double n to 1600, keeping $p = 2000$ and $b = 4$ fixed; see the right panel of Table 6. The bootstrap approximation and null calibration generally improve. These results illustrate the finite-sample limitation near $\beta = 3/4$ and are qualitatively consistent with the joint (n, p, β) scaling in our bounds.

Table 6: Extension 2: varying temporal dependence with $n = 800$ (left) and $n = 1600$ (right), with $p = 2000$ and $b = 4$. Each configuration uses 200 repetitions. Rejection rates use the nominal level $\alpha = 0.10$.

	$n = 800$					$n = 1600$				
	Gaussian		Bootstrap		Rejection rate	Gaussian		Bootstrap		Rejection rate
β	d_K	W_1	d_K	W_1		d_K	W_1	d_K	W_1	
2	0.20	0.11	0.06	0.04	0.10	0.14	0.07	0.07	0.05	0.07
0.9	0.36	0.76	0.10	0.15	0.08	0.56	1.19	0.36	1.10	0.02
0.85	0.27	0.56	0.37	1.14	0.30	0.48	1.20	0.07	0.26	0.05
0.8	0.29	0.63	0.64	2.12	0.56	0.27	0.70	0.36	1.21	0.25

C.3 Sherlock fMRI Experiment Result

Table 7: Movie self-selected graph and recall containment test results at $\alpha = 0.05$. Here $\hat{\nu}$ is the movie-phase threshold used to construct the selected graph, d_i is the degree of node i in the selected movie graph, $q_{0.25}(d)$ and $q_{0.75}(d)$ are the empirical 25% and 75% quartiles of the node degrees, and $\bar{s}$ is the average column-wise support size, with $s_i = d_i + 1$. The column p_{mov} is the movie self-diagnostic p-value used for graph selection. The column p_{rec} is the raw recall-phase containment-test p-value. The column $p_{\mathrm{rec}}^{\mathrm{Holm}}$ is the recall-phase p-value after Holm correction across the 17 subjects.

Subject	$\hat{\nu}$	$q_{0.25}(d)$	$\bar{s}$	$q_{0.75}(d)$	p_{mov}	p_{rec}	$p_{\mathrm{rec}}^{\mathrm{Holm}}$
1	0.1956	22.5	35.9	42.5	0.0692	0.00217	0.0145
2	0.1567	26.0	37.8	44.5	0.0648	0.00121	0.0145
3	0.1089	53.0	63.5	71.0	0.0516	0.00157	0.0145
4	0.1219	36.0	45.5	52.0	0.0835	0.00262	0.0145
5	0.1083	47.0	56.3	62.0	0.0549	0.00196	0.0145
6	0.1511	23.0	32.8	39.0	0.0643	0.00126	0.0145
7	0.1340	47.5	59.5	68.0	0.0511	0.00166	0.0145
8	0.1242	39.0	48.1	53.0	0.0582	0.00110	0.0143
9	0.1590	19.0	29.0	34.0	0.0725	0.00162	0.0145
10	0.1129	45.0	55.0	61.0	0.1153	0.00132	0.0145
11	0.1448	33.0	44.9	53.8	0.0637	0.00217	0.0145
12	0.1302	39.0	50.9	59.0	0.0571	0.000899	0.0129
13	0.1307	40.0	52.2	58.0	0.2202	0.000596	0.0101
14	0.1257	33.0	41.9	46.0	0.0670	0.000861	0.0129
15	0.1456	24.0	33.7	39.0	0.0544	0.00135	0.0145
16	0.1729	31.0	43.8	52.0	0.0511	0.00128	0.0145
17	0.1322	33.0	44.6	50.0	0.0577	0.000760	0.0122

C.4 Long Memory of Sherlock fMRI Time Series

The time series in the Sherlock fMRI dataset exhibit significant long memory. To demonstrate this, we plot the histogram of Hurst exponent of all 17×268 movie-phase time series in Figure 4. The Hurst exponent H, ranging from zero to one, measures the scaling of temporal dependence in a time series, with $H = 1/2$ corresponding to the short-memory benchmark associated with uncorrelated or rapidly decorrelating fluctuations. Values $H > 1/2$ indicate persistence: positive deviations tend to be followed by positive deviations, and negative deviations tend to be followed by negative deviations

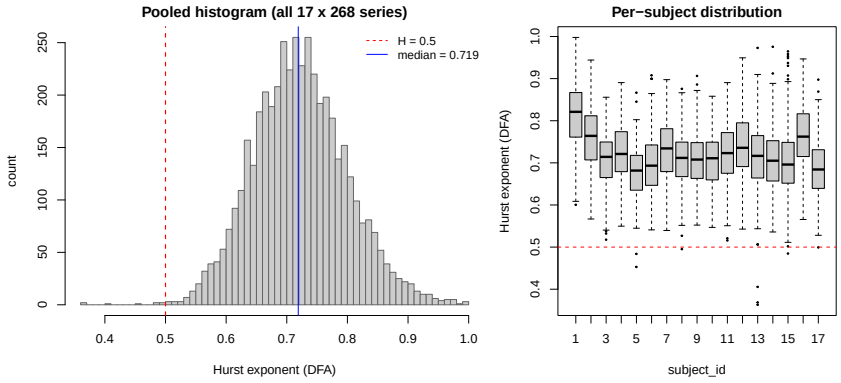

Figure 4: Distribution of Hurst exponents across all subject-ROI movie time series. Vertical dashed line: $H = 0.5$. Values $H > 0.5$ are considered consistent with long-range dependence.

over long time scales. For fractional long-memory processes, the autocorrelation typically satisfies

$$\rho(k) \asymp k^{2H-2},$$

so when $H > 1/2$, we have $2H - 2 \in (-1, 0)$. Consequently, the autocorrelation decays slowly and is not summable, which means that dependence persists across distant lags, i.e., long-memory. We can see from Figure 4 that 99.8% of all time series have $H > 1/2$, i.e., exhibiting long-memory. The median Hurst exponent is at 0.719. For each subject, we have also provided a boxplot. All subjects have a confident interval of Hurst exponent that completely exceeds 0.5, a strong sign of these time series being long-memory.

The ACF plots in Figure 5 provide a complementary diagnostic. In a short-memory process, the autocorrelation decays rapidly toward zero and usually becomes negligible after a small number of lags. By contrast, a slowly decaying or effectively non-decaying ACF over many lags indicates that distant observations remain correlated. This persistent autocorrelation is the empirical signature of long-range dependence. Thus, the combination of $H > 1/2$ and significant autocorrelations across long lags provides evidence that the Sherlock movie time series exhibit long-memory behavior. We selected the time series at the 50th, 75th, 90th, and 95th percentile of Hurst exponents. It appears that all these four time series have significant ACF lags near 100. For the latter two, particularly, the ACF is decaying very slowly. These time series are typically long-range dependent.

D Proof for Section 3

Consider the following examples satisfying Conditions 2–3. Let $\Omega = I_p + W$, where W is symmetric with zero diagonal and $\max_i \sum_j |W_{ij}| \le \rho$ for a fixed $0 < \rho < 1$. Then Ω is positive definite, and the Neumann series gives

$$|\Omega|_1 \le 1 + \rho, \qquad |\Sigma|_1 = |(I_p + W)^{-1}|_1 \le \frac{1}{1-\rho}.$$

For any nonempty proper subset $S \subset \{1, \dots, p\}$, the Schur complement identity gives

$$(\Sigma_{SS})^{-1} = \Omega_{SS} - \Omega_{SS^c}(\Omega_{S^cS^c})^{-1}\Omega_{S^cS}.$$

The two cross-block matrix norms are at most ρ, while $|(\Omega_{S^cS^c})^{-1}|_1 \le (1-\rho)^{-1}$. Consequently,

$$|(\Sigma_{SS})^{-1}|_1 \le 1 + \rho + \frac{\rho^2}{1-\rho} = \frac{1}{1-\rho}.$$

The same bound holds for $S = \{1, \dots, p\}$. Thus Conditions 2–3 hold with $C_\Sigma = C_A = (1-\rho)^{-1}$ and $C_\Omega = 1 + \rho$, uniformly in the dimension and the support sizes.

For a banded example, take $W_{ij} = \rho/(2b)$ when $0 < |i - j| \le b$ and zero otherwise, where $b \ge 1$ is the bandwidth. For a block-diagonal example, partition the coordinates into blocks and, within each

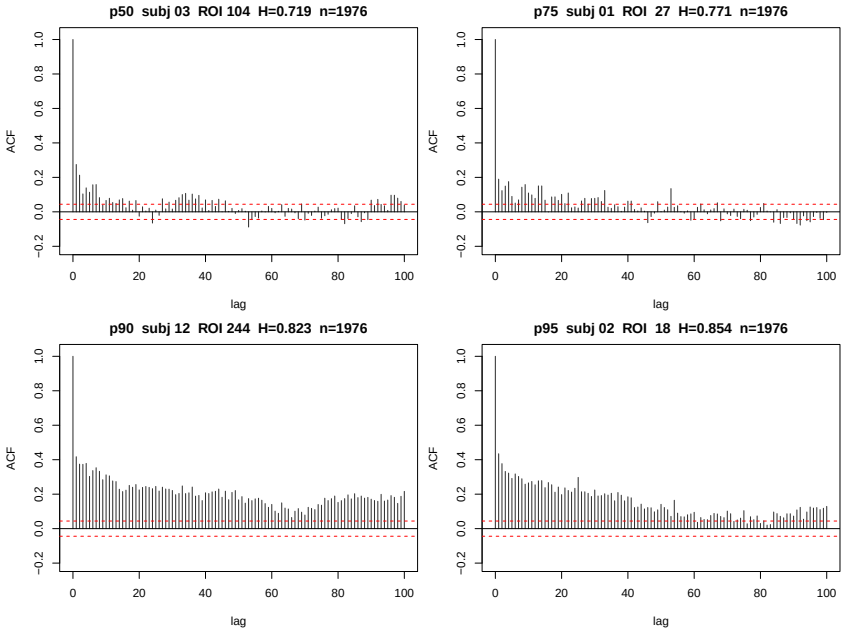

Figure 5: Sample autocorrelation functions (ACF) for movie time series at Hurst exponent distribution percentiles (50th, 75th, 90th, 95th). Each panel shows ACF up to lag 100; dashed bands at $\pm 1.96/\sqrt{n}$. Slowly-decaying significant ACF coefficients are evidence consistent with long-range dependence.

block of size $m \geq 2$, take $W_{ij} = \rho/(m-1)$ for $i \neq j$; all other entries are zero. Singleton blocks have precision entry one. Both constructions satisfy the row-sum bound above. The bandwidth or block size may grow with n, subject to the sparsity–dimension rates in Theorems 1 and 2, while the constants in Conditions 2–3 remain unchanged.

D.1 Proof of Proposition 1

For brevity, write

$$\Delta = \hat{\Sigma}_n - \Sigma, \quad A_j = B_{0,j}^\top \Sigma B_{0,j}, \quad \hat{A}_j = B_{0,j}^\top \hat{\Sigma}_n B_{0,j}, \quad b_j = B_{0,j}^\top e_j.$$

Throughout the proof, the constant $C > 0$ may vary from line to line and depends only on C_A, C_Σ, and C_Ω. On the event $2C_A s_0 |\Delta|_\infty \leq 1$, Condition 3 gives

$$|A_j^{-1}(\hat{A}_j - A_j)|_1 \leq |A_j^{-1}|_1 |\hat{A}_j - A_j|_1 \leq C_A s_{0,j} |\Delta|_\infty \leq \frac{1}{2}.$$

Hence $\hat{A}_j$ is invertible for every $1 \leq j \leq p$, and the exceptional definition of $\hat{\boldsymbol{\omega}}_{0,j}$ being zero when $\hat{A}_j$ being degenerate is then inactive on this event. Moreover,

$$|\hat{A}_j^{-1}|_1 \leq \frac{|A_j^{-1}|_1}{1 - |A_j^{-1}(\hat{A}_j - A_j)|_1} \leq 2C_A.$$

The inverse identity gives

$$\hat{A}_j^{-1} - A_j^{-1} = -A_j^{-1}(\hat{A}_j - A_j)A_j^{-1} + A_j^{-1}(\hat{A}_j - A_j)A_j^{-1}(\hat{A}_j - A_j)\hat{A}_j^{-1}.$$

Since $\boldsymbol{\omega}_{0,j} = B_{0,j} A_j^{-1} b_j$ and $\hat{\boldsymbol{\omega}}_{0,j} = B_{0,j} \hat{A}_j^{-1} b_j$, we obtain

$$\hat{\boldsymbol{\omega}}_{0,j} - \boldsymbol{\omega}_{0,j} = -B_{0,j} A_j^{-1} B_{0,j}^\top \Delta \boldsymbol{\omega}_{0,j} + r_{j,n}^{(\omega)},$$

where

$$r_{j,n}^{(\omega)} = B_{0,j} A_j^{-1}(\hat{A}_j - A_j)A_j^{-1}(\hat{A}_j - A_j)\hat{A}_j^{-1} b_j.$$

Because $|b_j|_1 = 1$, $|\hat{A}_j^{-1}|_1 \le 2C_A$, and $|\hat{A}_j - A_j|_1 \le s_{0,j}|\Delta|_\infty$, we have

$$\max_{1\le j\le p} |r_{j,n}^{(\omega)}|_1 \le 2C_A^3 s_0^2 |\Delta|_\infty^2.$$

Under H_0, $\boldsymbol{\omega}_{0,j}$ is the jth column of Ω. Thus Condition 2 implies $\max_{1\le j\le p} |\boldsymbol{\omega}_{0,j}|_1 \le |\Omega|_1 \le C_\Omega$. Consequently, under the event $2C_A s_0|\Delta|_\infty \le 1$,

$$\begin{aligned}
|\hat{\boldsymbol{\omega}}_{0,j} - \boldsymbol{\omega}_{0,j}|_1 &\le |A_j^{-1}|_1 |B_{0,j}^\top \Delta \boldsymbol{\omega}_{0,j}|_1 + |r_{j,n}^{(\omega)}|_1 \\
&\le C_A s_{0,j} |\Delta|_\infty |\boldsymbol{\omega}_{0,j}|_1 + 2C_A^3 s_0^2 |\Delta|_\infty^2 \\
&\le C_1 s_0 |\Delta|_\infty.
\end{aligned}$$

Now fix $(i,j) \in \mathcal{J}_0$. Since $i \notin S_{0,j}$ and we work under H_0, we have $e_i^\top \Sigma \boldsymbol{\omega}_{0,j} = 0$. Therefore

$$\begin{aligned}
\hat{r}_{ij} &= e_i^\top \hat{\Sigma}_n \hat{\boldsymbol{\omega}}_{0,j} \\
&= e_i^\top (\Sigma + \Delta)(\boldsymbol{\omega}_{0,j} + \hat{\boldsymbol{\omega}}_{0,j} - \boldsymbol{\omega}_{0,j}) \\
&= e_i^\top \Delta \boldsymbol{\omega}_{0,j} + e_i^\top \Sigma (\hat{\boldsymbol{\omega}}_{0,j} - \boldsymbol{\omega}_{0,j}) + e_i^\top \Delta (\hat{\boldsymbol{\omega}}_{0,j} - \boldsymbol{\omega}_{0,j}).
\end{aligned}$$

Substituting the expansion of $\hat{\boldsymbol{\omega}}_{0,j} - \boldsymbol{\omega}_{0,j}$ yields

$$\hat{r}_{ij} = \boldsymbol{u}_{ij}^\top \Delta \boldsymbol{\omega}_{0,j} + \rho_{ij,n},$$

where

$$\rho_{ij,n} = e_i^\top \Sigma r_{j,n}^{(\omega)} + e_i^\top \Delta (\hat{\boldsymbol{\omega}}_{0,j} - \boldsymbol{\omega}_{0,j}).$$

By Condition 2 and the previous bounds,

$$\max_{(i,j)\in\mathcal{J}_0} |\rho_{ij,n}| \le |\Sigma|_1 \max_{1\le j\le p} |r_{j,n}^{(\omega)}|_1 + |\Delta|_\infty \max_{1\le j\le p} |\hat{\boldsymbol{\omega}}_{0,j} - \boldsymbol{\omega}_{0,j}|_1 \le C_1 s_0^2 |\Delta|_\infty^2.$$

Since $\boldsymbol{u}_{ij}^\top \Delta \boldsymbol{\omega}_{0,j} = \frac{1}{n}\sum_{t=1}^n \mathcal{X}_{t,ij}$, we obtain

$$|\hat{T}_n - T_n^{\mathrm{L}}| \le \sqrt{n} \max_{(i,j)\in\mathcal{J}_0} |\rho_{ij,n}| \le C_1 \sqrt{n} s_0^2 |\Delta|_\infty^2.$$

It remains to bound the linear term. Condition 2 gives $|B_{0,j}^\top \Sigma e_i|_1 \le |\Sigma e_i|_1 \le |\Sigma|_1 \le C_\Sigma$. Together with Condition 3, this implies

$$|\boldsymbol{u}_{ij}|_1 \le 1 + |A_j^{-1}|_1 |B_{0,j}^\top \Sigma e_i|_1 \le 1 + C_A C_\Sigma.$$

Since $|\boldsymbol{\omega}_{0,j}|_1 \le C_\Omega$, we have

$$T_n^{\mathrm{L}} = \sqrt{n} \max_{(i,j)\in\mathcal{J}_0} |\boldsymbol{u}_{ij}^\top \Delta \boldsymbol{\omega}_{0,j}| \le C_1 \sqrt{n} |\Delta|_\infty.$$

Combining this bound with the bound for $|\hat{T}_n - T_n^{\mathrm{L}}|$ gives

$$\hat{T}_n \le C_1 \sqrt{n} |\Delta|_\infty + C_1 \sqrt{n} s_0^2 |\Delta|_\infty^2.$$

This proves Proposition 1.

D.2 Proof of Theorem 1

The following lemma connects Proposition 1 to Theorems 1 and 2 by representing the oracle scores as selected covariance coordinates of $Y_t = HX_t$ and verifying the assumptions of Theorems 1 and 2 of Zhai et al. (2026).

Lemma 3 (Selected covariance coordinates)**.** *Assume Conditions 1, 2, 3, and 5 under H_0. Define*

$$Y_t = \left(\{\boldsymbol{u}_{ij}^\top X_t\}_{(i,j)\in\mathcal{J}_0}, \{\boldsymbol{\omega}_{0,j}^\top X_t\}_{1\le j\le p} \right)^\top.$$

Then $Y_t = HX_t = \sum_{\ell=0}^\infty H A_\ell \epsilon_{t-\ell}$ is a Gaussian linear process of dimension $q = |\mathcal{J}_0| + p \le p^2 + p$, and every $\mathcal{X}_{t,ij}$ is a selected coordinate of $Y_t Y_t^\top - \mathbb{E}[Y_t Y_t^\top]$. The number of selected coordinates is $|\mathcal{J}_0| \le p^2$. With $C_H = \max\{1 + C_A C_\Sigma, C_\Omega\}$, the loading and coefficient bounds are

$$\max_r |H_{r\cdot}|_1 \le C_H, \qquad \max_r |(HA_\ell)_{r\cdot}| \le C_0 C_H (1 \vee \ell)^{-\beta}.$$

For $\Xi_m = \mathrm{Var}(m^{-1/2}\sum_{t=1}^m \mathcal{X}_t)$, there is a constant $C > 0$ such that, for $m \geq 2$,

$$|\Xi_m - \Xi|_\infty \leq C m^{-(4\beta-3)\wedge 1} \log^3(m). \tag{11}$$

In particular, $\min_{a\in\mathcal{J}_0} \Xi_{aa} \geq c_1$ and $\min_{a\in\mathcal{J}_0}(\Xi_m)_{aa} \geq c_1/2$ for all sufficiently large m. These bounds hold uniformly over the selected coordinates, with constants depending only on the fixed β and $C_0, c_1, C_\Sigma, C_\Omega, C_A$, not on $|\mathcal{J}_0|$ or s_0. They verify the coefficient-decay and nondegeneracy assumptions needed by the Gaussian-approximation and block-bootstrap arguments of Zhai et al. (2026) when the maxima are restricted to these selected covariance coordinates; all logarithmic dimension factors remain $O(\log p)$. No nondegeneracy is required of the unselected covariance coordinates of Y_t.

Proof. The rows of H are the vectors $\boldsymbol{u}_{ij}^\top$ and $\boldsymbol{\omega}_{0,j}^\top$. For each $(i,j) \in \mathcal{J}_0$,

$$\mathcal{X}_{t,ij} = (\boldsymbol{u}_{ij}^\top X_t)(\boldsymbol{\omega}_{0,j}^\top X_t) - \mathbb{E}[(\boldsymbol{u}_{ij}^\top X_t)(\boldsymbol{\omega}_{0,j}^\top X_t)],$$

which proves the selected-coordinate representation. Under H_0, $\boldsymbol{\omega}_{0,j}$ is the jth column of Ω, and therefore Condition 2 gives $\max_j |\boldsymbol{\omega}_{0,j}|_1 \leq C_\Omega$. Moreover, by Conditions 2 and 3,

$$|\boldsymbol{u}_{ij}|_1 \leq 1 + |(B_{0,j}^\top \Sigma B_{0,j})^{-1}|_1 |B_{0,j}^\top \Sigma e_i|_1 \leq 1 + C_A C_\Sigma.$$

Consequently, the coefficient matrices satisfy

$$\max_r |(HA_\ell)_{r\cdot}| \leq \max_r |H_{r\cdot}|_1 \max_s |(A_\ell)_{s\cdot}| \leq C_0 C_H (1\vee\ell)^{-\beta}.$$

For $a, b \in \mathcal{J}_0$, write $\gamma_{ab}(k) = \mathrm{Cov}(\mathcal{X}_{0,a}, \mathcal{X}_{k,b})$. By Isserlis' formula and the loading bounds,

$$|\gamma_{ab}(k)| \leq C\left(\sum_{\ell=0}^\infty (1\vee\ell)^{-\beta}(1\vee(\ell+|k|))^{-\beta}\right)^2.$$

Since $\beta > 3/4$, the right-hand side is summable in k, uniformly in a, b. Using

$$\Xi_m = \sum_{|k|<m}\left(1 - \frac{|k|}{m}\right)\mathrm{Cov}(\mathcal{X}_0, \mathcal{X}_k),$$

the tail sum and the Bartlett-weight error give

$$\begin{aligned}|\Xi_m - \Xi|_\infty &\leq \max_{a,b}\sum_{|k|\geq m}|\gamma_{ab}(k)| + \frac{1}{m}\max_{a,b}\sum_{1\leq|k|<m}|k||\gamma_{ab}(k)| \\ &\leq C m^{-(4\beta-3)\wedge 1}\log^3(m),\end{aligned}$$

where C may depend on the fixed β. Condition 5 now yields

$$\min_{a\in\mathcal{J}_0}(\Xi_m)_{aa} \geq c_1 - |\Xi_m - \Xi|_\infty \geq c_1/2$$

for all sufficiently large m, uniformly in the selected set and in s_0. The dimensions $q \leq p^2 + p$ and $|\mathcal{J}_0| \leq p^2$ imply $\log(qm) = O(\log(pm))$. In the arguments of Theorems 1 and 2 of Zhai et al. (2026), the coefficient bounds are uniform over the participating coordinates and the variance lower bounds are needed for the coordinates entering the maximum. Restricting their coordinate indices to the selected entries therefore preserves these bounds, without imposing nondegeneracy on any unselected entry. □

By Proposition 1, for every $\eta > 0$,

$$\mathbb{P}(|\hat{T}_n - T_n^{\mathrm{L}}| > \eta) \leq \mathbb{P}(2C_A s_0 |\hat{\Sigma}_n - \Sigma|_\infty > 1) + \mathbb{P}(C\sqrt{n} s_0^2 |\hat{\Sigma}_n - \Sigma|_\infty^2 > \eta).$$

We shall now approximate T_n^{L} by a Gaussian maximum. By Lemma 3, Theorem 1 of Zhai et al. (2026) applies to our selected score vector $\mathcal{X}_t$, in that there exists a constant C' related to the stated class constants and the fixed β such that

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}(T_n^{\mathrm{L}} \leq u) - \mathbb{P}(|G_n|_\infty \leq u)\right| \leq C'\Psi(p,n).$$

Here, G_n is a centered Gaussian vector with the covariance matrix

$$\Xi_n = \mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathcal{X}_t\right) = \sum_{|k|<n}\left(1-\frac{|k|}{n}\right)\mathrm{Cov}(\mathcal{X}_0,\mathcal{X}_k).$$

It remains to replace the finite-sample covariance of the Gaussian approximation by its long-run limit. By (11),

$$|\Xi_n - \Xi|_\infty \le Cn^{-(4\beta-3)\wedge 1}\log^3(n) := \Delta_{\Xi,n}.$$

We now apply the Gaussian comparison inequality of Chernozhukov et al. (2015), which leads to

$$\sup_{u\in\mathbb{R}}|\mathbb{P}(|G_n|_\infty \le u) - \mathbb{P}(|G|_\infty \le u)| \le C\Delta_{\Xi,n}^{1/3}\left\{1\vee\log\left(\frac{2|\mathcal{J}_0|}{\Delta_{\Xi,n}}\right)\right\}^{2/3}.$$

Note that the right-hand side is of smaller polynomial order than $\Psi(p,n)$, and can be absorbed into the constant multiplying $\Psi(p,n)$. Finally, Lemma 3 gives $\min_{a\in\mathcal{J}_0}\Xi_{aa} \ge c_1$ and $\min_{a\in\mathcal{J}_0}(\Xi_n)_{aa} \ge c_1/2$ for all sufficiently large n. Thus the finite-sample Gaussian target G_n is nondegenerate, and the anti-concentration and Gaussian comparison constants depend only on the stated class constants and the fixed β. Therefore, we arrive at the fact that there exists a constant C' related to $C_0, c_1, C_\Sigma, C_\Omega, C_A$ such that

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}(T_n^{\mathrm{L}} \le u) - \mathbb{P}(|G|_\infty \le u)\right| \le C'\Psi(p,n). \tag{12}$$

We now transfer the Gaussian approximation from T_n^{L} to $\hat{T}_n$. For every $u\in\mathbb{R}$ and every $\eta>0$,

$$\mathbb{P}(\hat{T}_n \le u) \le \mathbb{P}(T_n^{\mathrm{L}} \le u+\eta) + \mathbb{P}(|\hat{T}_n - T_n^{\mathrm{L}}| > \eta),$$

and similarly,

$$\mathbb{P}(\hat{T}_n \le u) \ge \mathbb{P}(T_n^{\mathrm{L}} \le u-\eta) - \mathbb{P}(|\hat{T}_n - T_n^{\mathrm{L}}| > \eta).$$

Combining these inequalities give

$$\begin{aligned}\sup_{u\in\mathbb{R}}\left|\mathbb{P}(\hat{T}_n \le u) - \mathbb{P}(|G|_\infty \le u)\right| &\le C'\Psi(p,n) + \mathbb{P}(2C_A s_0|\hat{\Sigma}_n - \Sigma|_\infty > 1)\\ &\quad + \mathbb{P}(C\sqrt{n}s_0^2|\hat{\Sigma}_n - \Sigma|_\infty^2 > \eta) + \sup_{u\in\mathbb{R}}\mathbb{P}(u < |G|_\infty \le u+\eta).\end{aligned} \tag{13}$$

By Condition 5, every coordinate of G has variance at least c_1. The upper variance bound follows from Condition 1 applied to the derived process Y_t, with coefficient constant C_0C_H from Lemma 3. Therefore Nazarov's anti-concentration inequality implies that there exists a constant C' related to $C_0, c_1, C_\Sigma, C_\Omega, C_A$ such that

$$\sup_{u\in\mathbb{R}}\mathbb{P}(u < |G|_\infty \le u+\eta) \le C'\eta\sqrt{\log p}.$$

It remains to control the two probability terms in (13). Let Z_Σ be the centered Gaussian vector in the covariance Gaussian approximation theorem of Zhai et al. (2026), so that under Conditions 1 and 4,

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}(\sqrt{n}|\hat{\Sigma}_n - \Sigma|_\infty \ge u) - \mathbb{P}(|Z_\Sigma|_\infty \ge u)\right| \le C_{\mathrm{cov}}\Psi(p,n),$$

where C_{cov} is related to C_0 and c_0. By the upper variance bound implied by Condition 1, there exists $v_\Sigma = v_\Sigma(C_0)$ such that every coordinate of Z_Σ has variance at most v_Σ. Hence, for all $x>0$,

$$\mathbb{P}(|Z_\Sigma|_\infty > x) \le 2p^2\exp\left(-\frac{x^2}{2v_\Sigma}\right).$$

Consequently,

$$\begin{aligned}\mathbb{P}(2C_A s_0|\hat{\Sigma}_n - \Sigma|_\infty > 1) &= \mathbb{P}\left(\sqrt{n}|\hat{\Sigma}_n - \Sigma|_\infty > \frac{\sqrt{n}}{2C_A s_0}\right)\\ &\le 2p^2\exp\left(-\frac{n}{8v_\Sigma C_A^2 s_0^2}\right) + C_{\mathrm{cov}}\Psi(p,n),\end{aligned}$$

and

$$\mathbb{P}(C\sqrt{n}s_0^2|\hat{\Sigma}_n - \Sigma|_\infty^2 > \eta) = \mathbb{P}\left(\sqrt{n}|\hat{\Sigma}_n - \Sigma|_\infty > \left(\frac{\eta\sqrt{n}}{Cs_0^2}\right)^{1/2}\right)$$
$$\leq 2p^2 \exp\left(-\frac{\eta\sqrt{n}}{2v_\Sigma C s_0^2}\right) + C_{\mathrm{cov}}\Psi(p,n).$$

Substituting the last three bounds into (13), we obtain

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}(\hat{T}_n \leq u) - \mathbb{P}(|G|_\infty \leq u)\right| \leq (C' + 2C_{\mathrm{cov}})\Psi(p,n) + 2p^2\exp\left(-\frac{n}{8v_\Sigma C_A^2 s_0^2}\right)$$
$$+ 2p^2\exp\left(-\frac{\eta\sqrt{n}}{2v_\Sigma C s_0^2}\right) + C'\eta\sqrt{\log p}. \quad (14)$$

We now choose η to balance the last two terms. Write $a = (2v_\Sigma C)^{-1}\sqrt{n}/s_0^2$, $b = C'\sqrt{\log p}$, and $M = 2p^2$. Take $\eta_* = a^{-1}\log(e + Ma/b)$. Then we have

$$M\exp(-a\eta_*) + b\eta_* \leq \frac{b}{a}\left[1 + \log\left(e + \frac{Ma}{b}\right)\right].$$

Since $b/a \leq C_* s_0^2\sqrt{\log p}/\sqrt{n}$ and $Ma/b \leq C_* p^2\sqrt{n}/(s_0^2\sqrt{\log p})$, for a constant C_* depending only on C_0, c_1, C_Σ, C_Ω, and C_A, we have

$$M\exp(-a\eta_*) + b\eta_* \leq C_*\frac{s_0^2\sqrt{\log p}}{\sqrt{n}}\left[1 + \log\left(e + \frac{p^2\sqrt{n}}{s_0^2\sqrt{\log p}}\right)\right].$$

Using $\log p \leq \log(pn)$ and

$$\log\left(e + \frac{p^2\sqrt{n}}{s_0^2\sqrt{\log p}}\right) \leq C\log(pn),$$

the last display is bounded by

$$C_*\frac{s_0^2\log^{3/2}(pn)}{\sqrt{n}}.$$

Substituting $\eta = \eta_*$ into (14) and enlarging constants, we obtain

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}(\hat{T}_n \leq u) - \mathbb{P}(|G|_\infty \leq u)\right| \leq C\Psi(p,n) + Cp^2\exp\left(-\frac{cn}{s_0^2}\right) + C\frac{s_0^2\log^{3/2}(pn)}{\sqrt{n}},$$

where C depends only on C_0, c_0, c_1, C_Σ, C_Ω, and C_A. This finishes the proof of Theorem 1.

D.3 Proof of Proposition 2

Recall the proof in Appendix D.1 for the notation

$$\Delta = \hat{\Sigma}_n - \Sigma, \qquad A_j = B_{0,j}^\top \Sigma B_{0,j}, \qquad \hat{A}_j = B_{0,j}^\top \hat{\Sigma}_n B_{0,j}, \qquad b_j = B_{0,j}^\top e_j.$$

In this proof, the constant $C_1' > 0$ may vary from line to line, but they depend only on C_A and C_Σ. Recall the proof in in Appendix D.1 that, on the event $2C_A s_0|\Delta|_\infty \leq 1$, the matrices $\hat{A}_j$ are invertible and $\max_{1\leq j\leq p}|\hat{A}_j^{-1}|_1 \leq 2C_A$. Moreover, by the inverse identity, we get

$$\hat{A}_j^{-1} - A_j^{-1} = -A_j^{-1}(\hat{A}_j - A_j)\hat{A}_j^{-1}.$$

And since $|\hat{A}_j - A_j|_1 \leq s_{0,j}|\Delta|_\infty$, we have

$$\max_{1\leq j\leq p}|\hat{A}_j^{-1} - A_j^{-1}|_1 \leq C_1' s_0|\Delta|_\infty.$$

Therefore, since $\boldsymbol{\omega}_{0,j} = B_{0,j}A_j^{-1}b_j$, $\hat{\boldsymbol{\omega}}_{0,j} = B_{0,j}\hat{A}_j^{-1}b_j$, and $|b_j|_1 = 1$, we have

$$\max_{1\leq j\leq p}|\hat{\boldsymbol{\omega}}_{0,j} - \boldsymbol{\omega}_{0,j}|_1 \leq C_1' s_0|\Delta|_\infty, \qquad \max_{1\leq j\leq p}\{|\boldsymbol{\omega}_{0,j}|_1 + |\hat{\boldsymbol{\omega}}_{0,j}|_1\} \leq C_1'.$$

We now control the perturbation of $\boldsymbol{u}_{ij}$. For $(i,j) \in \mathcal{J}_0$,

$$\hat{\boldsymbol{u}}_{ij} - \boldsymbol{u}_{ij} = -B_{0,j}(\hat{A}_j^{-1} - A_j^{-1})B_{0,j}^\top \Sigma e_i - B_{0,j}\hat{A}_j^{-1}B_{0,j}^\top \Delta e_i.$$

Condition 2 gives

$$|B_{0,j}^\top \Sigma e_i|_1 \le |\Sigma e_i|_1 \le |\Sigma|_1 \le C_\Sigma,$$

meanwhile $|B_{0,j}^\top \Delta e_i|_1 \le s_{0,j}|\Delta|_\infty$. Combining these bounds with the preceding inverse bounds yields

$$\max_{(i,j)\in\mathcal{J}_0} |\hat{\boldsymbol{u}}_{ij} - \boldsymbol{u}_{ij}|_1 \le C_1' s_0 |\Delta|_\infty.$$

This proves the first assertion. The same argument also gives

$$\max_{(i,j)\in\mathcal{J}_0} \{|\boldsymbol{u}_{ij}|_1 + |\hat{\boldsymbol{u}}_{ij}|_1\} \le C_1',$$

because

$$|\boldsymbol{u}_{ij}|_1 \le 1 + |A_j^{-1}|_1 |B_{0,j}^\top \Sigma e_i|_1 \le 1 + C_A C_\Sigma.$$

Next compare the plug-in and oracle scores. Note that $\mathcal{X}_{t,ij} - \bar{\mathcal{X}}_{n,ij} = \boldsymbol{u}_{ij}^\top (X_t X_t^\top - \hat{\Sigma}_n)\boldsymbol{\omega}_{0,j}$. Therefore,

$$\begin{aligned}\hat{\mathcal{X}}_{t,ij} - (\mathcal{X}_{t,ij} - \bar{\mathcal{X}}_{n,ij}) &= \hat{\boldsymbol{u}}_{ij}^\top (X_t X_t^\top - \hat{\Sigma}_n)\hat{\boldsymbol{\omega}}_{0,j} - \boldsymbol{u}_{ij}^\top (X_t X_t^\top - \hat{\Sigma}_n)\boldsymbol{\omega}_{0,j} \\ &= (\hat{\boldsymbol{u}}_{ij} - \boldsymbol{u}_{ij})^\top (X_t X_t^\top - \hat{\Sigma}_n)\hat{\boldsymbol{\omega}}_{0,j} + \boldsymbol{u}_{ij}^\top (X_t X_t^\top - \hat{\Sigma}_n)(\hat{\boldsymbol{\omega}}_{0,j} - \boldsymbol{\omega}_{0,j}).\end{aligned}$$

On the same aforementioned event, we have

$$|X_t X_t^\top - \hat{\Sigma}_n|_\infty \le |X_t|_\infty^2 + |\Sigma|_\infty + |\Delta|_\infty \le C_1'(M_n^2 + 1),$$

where we used $|\Sigma|_\infty \le |\Sigma|_1 \le C_\Sigma$ and $|\Delta|_\infty \le (2C_A s_0)^{-1}$. Combining the preceding bounds gives

$$\max_{1\le t\le n} \max_{(i,j)\in\mathcal{J}_0} |\hat{\mathcal{X}}_{t,ij} - (\mathcal{X}_{t,ij} - \bar{\mathcal{X}}_{n,ij})| \le C_1' s_0 |\Delta|_\infty (M_n^2 + 1).$$

This finishes the proof of Proposition 2. We would like to point out that, although this consequence is not part of the proposition statement, the same pointwise bound implies that for every $l \le r \le n$,

$$l^{-1/2} \left| \sum_{t=r-l+1}^{r} [\hat{\mathcal{X}}_t - (\mathcal{X}_t - \bar{\mathcal{X}}_n)] \right|_\infty \le C_1' \sqrt{l} s_0 |\hat{\Sigma}_n - \Sigma|_\infty (M_n^2 + 1).$$

D.4 Proof of Theorem 2

We first apply the block-bootstrap theorem of Zhai et al. (2026) to the oracle score process. Lemma 3 verifies the selected-coordinate representation, coefficient decay, and nondegeneracy conditions, including the finite-block variance lower bound for $m = n$ and $m = l$ when these lengths are sufficiently large. The selected-coordinate block-bootstrap argument of Zhai et al. therefore applies to $\mathcal{X}_t$, with logarithmic dimension factors unchanged up to constants. Define the oracle overlapping block sums $B_{r,l}^{\mathrm{L}} = \sum_{t=r-l+1}^{r} (\mathcal{X}_t - \bar{\mathcal{X}}_n)$ for $r = l, \ldots, n$, and define the corresponding oracle empirical CDF by

$$\hat{F}_{n,l}^{\mathrm{L}}(u) = \frac{1}{n-l+1} \sum_{r=l}^{n} \mathbf{1}\left\{ l^{-1/2} |B_{r,l}^{\mathrm{L}}|_\infty \le u \right\}.$$

By Theorem 2 of Zhai et al. (2026), there exists an event $\mathcal{A}_n$, at a probability no less than $1 - 3\Psi_B^\epsilon(p,n,\epsilon)$, such that

$$\sup_{u\in\mathbb{R}} |\hat{F}_{n,l}^{\mathrm{L}}(u) - \mathbb{P}(T_n^{\mathrm{L}} \le u)| \le C\Psi_B^{1-\epsilon}(p,n,\epsilon),$$

where

$$\Psi_B(p,n,\epsilon) = \frac{\log^{(5\tilde{\beta}+12)/[(6-2\epsilon)\tilde{\beta}+8]}(pn)}{n^{\tilde{\beta}/[(6-2\epsilon)\tilde{\beta}+8]}}.$$

Next, we shall compare the plug-in block CDF with the oracle one. Let $\Lambda_{n,l} = \sup_{l \le r \le n} l^{-1/2} |\hat{B}_{r,l} - B^{\mathrm{L}}_{r,l}|_\infty$. On the event $\Lambda_{n,l} \le \eta_2$, we have $\hat{F}^{\mathrm{L}}_{n,l}(u - \eta_2) \le \hat{F}_{n,l}(u) \le \hat{F}^{\mathrm{L}}_{n,l}(u + \eta_2)$ for every $u \in \mathbb{R}$. Hence on the event $\mathcal{A}_n \cap \{\Lambda_{n,l} \le \eta_2\}$,

$$\sup_{u \in \mathbb{R}} |\hat{F}_{n,l}(u) - \mathbb{P}(T^{\mathrm{L}}_n \le u)| \le C\Psi_B^{1-\epsilon}(p, n, \epsilon) + \sup_{u \in \mathbb{R}} \mathbb{P}(u - \eta_2 < T^{\mathrm{L}}_n \le u + \eta_2).$$

The increment probability of T^{L}_n is controlled by the linearized Gaussian approximation established in the proof of Theorem 1. Together with Nazarov's inequality, we get

$$\sup_{u \in \mathbb{R}} \mathbb{P}(u - \eta_2 < T^{\mathrm{L}}_n \le u + \eta_2) \le C\Psi(p, n) + C\eta_2 \sqrt{\log p}.$$

On the event $\mathcal{A}_n \cap \{\Lambda_{n,l} \le \eta_2\}$,

$$\sup_{u \in \mathbb{R}} |\hat{F}_{n,l}(u) - \mathbb{P}(T^{\mathrm{L}}_n \le u)| \le C\Psi_B^{1-\epsilon}(p, n, \epsilon) + C\Psi(p, n) + C\eta_2 \sqrt{\log p}.$$

We now transfer from T^{L}_n to $\hat{T}_n$. The same smoothing argument used in the proof of Theorem 1 gives, for every $\eta_1 > 0$,

$$\sup_{u \in \mathbb{R}} |\mathbb{P}(T^{\mathrm{L}}_n \le u) - \mathbb{P}(\hat{T}_n \le u)| \le \mathbb{P}(|\hat{T}_n - T^{\mathrm{L}}_n| > \eta_1) + C\Psi(p, n) + C\eta_1 \sqrt{\log p}.$$

Combining the preceding displays, on the event $\mathcal{A}_n \cap \{\Lambda_{n,l} \le \eta_2\}$, which is of probability no less than $1 - C\Psi_B^{\epsilon}(p, n, \epsilon) - \mathbb{P}(\Lambda_{n,l} > \eta_2)$, we obtain

$$\begin{aligned} &\sup_{u \in \mathbb{R}} |\hat{F}_{n,l}(u) - \mathbb{P}(\hat{T}_n \le u)| \\ &\qquad \le C\Psi_B^{1-\epsilon}(p, n, \epsilon) + C\Psi(p, n) + \mathbb{P}(|\hat{T}_n - T^{\mathrm{L}}_n| > \eta_1) + C(\eta_1 + \eta_2)\sqrt{\log p}. \end{aligned} \tag{15}$$

Take $\eta_1 = K_1 s_0^2 \log(pn)/\sqrt{n}$ and $\eta_2 = K_2 s_0 \sqrt{l/n} \log^{3/2}(pn)$, where K_1 and K_2 are large fixed constants. By Proposition 1,

$$\mathbb{P}(|\hat{T}_n - T^{\mathrm{L}}_n| > \eta_1) \le \mathbb{P}(2C_A s_0 |\hat{\Sigma}_n - \Sigma|_\infty > 1) + \mathbb{P}(C\sqrt{n} s_0^2 |\hat{\Sigma}_n - \Sigma|_\infty^2 > \eta_1).$$

The covariance Gaussian approximation argument used in the proof of Theorem 1 gives

$$\mathbb{P}(2C_A s_0 |\hat{\Sigma}_n - \Sigma|_\infty > 1) \le C\Psi(p, n) + Cp^2 \exp\left(-\frac{cn}{s_0^2}\right),$$

and

$$\mathbb{P}(C\sqrt{n} s_0^2 |\hat{\Sigma}_n - \Sigma|_\infty^2 > \eta_1) \le C\Psi(p, n) + Cp^2 \exp(-cK_1 \log(pn)).$$

With the above choice of K_1, the last exponential term is bounded by $C(pn)^{-1}$. Thus

$$\mathbb{P}(|\hat{T}_n - T^{\mathrm{L}}_n| > \eta_1) \le C\Psi(p, n) + Cp^2 \exp\left(-\frac{cn}{s_0^2}\right) + C(pn)^{-1}. \tag{16}$$

We next control the plug-in block-score error. By Proposition 2, summing the pointwise bound over at most l terms gives

$$\Lambda_{n,l} \le C_1' \sqrt{l} s_0 |\hat{\Sigma}_n - \Sigma|_\infty (M_n^2 + 1)$$

on the event $2C_A s_0 |\hat{\Sigma}_n - \Sigma|_\infty \le 1$. Therefore

$$\mathbb{P}(\Lambda_{n,l} > \eta_2) \le \mathbb{P}(2C_A s_0 |\hat{\Sigma}_n - \Sigma|_\infty > 1) + \mathbb{P}(C_1' \sqrt{l} s_0 |\hat{\Sigma}_n - \Sigma|_\infty (M_n^2 + 1) > \eta_2).$$

Define

$$\mathcal{H}_n = \left\{\sqrt{n} |\hat{\Sigma}_n - \Sigma|_\infty \le C_1 \sqrt{\log(pn)}\right\} \cap \left\{M_n^2 + 1 \le C_2 \log(pn)\right\},$$

where C_1 and C_2 are large fixed constants. By the covariance Gaussian approximation theorem of Zhai et al. (2026) and the Gaussian tail bound for the corresponding approximating vector,

$$\mathbb{P}\left(\sqrt{n} |\hat{\Sigma}_n - \Sigma|_\infty > C_1 \sqrt{\log(pn)}\right) \le C\Psi(p, n) + C(pn)^{-1}.$$

Also, Condition 1 implies that each coordinate of X_t is a centered Gaussian variable with variance uniformly bounded by a constant depending only on C_0. Hence a union bound gives $\mathbb{P}(M_n > x) \leq 2pn\exp(-cx^2)$, and therefore $\mathbb{P}(M_n^2 + 1 > C_2\log(pn)) \leq C(pn)^{-1}$. On $\mathcal{H}_n$,

$$C_1'\sqrt{l}s_0|\hat{\Sigma}_n - \Sigma|_\infty(M_n^2+1) \leq Cs_0\sqrt{\frac{l}{n}}\log^{3/2}(pn) \leq \eta_2.$$

It follows that

$$\mathbb{P}(\Lambda_{n,l} > \eta_2) \leq C\Psi(p,n) + Cp^2\exp\left(-\frac{cn}{s_0^2}\right) + C(pn)^{-1}. \tag{17}$$

Finally, $\eta_1\sqrt{\log p} \leq Cs_0^2\log^{3/2}(pn)/\sqrt{n}$ and $\eta_2\sqrt{\log p} \leq Cs_0\sqrt{l/n}\log^2(pn)$.

Combining these bounds with (15), (16), and (17), we obtain

$$\begin{aligned}\mathbb{P}\Bigg(\sup_{u\in\mathbb{R}}|\hat{F}_{n,l}(u) - \mathbb{P}(\hat{T}_n \leq u)| &> C\Psi_B^{1-\epsilon}(p,n,\epsilon) + C\Psi(p,n) + Cp^2\exp\left(-\frac{cn}{s_0^2}\right) \\ &\quad + C(pn)^{-1} + C\frac{s_0^2\log^{3/2}(pn)}{\sqrt{n}} + Cs_0\sqrt{\frac{l}{n}}\log^2(pn)\Bigg) \\ &\leq 3\Psi_B^\epsilon(p,n,\epsilon) + C\left\{\Psi(p,n) + p^2\exp\left(-\frac{cn}{s_0^2}\right) + (pn)^{-1}\right\}.\end{aligned}$$

This is already a valid probability bound, but we shall do the following simplification. For brevity, write $R_f = \Psi(p,n) + p^2\exp(-cn/s_0^2) + (pn)^{-1}$. Also note that the $(pn)^{-1}$ term can be absorbed by the $s_0^2\log^{3/2}(pn)/n^{1/2}$ term, and the $\Psi(p,n)$ term can be absorbed by the $\Psi_B(p,n,\epsilon)$ term for any $\epsilon \in (0,1)$. The threshold on the left-hand side of the preceding display is thus bounded by $C\mathfrak{R}_{n,l}^{1-\epsilon}(\epsilon)$ whenever $\mathfrak{R}_{n,l}(\epsilon) \leq 1$. Meanwhile, since $\mathfrak{R}_{n,l}(\epsilon) \geq \Psi_B(p,n,\epsilon) + R_f$, the concavity of $t \mapsto t^\epsilon$ gives

$$\mathfrak{R}_{n,l}^\epsilon(\epsilon) - \Psi_B^\epsilon(p,n,\epsilon) \geq \epsilon\mathfrak{R}_{n,l}^{\epsilon-1}(\epsilon)[\mathfrak{R}_{n,l}(\epsilon) - \Psi_B(p,n,\epsilon)] \geq \epsilon\mathfrak{R}_{n,l}^{\epsilon-1}(\epsilon)R_f.$$

There exists a $r_\epsilon \in (0,1]$ small enough (concretely, we may choose $r_\epsilon = 1 \wedge (3\epsilon/C)^{1/(1-\epsilon)}$) such that $CR_f \leq 3[\mathfrak{R}_{n,l}^\epsilon(\epsilon) - \Psi_B^\epsilon(p,n,\epsilon)]$ whenever $\mathfrak{R}_{n,l}(\epsilon) \leq r_\epsilon$. Then we have

$$\mathbb{P}\left(\sup_{u\in\mathbb{R}}|\hat{F}_{n,l}(u) - \mathbb{P}(\hat{T}_n \leq u)| > C\mathfrak{R}_{n,l}(\epsilon)^{1-\epsilon}\right) \leq 3\mathfrak{R}_{n,l}(\epsilon)^\epsilon.$$

This finishes the proof of Theorem 2.

Validity of the bootstrap critical value. We now establish the level statement following Theorem 2 for a nonempty tested coordinate set $\mathcal{J}_0$ and a fixed $0 < \alpha < 1$. Write

$$F_n(u) = \mathbb{P}_{H_0}(\hat{T}_n \leq u), \qquad H_n(u) = \mathbb{P}(|G|_\infty \leq u), \qquad a_n = \sup_{u\in\mathbb{R}}|F_n(u) - H_n(u)|.$$

Condition 5 ensures that H_n is continuous, even if the covariance matrix of G is singular. Let $q_\gamma^G = \inf\{u \in \mathbb{R} : H_n(u) \geq \gamma\}$ for $0 < \gamma < 1$, so $H_n(q_\gamma^G) = \gamma$. Also define

$$\Delta_{n,l} = \sup_{u\in\mathbb{R}}|\hat{F}_{n,l}(u) - F_n(u)|, \qquad \eta_n = C\mathfrak{R}_{n,l}(\epsilon)^{1-\epsilon}, \qquad d_n = \eta_n + a_n.$$

On the event $\Delta_{n,l} \leq \eta_n$, we have $\sup_u|\hat{F}_{n,l}(u) - H_n(u)| \leq d_n$. Whenever $d_n < \min\{\alpha, 1-\alpha\}$, the definitions of the quantiles therefore imply

$$q_{1-\alpha-d_n}^G \leq \hat{q}_{1-\alpha} \leq q_{1-\alpha+d_n}^G.$$

Comparing with these deterministic quantiles and using $|F_n - H_n|_\infty \leq a_n$ gives

$$\alpha - d_n - a_n - \mathbb{P}(\Delta_{n,l} > \eta_n) \leq \mathbb{P}_{H_0}(\hat{T}_n > \hat{q}_{1-\alpha}) \leq \alpha + d_n + a_n + \mathbb{P}(\Delta_{n,l} > \eta_n).$$

This argument requires continuity only of the Gaussian target, not of F_n, and does not require independence between the statistic and the bootstrap critical value. Theorem 1 gives $a_n \leq C\mathfrak{R}_{n,l}(\epsilon)$ when $\mathfrak{R}_{n,l}(\epsilon) \leq 1$, and Theorem 2 gives $\mathbb{P}(\Delta_{n,l} > \eta_n) \leq 3\mathfrak{R}_{n,l}(\epsilon)^\epsilon$. Hence, after enlarging C if necessary, whenever $\mathfrak{R}_{n,l}(\epsilon) \to 0$,

$$\left|\mathbb{P}_{H_0}(\hat{T}_n > \hat{q}_{1-\alpha}) - \alpha\right| \leq C\mathfrak{R}_{n,l}(\epsilon)^{1-\epsilon} + 3\mathfrak{R}_{n,l}(\epsilon)^\epsilon \to 0.$$

E Proof of Supplemental Theory

E.1 Proof of Corollary 5

Proof. We reduce the random-null problem to the fixed-null problem covered by Theorems 1 and 2. Let

$$\mathcal{A}_{n_1} = \{\hat{E}^{(1)} = E^{(1)}\}.$$

By assumption, $\mathbb{P}(\mathcal{A}_{n_1}^c) = \pi_{n_1} \to 0$. Under $H_0^{1\to2}$, on $\mathcal{A}_{n_1}$ we have $E^{(2)} \subseteq E^{(1)} = \hat{E}^{(1)}$. Thus $\hat{T}^{1\to2}$ is exactly the direct statistic computed from the second sample with the fixed null graph $E^{(1)}$. Denote this fixed-null statistic by $T_{n_2}^{1\to2}$. The bootstrap CDF, its critical value, and the bootstrap p-value likewise coincide with their fixed-reference counterparts on $\mathcal{A}_{n_1}$. These pathwise identities and the resulting unconditional probability bounds hold without cross-sample independence. For every $u \in \mathbb{R}$,

$$|\mathbb{P}_{H_0^{1\to2}}(\hat{T}^{1\to2} \le u) - \mathbb{P}(T_{n_2}^{1\to2} \le u)| \le \mathbb{P}(\mathcal{A}_{n_1}^c) = \pi_{n_1}.$$

Therefore,

$$\begin{aligned}\sup_{u\in\mathbb{R}} &|\mathbb{P}_{H_0^{1\to2}}(\hat{T}^{1\to2} \le u) - \mathbb{P}(|G^{1\to2}|_\infty \le u)| \\ &\le \sup_{u\in\mathbb{R}} |\mathbb{P}(T_{n_2}^{1\to2} \le u) - \mathbb{P}(|G^{1\to2}|_\infty \le u)| + \pi_{n_1}.\end{aligned}$$

Theorem 1, applied to the second sample with the fixed null graph $E^{(1)}$, gives

$$\begin{aligned}\sup_{u\in\mathbb{R}} &|\mathbb{P}(T_{n_2}^{1\to2} \le u) - \mathbb{P}(|G^{1\to2}|_\infty \le u)| \\ &\le C\Psi(p, n_2) + Cp^2 \exp\{-cn_2/(s_0^{(1)})^2\} + C\frac{(s_0^{(1)})^2 \log^{3/2}(pn_2)}{\sqrt{n_2}}.\end{aligned}$$

This proves the Gaussian approximation bound.

We next prove the bootstrap bound. Let $F_{n_2,l}^{(1)}$ be the block bootstrap CDF computed from the second sample with fixed null graph $E^{(1)}$. Write $p_{1\to2}^\circ = 1 - F_{n_2,l}^{(1)}(T_{n_2}^{1\to2}-)$. The pathwise equality $p_{1\to2} = p_{1\to2}^\circ$ on $\mathcal{A}_{n_1}$ gives, for every $u \in [0, 1]$,

$$|\mathbb{P}(p_{1\to2} \le u) - \mathbb{P}(p_{1\to2}^\circ \le u)| \le \pi_{n_1}.$$

Let $\mathcal{D}_{n,l}^{1\to2} = \sup_{u\in\mathbb{R}} |\hat{F}_l^{1\to2}(u) - \mathbb{P}_{H_0^{1\to2}}(\hat{T}^{1\to2} \le u)|$. Define

$$D_{n,l}^\circ = \sup_{u\in\mathbb{R}} |F_{n_2,l}^{(1)}(u) - \mathbb{P}(T_{n_2}^{1\to2} \le u)|.$$

On $\mathcal{A}_{n_1}$, we have $\hat{F}_l^{1\to2} = F_{n_2,l}^{(1)}$. Hence, on $\mathcal{A}_{n_1}$,

$$\begin{aligned}\mathcal{D}_{n,l}^{1\to2} &\le D_{n,l}^\circ + \sup_{u\in\mathbb{R}} |\mathbb{P}_{H_0^{1\to2}}(\hat{T}^{1\to2} \le u) - \mathbb{P}(T_{n_2}^{1\to2} \le u)| \\ &\le D_{n,l}^\circ + \pi_{n_1}.\end{aligned}$$

Let

$$\begin{aligned}R_{n_2,l}^{(1)}(\epsilon) = \Psi_B(p, n_2, \epsilon) + \Psi(p, n_2) + p^2 \exp\{-cn_2/(s_0^{(1)})^2\} + (pn_2)^{-1} \\ + \frac{(s_0^{(1)})^2 \log^{3/2}(pn_2)}{\sqrt{n_2}} + s_0^{(1)}\sqrt{\frac{l}{n_2}} \log^2(pn_2).\end{aligned}$$

Theorem 2, applied to the second sample with fixed null graph $E^{(1)}$, implies

$$\mathbb{P}(D_{n,l}^\circ > CR_{n_2,l}^{(1)}(\epsilon)^{1-\epsilon}) \le CR_{n_2,l}^{(1)}(\epsilon)^\epsilon.$$

Therefore,

$$\begin{aligned}&\mathbb{P}(\mathcal{D}_{n,l}^{1\to2} > C\pi_{n_1} + CR_{n_2,l}^{(1)}(\epsilon)^{1-\epsilon}) \\ &\quad\le \mathbb{P}(\mathcal{A}_{n_1}^c) + \mathbb{P}(D_{n,l}^\circ > CR_{n_2,l}^{(1)}(\epsilon)^{1-\epsilon}) \\ &\quad\le \pi_{n_1} + CR_{n_2,l}^{(1)}(\epsilon)^\epsilon.\end{aligned}$$

By the definition of $\mathfrak{R}_{n,l}^{1\to 2}(\epsilon)$ and by enlarging C if necessary, the last display implies

$$\mathbb{P}(\mathcal{D}_{n,l}^{1\to 2} > C\mathfrak{R}_{n,l}^{1\to 2}(\epsilon)^{1-\epsilon}) \le C\mathfrak{R}_{n,l}^{1\to 2}(\epsilon)^{\epsilon}$$

whenever $\mathfrak{R}_{n,l}^{1\to 2}(\epsilon) \le 1$. This proves the bootstrap bound.

It remains to prove the exact asymptotic level. Let

$$F_n^{1\to 2}(u) = \mathbb{P}_{H_0^{1\to 2}}(\hat{T}^{1\to 2} \le u), \qquad q_n^{1\to 2}(\gamma) = \inf\{u \in \mathbb{R} : F_n^{1\to 2}(u) \ge \gamma\}.$$

The bootstrap bound implies

$$\sup_{u\in\mathbb{R}} |\hat{F}_l^{1\to 2}(u) - F_n^{1\to 2}(u)| \to 0$$

in probability. Fix $\eta > 0$. With probability tending to one,

$$q_n^{1\to 2}(1-\alpha-\eta) \le \hat{q}_{1-\alpha}^{1\to 2} \le q_n^{1\to 2}(1-\alpha+\eta).$$

Hence

$$\alpha - \eta + o(1) \le \mathbb{P}_{H_0^{1\to 2}}(\hat{T}^{1\to 2} > \hat{q}_{1-\alpha}^{1\to 2}) \le \alpha + \eta + o(1).$$

The Gaussian approximation already proved shows that $F_n^{1\to 2}$ converges uniformly to the distribution function of $|G^{1\to 2}|_\infty$. Since this limiting distribution is continuous at its $(1-\alpha)$-quantile, we may let $\eta \downarrow 0$. This gives

$$\mathbb{P}_{H_0^{1\to 2}}(\hat{T}^{1\to 2} > \hat{q}_{1-\alpha}^{1\to 2}) \to \alpha.$$

This finishes the proof. □

E.2 Proof of Lemma 1

We prove the result by comparing the plug-in block estimator with an oracle block estimator for the local-inverse score process. Let

$$d_\omega = \sum_{j=1}^{p} |S_{0,j}| \le ps_0$$

be the number of local-inverse coordinates. Define the sample mean $\bar{\mathcal{X}}_{ij}^\omega = n^{-1}\sum_{t=1}^n \mathcal{X}_{t,ij}^\omega$ and the oracle centered block estimator as an intermediate quantity,

$$\bar{L}_{ij}^\omega = \frac{1}{n-l+1}\sum_{r=l}^{n}\left(l^{-1/2}\sum_{t=r-l+1}^{r}(\mathcal{X}_{t,ij}^\omega - \bar{\mathcal{X}}_{ij}^\omega)\right)^2.$$

The first step of the proof is to show that $\max_{1\le j\le p}\max_{i\in S_{0,j}}|\bar{L}_{ij}^\omega - L_{ij}^\omega| = o_{\mathbb{P}}(1)$. Too see this, for $i \in S_{0,j}$ and $i' \in S_{0,j'}$, write $\gamma_{ij,i'j'}^\omega(h) = \mathrm{Cov}(\mathcal{X}_{0,ij}^\omega, \mathcal{X}_{h,i'j'}^\omega)$. By Isserlis' formula,

$$\gamma_{ij,i'j'}^\omega(h) = (v_{0,ij}^\top\Gamma_h v_{0,i'j'})(\omega_{0,j}^\top\Gamma_h\omega_{0,j'}) + (v_{0,ij}^\top\Gamma_h\omega_{0,j'})(\omega_{0,j}^\top\Gamma_h v_{0,i'j'}).$$

Under Conditions 2 and 3, the vectors $v_{0,ij}$ and $\omega_{0,j}$ have uniformly bounded ℓ_1-norms. Combining this fact with Condition 1 gives

$$\sup_{1\le j,j'\le p}\sup_{i\in S_{0,j},i'\in S_{0,j'}}|\gamma_{ij,i'j'}^\omega(h)| \le C(1\vee|h|)^{-2[\beta\wedge(2\beta-1)]}\log^2(e+|h|).$$

Since $\beta > 3/4$, the summation of the right-hand side over $h \in \mathbb{N}$ is finite, and thus the long-run variance L_{ij}^ω is well defined. Let $Z_{r,ij}^0 = l^{-1/2}\sum_{t=r-l+1}^r \mathcal{X}_{t,ij}^\omega$. By stationarity,

$$\mathbb{E}[(Z_{r,ij}^0)^2] = \sum_{|h|<l}\left(1-\frac{|h|}{l}\right)\gamma_{ij,ij}^\omega(h).$$

By simple calculation, the deterministic bias of the block variance satisfies

$$\max_{1\le j\le p}\max_{i\in S_{0,j}}\left|\mathbb{E}[(Z_{r,ij}^0)^2] - L_{ij}^\omega\right| \le Cl^{-[(4\beta-3)\wedge 1]}\log^3(l).$$

It remains to control the stochastic fluctuation of the overlapping block average. The variables $(Z^0_{r,ij})^2 - \mathbb{E}[(Z^0_{r,ij})^2] := W_{r,ij}$ are centered Gaussian chaos of order at most four. By hypercontractivity for Gaussian chaoses, for every $q \geq 2$,

$$\left\| \frac{1}{n-l+1} \sum_{r=l}^{n} W_{r,ij} \right\|_q \leq Cq^2 \left\| \frac{1}{n-l+1} \sum_{r=l}^{n} W_{r,ij} \right\|_2 .$$

It remains to bound the right-hand side. By Isserlis' formula and the covariance decay bound derived above, the covariance sequence of the block-chaos variables has total mass of order l, uniformly over i, j. More precisely,

$$\sup_{1\leq j\leq p} \sup_{i\in S_{0,j}} \sum_{s=l}^{n} |\mathrm{Cov}(W_{r,ij}, W_{s,ij})| \leq Cl, \qquad l \leq r \leq n.$$

Indeed, there are $O(l)$ indices s for which the two blocks overlap or are within distance l, and for these indices the covariance is bounded by a constant. For the remaining indices, every cross-block covariance factor contains the separation between the two blocks, and Isserlis' formula reduces the bound to the summable autocovariance envelope of the local-inverse score process. Consequently,

$$\mathrm{Var}\left(\frac{1}{n-l+1} \sum_{r=l}^{n} W_{r,ij} \right) \leq \frac{1}{(n-l+1)^2} \sum_{r,s=l}^{n} |\mathrm{Cov}(W_{r,ij}, W_{s,ij})| \leq C \frac{l}{n},$$

and therefore

$$\max_{1\leq j\leq p} \max_{i\in S_{0,j}} \left\| \frac{1}{n-l+1} \sum_{r=l}^{n} W_{r,ij} \right\|_q \leq Cq^2 \sqrt{\frac{l}{n}}.$$

Taking $q = 2\log(ps_0n)$, using $s_0 \leq p$, and applying Markov's inequality gives

$$\max_{1\leq j\leq p} \max_{i\in S_{0,j}} \left| \frac{1}{n-l+1} \sum_{r=l}^{n} \left\{ (Z^0_{r,ij})^2 - \mathbb{E}[(Z^0_{r,ij})^2] \right\} \right| = O_{\mathbb{P}} \left(\sqrt{\frac{l}{n}} \log^2(pn) \right).$$

We now account for the sample centering by $\bar{\mathcal{X}}^{\omega}_{n,ij} = n^{-1} \sum_{t=1}^{n} \mathcal{X}^{\omega}_{t,ij}$. The Gaussian approximation argument used for the local-inverse score vector gives

$$\mathfrak{A}_n := \max_{1\leq j\leq p} \max_{i\in S_{0,j}} |\bar{\mathcal{X}}^{\omega}_{n,ij}| = O_{\mathbb{P}} \left(\sqrt{\frac{\log(pn)}{n}} \right).$$

For each r, we have

$$l^{-1/2} \sum_{t=r-l+1}^{r} (\mathcal{X}^{\omega}_{t,ij} - \bar{\mathcal{X}}^{\omega}_{n,ij}) = Z^0_{r,ij} - \sqrt{l}\, \bar{\mathcal{X}}^{\omega}_{n,ij}.$$

Therefore,

$$\begin{aligned} &\frac{1}{n-l+1} \sum_{r=l}^{n} \left[\left(l^{-1/2} \sum_{t=r-l+1}^{r} (\mathcal{X}^{\omega}_{t,ij} - \bar{\mathcal{X}}^{\omega}_{n,ij}) \right)^2 - (Z^0_{r,ij})^2 \right] \\ &\quad = -2\sqrt{l}\, \bar{\mathcal{X}}^{\omega}_{n,ij} \left(\frac{1}{n-l+1} \sum_{r=l}^{n} Z^0_{r,ij} \right) + l(\bar{\mathcal{X}}^{\omega}_{n,ij})^2. \end{aligned}$$

By Cauchy–Schwarz,

$$\left| \frac{1}{n-l+1} \sum_{r=l}^{n} Z^0_{r,ij} \right| \leq \left(\frac{1}{n-l+1} \sum_{r=l}^{n} (Z^0_{r,ij})^2 \right)^{1/2}.$$

The bias and stochastic fluctuation bounds above, together with the uniform upper bound in Condition 6, imply

$$\mathfrak{B}_n := \max_{1\leq j\leq p} \max_{i\in S_{0,j}} \left(\frac{1}{n-l+1} \sum_{r=l}^{n} (Z^0_{r,ij})^2 \right)^{1/2} = O_{\mathbb{P}}(1).$$

Consequently,

$$\begin{aligned}\max_{1\le j\le p}\max_{i\in S_{0,j}}\left|\frac{1}{n-l+1}\sum_{r=l}^{n}\left[\left(l^{-1/2}\sum_{t=r-l+1}^{r}(\mathcal{X}^{\omega}_{t,ij}-\bar{\mathcal{X}}^{\omega}_{n,ij})\right)^2-(Z^0_{r,ij})^2\right]\right| \\ \le 2\sqrt{l}\mathfrak{A}_n\mathfrak{B}_n+l\mathfrak{A}_n^2 \\ = O_{\mathbb{P}}\left(\sqrt{\frac{l\log(pn)}{n}}+\frac{l\log(pn)}{n}\right).\end{aligned}$$

Combining the bias, stochastic fluctuation, and sample-centering terms gives

$$\max_{1\le j\le p}\max_{i\in S_{0,j}}|\bar{L}^{\omega}_{ij}-L^{\omega}_{ij}|=O_{\mathbb{P}}\left(l^{-[(4\beta-3)\wedge 1]}\log^3(l)+\sqrt{\frac{l}{n}}\log^2(pn)\right)=o_{\mathbb{P}}(1).$$

This finishes the first step of the proof. We now move on to the second step, i.e., comparing the plug-in block estimator $\hat{L}^{\omega}_{ij}$ with the oracle block estimator $\bar{L}^{\omega}_{ij}$. Define

$$Z_{r,ij}=l^{-1/2}\sum_{t=r-l+1}^{r}(\mathcal{X}^{\omega}_{t,ij}-\bar{\mathcal{X}}^{\omega}_{n,ij}),\qquad \hat{Z}_{r,ij}=l^{-1/2}\sum_{t=r-l+1}^{r}\hat{\mathcal{X}}^{\omega}_{t,ij}.$$

Then

$$\bar{L}^{\omega}_{ij}=\frac{1}{n-l+1}\sum_{r=l}^{n}Z^2_{r,ij},\qquad \hat{L}^{\omega}_{ij}=\frac{1}{n-l+1}\sum_{r=l}^{n}\hat{Z}^2_{r,ij}.$$

On the event $2C_As_0|\hat{\Sigma}_n-\Sigma|_\infty\le 1$, the same inverse perturbation argument as in Proposition 1 gives

$$\max_{1\le j\le p}|\hat{\omega}_{0,j}-\omega_{0,j}|_1\le Cs_0|\hat{\Sigma}_n-\Sigma|_\infty,$$

and

$$\max_{1\le j\le p}\max_{i\in S_{0,j}}|\hat{v}_{0,ij}-v_{0,ij}|_1\le Cs_0|\hat{\Sigma}_n-\Sigma|_\infty.$$

Moreover,

$$\mathcal{X}^{\omega}_{t,ij}-\bar{\mathcal{X}}^{\omega}_{n,ij}=v_{0,ij}^\top(X_tX_t^\top-\hat{\Sigma}_n)\omega_{0,j}.$$

Therefore,

$$\hat{\mathcal{X}}^{\omega}_{t,ij}-(\mathcal{X}^{\omega}_{t,ij}-\bar{\mathcal{X}}^{\omega}_{n,ij})=(\hat{v}_{0,ij}-v_{0,ij})^\top(X_tX_t^\top-\hat{\Sigma}_n)\hat{\omega}_{0,j}+v_{0,ij}^\top(X_tX_t^\top-\hat{\Sigma}_n)(\hat{\omega}_{0,j}-\omega_{0,j}).$$

Using the uniform ℓ_1-bounds on $v_{0,ij}$, $\omega_{0,j}$, $\hat{v}_{0,ij}$, and $\hat{\omega}_{0,j}$, together with

$$|X_tX_t^\top-\hat{\Sigma}_n|_\infty\le C(M_n^2+1),$$

we obtain

$$\max_{1\le t\le n}\max_{1\le j\le p}\max_{i\in S_{0,j}}\left|\hat{\mathcal{X}}^{\omega}_{t,ij}-(\mathcal{X}^{\omega}_{t,ij}-\bar{\mathcal{X}}^{\omega}_{n,ij})\right|\le Cs_0|\hat{\Sigma}_n-\Sigma|_\infty(M_n^2+1).$$

Consequently,

$$\max_{l\le r\le n}\max_{1\le j\le p}\max_{i\in S_{0,j}}|\hat{Z}_{r,ij}-Z_{r,ij}|\le C\sqrt{l}s_0|\hat{\Sigma}_n-\Sigma|_\infty(M_n^2+1).$$

By the Gaussian approximation bound in Theorem 1 of Zhai et al. (2026) and the Gaussian maximal tail bound,

$$|\hat{\Sigma}_n-\Sigma|_\infty=O_{\mathbb{P}}\left(\sqrt{\frac{\log(pn)}{n}}\right),\qquad M_n^2+1=O_{\mathbb{P}}(\log(pn)).$$

Hence

$$\max_{l\le r\le n}\max_{1\le j\le p}\max_{i\in S_{0,j}}|\hat{Z}_{r,ij}-Z_{r,ij}|=O_{\mathbb{P}}\left(s_0\sqrt{\frac{l}{n}}\log^2(pn)\right)=o_{\mathbb{P}}(1),$$

where the last step follows from $\mathfrak{R}_{n,l}(\epsilon)\to 0$.

Since the oracle bound above and Condition 6 imply

$$\max_{1\le j\le p}\max_{i\in S_{0,j}}\frac{1}{n-l+1}\sum_{r=l}^{n}Z_{r,ij}^2=O_{\mathbb{P}}(1),$$

we have

$$\begin{aligned}
&\max_{1\le j\le p}\max_{i\in S_{0,j}}|\hat{L}_{ij}^{\omega}-\bar{L}_{ij}^{\omega}|\\
&\quad\le 2\max_{1\le j\le p}\max_{i\in S_{0,j}}\left(\frac{1}{n-l+1}\sum_{r=l}^{n}Z_{r,ij}^2\right)^{1/2}\max_{1\le j\le p}\max_{i\in S_{0,j}}\left(\frac{1}{n-l+1}\sum_{r=l}^{n}(\hat{Z}_{r,ij}-Z_{r,ij})^2\right)^{1/2}\\
&\qquad+\max_{1\le j\le p}\max_{i\in S_{0,j}}\frac{1}{n-l+1}\sum_{r=l}^{n}(\hat{Z}_{r,ij}-Z_{r,ij})^2\\
&\quad=o_{\mathbb{P}}(1).
\end{aligned}$$

This finishes the second step of the proof. Combining this result with the first step's result gives

$$\max_{1\le j\le p}\max_{i\in S_{0,j}}|\hat{L}_{ij}^{\omega}-L_{ij}^{\omega}|=o_{\mathbb{P}}(1).$$

Finally, Condition 6 gives $0<c_\omega\le L_{ij}^{\omega}\le C_\omega<\infty$ uniformly over $1\le j\le p$ and $i\in S_{0,j}$. Since the square-root map is Lipschitz on compact subsets of $(0,\infty)$,

$$\max_{1\le j\le p}\max_{i\in S_{0,j}}\left|\frac{\hat{\tau}_{ij}}{\tau_{ij}}-1\right|=\max_{1\le j\le p}\max_{i\in S_{0,j}}\left|\sqrt{\frac{\hat{L}_{ij}^{\omega}}{L_{ij}^{\omega}}}-1\right|=o_{\mathbb{P}}(1).$$

This finishes the proof of Lemma 1.

E.3 Proof of Lemma 2

We prove the two screening claims separately for any sequence satisfying $\delta_n\to\infty$ and $\delta_n+\sqrt{\log p}=o(\sqrt{n})$, which includes the stated choice $\delta_n=\sqrt{\log n}$. The key point is that a nonzero graph-compatible coordinate is much larger than its long-run standard error, while a zero coordinate remains on the stochastic standard-error scale.

By Condition 6, $\tau_{ij}\asymp n^{-1/2}$ uniformly over $1\le j\le p$ and $i\in S_{0,j}$. The same Gaussian approximation argument used in Theorem 1, applied to the local-inverse score vector $(\mathcal{X}_{t,ij}^{\omega})_{i\in S_{0,j},1\le j\le p}$, gives

$$\max_{1\le j\le p}\max_{i\in S_{0,j}}\left|\frac{\hat{\omega}_{0,ij}-\omega_{0,ij}}{\tau_{ij}}\right|=O_{\mathbb{P}}(\sqrt{\log p}).$$

Together with Lemma 1, the same bound holds after replacing τ_{ij} by $\hat{\tau}_{ij}$.

We first work under $H_0: E=E_0$. By the fixed-signal condition, there exists a constant $c_*>0$ such that $\min_{1\le j\le p}\min_{i\in S_{0,j}}|\omega_{0,ij}|\ge c_*$. Since $\tau_{ij}\asymp n^{-1/2}$, this implies

$$\min_{1\le j\le p}\min_{i\in S_{0,j}}\frac{|\omega_{0,ij}|}{\tau_{ij}}\asymp\sqrt{n}.$$

Under $\mathfrak{R}_{n,l}(\epsilon)\to 0$, we have $\log p=o(n)$. By the assumed threshold condition, $\sqrt{\log p}+\delta_n=o(\sqrt{n})$. It follows that

$$\min_{1\le j\le p}\min_{i\in S_{0,j}}\frac{|\hat{\omega}_{0,ij}|}{\hat{\tau}_{ij}}>\delta_n$$

with probability tending to one. That si equivalent to $\hat{\mathbf{\Delta}}_j=0$ for every $1\le j\le p$ with probability tending to one.

Now consider a fixed included alternative $E\subsetneq E_0$. Because the true graph is contained in the proposed graph, the restricted population inverse $\boldsymbol{\omega}_{0,j}$ coincides with the corresponding true precision column. Because the inclusion is strict, there exist j_0 and $i_0\in S_{0,j_0}$ such that $\omega_{0,i_0j_0}=0$. For

this fixed coordinate, the local-inverse expansion and the variance condition (Condition 6) give $\hat{\omega}_{0,i_0j_0}/\tau_{i_0j_0} = O_{\mathbb{P}}(1)$, and furthermore by Lemma 1,

$$\frac{\hat{\omega}_{0,i_0j_0}}{\hat{\tau}_{i_0j_0}} = O_{\mathbb{P}}(1).$$

Since $\delta_n \to \infty$,

$$\mathbb{P}\left(\frac{|\hat{\omega}_{0,i_0j_0}|}{\hat{\tau}_{i_0j_0}} \le \delta_n\right) \to 1.$$

Thus $\hat{\Delta}_{i_0j_0} = C_\Delta\sqrt{\log p}$ with probability tending to one. Equivalently, $\hat{\mathbf{\Delta}}_{j_0} \neq 0$ with probability tending to one. This finishes the proof of Lemma 2.

E.4 Proof of Theorem 6

We first combine the uniform consistency of the long-run standard errors in Lemma 1 with the screening result in Lemma 2 to establish null validity, and then control the pumped statistic under fixed alternatives to prove power consistency.

We first focus on the null behavior. The first assertion of the theorem is directly implied by Lemma 2, where we have

$$\mathbb{P}_{H_0}\left(\hat{\mathbf{\Delta}}_j = 0 \text{ for all } 1 \le j \le p\right) \to 1.$$

On this event, $\tilde{\boldsymbol{\omega}}_{0,j} = \hat{\boldsymbol{\omega}}_{0,j}$ for every j. For every $i \in S_{0,j}$, the sample normal equations give $e_i^\top \hat{\Sigma}_n \hat{\boldsymbol{\omega}}_{0,j} - \mathbf{1}\{i=j\} = 0$. Therefore, the maximum in $\tilde{T}_n$ over all $1 \le i,j \le p$ is equivalent to the maximum over $(i,j) \in \mathcal{J}_0$. Thus $\tilde{T}_n = \hat{T}_n$ with probability tending to one under H_0. Consequently, the Gaussian approximation and block-bootstrap validity for $\tilde{T}_n$ now follow directly from Theorems 1 and 2. Indeed,

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}_{H_0}(\tilde{T}_n \le u) - \mathbb{P}(|G|_\infty \le u)\right| \le \sup_{u\in\mathbb{R}}\left|\mathbb{P}_{H_0}(\hat{T}_n \le u) - \mathbb{P}(|G|_\infty \le u)\right| + \mathbb{P}_{H_0}(\tilde{T}_n \neq \hat{T}_n),$$

and the right-hand side converges to zero. Similarly,

$$\sup_{u\in\mathbb{R}}\left|\hat{F}_{n,l}(u) - \mathbb{P}_{H_0}(\tilde{T}_n \le u)\right| \le \sup_{u\in\mathbb{R}}\left|\hat{F}_{n,l}(u) - \mathbb{P}_{H_0}(\hat{T}_n \le u)\right| + \mathbb{P}_{H_0}(\tilde{T}_n \neq \hat{T}_n),$$

which converges to zero in probability. The asymptotic size statement follows from the same quantile-bracketing argument used for the direct statistic.

It remains to prove the power guarantee. Recall the simplified notation that we defined earlier, $A_j = B_{0,j}^\top \Sigma B_{0,j}$ and $\hat{A}_j = B_{0,j}^\top \hat{\Sigma}_n B_{0,j}$. Consider the event $\mathcal{G}_n = \{2C_A s_0 |\hat{\Sigma}_n - \Sigma|_\infty \le 1\}$. By the covariance concentration bound used in Theorems 1 and 2, we have $\mathbb{P}(\mathcal{G}_n) \to 1$. On this event,

$$|A_j^{-1}(\hat{A}_j - A_j)|_1 \le |A_j^{-1}|_1 |\hat{A}_j - A_j|_1 \le C_A s_0 |\hat{\Sigma}_n - \Sigma|_\infty \le \frac{1}{2}.$$

Hence $\hat{A}_j$ is invertible, and the Neumann-series bound gives

$$|\hat{A}_j^{-1}|_1 \le \frac{|A_j^{-1}|_1}{1 - |A_j^{-1}(\hat{A}_j - A_j)|_1} \le 2C_A.$$

Since $\hat{A}_j$ is symmetric positive definite on $\mathcal{G}_n$,

$$\lambda_{\min}(\hat{A}_j) = |\hat{A}_j^{-1}|_{\text{op}}^{-1} \ge |\hat{A}_j^{-1}|_1^{-1} \ge \frac{1}{2C_A}.$$

Suppose that $\mathcal{G}_n$ holds and that $\hat{\mathbf{\Delta}}_j \neq 0$ for some j. Let $x_j = B_{0,j}^\top \hat{\mathbf{\Delta}}_j$. Then $x_j \neq 0$, and every nonzero coordinate of x_j is equal to C_n. By the identity $B_{0,j}^\top(\hat{\Sigma}_n \hat{\boldsymbol{\omega}}_{0,j} - e_j) = 0$, we have

$$B_{0,j}^\top(\hat{\Sigma}_n \tilde{\boldsymbol{\omega}}_{0,j} - e_j) = B_{0,j}^\top \hat{\Sigma}_n \hat{\mathbf{\Delta}}_j = \hat{A}_j x_j.$$

It follows that

$$\begin{aligned}\max_{1\le i\le p}\left|e_i^\top\hat{\Sigma}_n\tilde{\boldsymbol{\omega}}_{0,j}-\mathbf{1}\{i=j\}\right| &\ge |\hat{A}_j x_j|_\infty\\ &\ge \frac{x_j^\top\hat{A}_j x_j}{|x_j|_1}\\ &\ge \lambda_{\min}(\hat{A}_j)\frac{|x_j|_2^2}{|x_j|_1}\\ &\ge \frac{C_n}{2C_A}.\end{aligned}$$

Consequently, on the event $\mathcal{G}_n$,

$$\hat{\boldsymbol{\Delta}}_j\neq 0 \text{ for some } j \quad\Longrightarrow\quad \tilde{T}_n\ge\frac{\sqrt{n}C_n}{2C_A}. \tag{18}$$

Now we are ready to apply Lemma 2 to prove the power assertion. First consider a fixed included alternative $E\subsetneq E_0$. By Lemma 2, with probability tending to one, there exists j such that $\hat{\boldsymbol{\Delta}}_j\neq 0$. Combining this event with (18) gives

$$\tilde{T}_n\ge\frac{\sqrt{n}C_n}{2C_A}=\frac{\sqrt{n}C_\Delta\sqrt{\log p}}{2C_A}$$

with probability tending to one. Next, we claim that $\hat{q}_{1-\alpha}=O_{\mathbb{P}}(\sqrt{\log(pn)})$. To see this, let $\hat{Z}_r=l^{-1/2}\left|\sum_{t=r-l+1}^{r}\hat{\mathcal{X}}_t\right|_\infty$ for $l\le r\le n$. Then $\hat{q}_{1-\alpha}$ is the empirical $(1-\alpha)$-quantile of $\{\hat{Z}_r: l\le r\le n\}$. Let $Z_r^{\mathrm{L}}=l^{-1/2}\left|\sum_{t=r-l+1}^{r}(\mathcal{X}_t-\bar{\mathcal{X}}_n)\right|_\infty$. By Proposition 2 and the covariance maximal bound,

$$\max_{l\le r\le n}|\hat{Z}_r-Z_r^{\mathrm{L}}|\le C\sqrt{l}s_0|\hat{\Sigma}_n-\Sigma|_\infty(M_n^2+1)=o_{\mathbb{P}}(1).$$

The oracle block statistic Z_r^{L} is a maximum of selected covariance-score block sums of the same derived Gaussian linear process used in Theorem 2. The Gaussian maximal tail bound for these block sums, uniformly over r, gives

$$\mathbb{P}(Z_r^{\mathrm{L}}>K\sqrt{\log(pn)})\to 0$$

for a sufficiently large constant K. By Markov's inequality,

$$\mathbb{P}\left(\frac{1}{n-l+1}\sum_{r=l}^{n}\mathbf{1}\left\{Z_r^{\mathrm{L}}>K\sqrt{\log(pn)}\right\}>\alpha\right)\le\frac{1}{\alpha(n-l+1)}\sum_{r=l}^{n}\mathbb{P}\left(Z_r^{\mathrm{L}}>K\sqrt{\log(pn)}\right).$$

The same selected-coordinate block-tail bound applies uniformly in r, so the right-hand side converges to zero. Therefore we arrive at

$$\hat{q}_{1-\alpha}=O_{\mathbb{P}}(\sqrt{\log(pn)}).$$

Therefore

$$\frac{\sqrt{n}C_\Delta\sqrt{\log p}/(2C_A)}{\hat{q}_{1-\alpha}}\to\infty$$

in probability for every fixed $C_\Delta>0$. Therefore

$$\mathbb{P}_{H_1}(\tilde{T}_n>\hat{q}_{1-\alpha})\to 1$$

for every fixed included alternative $E\subsetneq E_0$.

There is, however, another case for $E\neq E_0$ not covered by $E\subsetneq E_0$. That is, a fixed alternative not included in the null graph, namely $E\not\subseteq E_0$. In this case, we cannot directly apply the second assertion of Lemma 2. Instead, we may control the power in the following way. Define

$$d_0=\max_{(i,j)\in\mathcal{J}_0}|e_i^\top\Sigma\boldsymbol{\omega}_{0,j}|.$$

We first show that $d_0>0$. Suppose otherwise that $d_0=0$. Then $e_i^\top\Sigma\boldsymbol{\omega}_{0,j}=0$ for every $i\notin S_{0,j}$ and every $1\le j\le p$. Meanwhile, by the definition of the population restricted inverse,

$$B_{0,j}^\top\Sigma\boldsymbol{\omega}_{0,j}=B_{0,j}^\top e_j.$$

Combining the inside-support equation with the off-support equation gives

$$\Sigma \boldsymbol{\omega}_{0,j} = e_j.$$

Since Σ is invertible, $\boldsymbol{\omega}_{0,j}$ must be the j-th column of the true precision matrix Ω. However, $\boldsymbol{\omega}_{0,j}$ is supported on $S_{0,j}$ by construction. Thus the j-th column of Ω is supported on $S_{0,j}$ for every j, which implies $E \subseteq E_0$. This contradicts $E \not\subseteq E_0$. Therefore $d_0 > 0$.

Next, recall the notation that $\Delta = \hat{\Sigma}_n - \Sigma$ and $b_j = B_{0,j}^\top e_j$ for each j. Since $\boldsymbol{\omega}_{0,j} = B_{0,j} A_j^{-1} b_j$ and $\hat{\boldsymbol{\omega}}_{0,j} = B_{0,j} \hat{A}_j^{-1} b_j$, we obtain $|\boldsymbol{\omega}_{0,j}|_1 \le C_A$ and $|\hat{\boldsymbol{\omega}}_{0,j} - \boldsymbol{\omega}_{0,j}|_1 \le 2C_A^2 s_0 |\Delta|_\infty$. Therefore, uniformly over $(i,j) \in \mathcal{J}_0$,

$$\begin{aligned}
\left| e_i^\top \hat{\Sigma}_n \hat{\boldsymbol{\omega}}_{0,j} - e_i^\top \Sigma \boldsymbol{\omega}_{0,j} \right| &\le \left| e_i^\top \Delta \boldsymbol{\omega}_{0,j} \right| + \left| e_i^\top \hat{\Sigma}_n (\hat{\boldsymbol{\omega}}_{0,j} - \boldsymbol{\omega}_{0,j}) \right| \\
&\le C_A |\Delta|_\infty + \left(|e_i^\top \Sigma B_{0,j}|_1 + |e_i^\top \Delta B_{0,j}|_1 \right) 2C_A^2 s_0 |\Delta|_\infty \\
&\le C_A |\Delta|_\infty + \left(C_\Sigma + s_0 |\Delta|_\infty \right) 2C_A^2 s_0 |\Delta|_\infty \\
&\le C s_0 |\Delta|_\infty,
\end{aligned}$$

where $C > 0$ depends only on C_A and C_Σ. Since $\mathbb{P}_{H_1}(\mathcal{G}_n) \to 1$ and $s_0 |\hat{\Sigma}_n - \Sigma|_\infty = o_{\mathbb{P}_{H_1}}(1)$, it follows that

$$\max_{(i,j) \in \mathcal{J}_0} \left| e_i^\top \hat{\Sigma}_n \hat{\boldsymbol{\omega}}_{0,j} - e_i^\top \Sigma \boldsymbol{\omega}_{0,j} \right| = o_{\mathbb{P}_{H_1}}(1).$$

Consequently, with the event $\mathcal{D}_n = \{\max_{(i,j) \in \mathcal{J}_0} |e_i^\top \hat{\Sigma}_n \hat{\boldsymbol{\omega}}_{0,j} - e_i^\top \Sigma \boldsymbol{\omega}_{0,j}| \le d_0/2\}$, we have $\mathbb{P}_{H_1}(\mathcal{D}_n) \to 1$. Also recall the aforementioned event $\mathcal{G}_n$, and that $\mathbb{P}_{H_1}(\mathcal{G}_n^c) \to 0$. Therefore, it suffices work on the lower bound of $\tilde{T}_n$ on the high-probability event $\mathcal{D}_n \cap \mathcal{G}_n$, which has probability tending to one. We further split this event by $\mathcal{P}_n = \{\hat{\boldsymbol{\Delta}}_j \neq 0 \text{ for some } 1 \le j \le p\}$.

On the event $\mathcal{P}_n$, which covers $\mathcal{P}_n \cap \mathcal{D}_n \cap \mathcal{G}_n$, we have already established that

$$\tilde{T}_n \ge \frac{\sqrt{n} C_\Delta \sqrt{\log p}}{2C_A}.$$

Meanwhile, on the event $\mathcal{P}_n^c \cap \mathcal{D}_n$, which covers $\mathcal{P}_n^c \cap \mathcal{D}_n \cap \mathcal{G}_n$, we have $\tilde{\boldsymbol{\omega}}_{0,j} = \hat{\boldsymbol{\omega}}_{0,j}$ for every j. Since $\mathcal{J}_0$ contains only off-support coordinates, we have $\mathbf{1}\{i = j\} = 0$ for every $(i,j) \in \mathcal{J}_0$. Thus, on $\mathcal{P}_n^c \cap \mathcal{D}_n \cap \mathcal{G}_n$,

$$\begin{aligned}
\tilde{T}_n &\ge \sqrt{n} \max_{(i,j) \in \mathcal{J}_0} |e_i^\top \hat{\Sigma}_n \hat{\boldsymbol{\omega}}_{0,j}| \\
&\ge \sqrt{n} \left[\max_{(i,j) \in \mathcal{J}_0} |e_i^\top \Sigma \boldsymbol{\omega}_{0,j}| - \max_{(i,j) \in \mathcal{J}_0} \left| e_i^\top \hat{\Sigma}_n \hat{\boldsymbol{\omega}}_{0,j} - e_i^\top \Sigma \boldsymbol{\omega}_{0,j} \right| \right] \\
&\ge \frac{\sqrt{n} d_0}{2}.
\end{aligned}$$

Combining the two sub-events, we obtain that on $\mathcal{G}_n \cap \mathcal{D}_n$,

$$\tilde{T}_n \ge \min \left\{ \frac{\sqrt{n} C_\Delta \sqrt{\log p}}{2C_A}, \frac{\sqrt{n} d_0}{2}. \right\}$$

Therefore,

$$\begin{aligned}
\mathbb{P}_{H_1}(\tilde{T}_n \le \hat{q}_{1-\alpha}) \le \mathbb{P}_{H_1}(\mathcal{G}_n^c) + \mathbb{P}_{H_1}(\mathcal{D}_n^c) \\
+ \mathbb{P}_{H_1} \left[\hat{q}_{1-\alpha} \ge \min \left\{ \frac{\sqrt{n} C_n}{2C_A}, \frac{\sqrt{n} d_0}{2} \right\} \right].
\end{aligned}$$

The first two terms converge to zero. For the last term, the bootstrap critical value satisfies $\hat{q}_{1-\alpha} = O_{\mathbb{P}_{H_1}}(\sqrt{\log(pn)})$ under the preceding conditions. Hence the third term also converges to zero Therefore, we have

$$\mathbb{P}_{H_1}(\tilde{T}_n > \hat{q}_{1-\alpha}) \to 1$$

for every fixed alternative satisfying $E \not\subseteq E_0$. Together with the included alternative case $E \subsetneq E_0$ handled above, this proves the power guarantee for every fixed alternative $E \neq E_0$. This finishes the proof of Theorem 6.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state the proposed graph-testing problem, the direct statistic, the finite-sample Gaussian approximation, the block-bootstrap calibration, and the two-sample extension. These claims are matched by the formal methodology in Section 2, the theoretical guarantees in Section 3, the supplemental theory, and the quantitative and fMRI studies.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper states the main scope and limitations through its assumptions, including the stationary Gaussian linear process model, the long-memory condition $\beta > 3/4$, sparsity and stability requirements, and nondegenerate long-run variance conditions. The quantitative study also includes an out-of-theory ultra-long-memory regime $\beta = 0.55$, where the approximation fails as expected, illustrating the boundary of the theory.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best

judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: All theoretical results are stated with explicit conditions in Section 3 and the supplemental theory. Complete proofs for the main Gaussian approximation theorem, block-bootstrap theorem, two-sample extension, and consistency-empowered correction are provided in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Sections 4 and C describe the simulation settings, including sample sizes, dimensions, temporal-dependence parameters, graph construction, signal strengths, number of Monte Carlo replications, bootstrap block length, and evaluation metrics. The real-data analysis also specifies the graph-estimation and testing pipeline used for the Sherlock fMRI study.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.

(b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: The paper provides detailed algorithms, simulation designs, and data-processing steps sufficient to reproduce the experiments in principle, but the current submission does not include an anonymized public code repository with executable reproduction scripts. The synthetic data-generating mechanism and the real-data testing pipeline are nevertheless described in the main text and appendix.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The experimental settings are specified in Sections 4.1, 4.2, and 4, including the grids of (n, p), values of β, graph bandwidth, precision-matrix construction, planted-edge strengths, Monte Carlo replication count, nominal testing level, block length, and correction procedure for the real-data p-values. Algorithm 1 gives the implementation steps for the proposed bootstrap calibration.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.

- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The experiments report distributional approximation errors through Kolmogorov and Wasserstein-1 distances, empirical rejection rates over 200 Monte Carlo replications, and nominal testing levels. In the real-data analysis, subject-level p-values are adjusted using the Holm procedure to control family-wise error rate across the 17 subjects.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The experiments use CPUs on the Mercury cluster provided by the University of Chicago Booth School of Business, an NVIDIA GeForce RTX 4090 GPU, and NVIDIA DGX Spark (GB10).

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

Justification: The research is methodological, uses synthetic data and an existing fMRI dataset for illustration, and does not introduce generative models, scraped datasets, surveillance systems, or other high-risk artifacts.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The work has positive potential impact by providing uncertainty-aware tests for conditional-dependence graphs in long-memory scientific time series, including neuroimaging applications. Possible negative impacts are limited but include overinterpretation of exploratory connectivity findings; the paper presents the fMRI study as a statistical comparison rather than a clinical or causal claim.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper does not release pretrained generative models, language models, scraped internet datasets, or other artifacts with a high risk of misuse. The proposed method is a statistical testing procedure evaluated on synthetic data and an existing fMRI dataset.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The paper cites the methodological works and uses the Sherlock fMRI dataset as an existing data asset. No third-party model or code asset is redistributed as part of the method.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not introduce or release a new dataset, benchmark, pretrained model, or software package as a new asset. The contribution is a statistical testing method with theoretical guarantees and empirical evaluation.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing and does not collect new human-subject data. The fMRI analysis uses an existing dataset, and no new participant instructions, compensation, or interaction with participants is part of this work.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not conduct new human-subject experiments or collect new participant data. The real-data component is a secondary analysis of an existing fMRI dataset, so no new IRB approval by the authors is applicable.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core methodology does not involve LLMs as an original, important, or non-standard component. Any use of writing or formatting tools, if any, is not part of the scientific method or empirical results.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.