Supplementary Material for "High-dimensional Detection of Spatial Interference Effects"

In this Supplementary Material, we illustrate the detailed potential outcome framework in Section S1, discuss the model assumptions and robustness of our proposed method in Section S2. Discussion on the convergence and computational time of the estimation algorithm, more explanations on BiRS algorithm, and the selection of tuning parameters are included in Section S3. Theoretical analysis, including lemmas and proofs, is included in Section S4–Section S5.

S1 Potential outcome framework and interference effect

We now use the potential outcome framework to introduce the causal estimands of interest. This framework is not intended to add methodological novelty, but rather to situate our proposed method within the well-established language of causal inference. By formulating our problem in terms of potential outcomes, we clarify how the direct and interference effects we estimate correspond to standard causal estimands. This connection helps readers interpret our results in familiar causal terms, and demonstrates that our estimation and detection procedures are compatible with the classical causal framework. Let M and Y be the corresponding RC-dim vectors in the whole experimental area. When SUTVA does not hold, Y_{rc} depends on M instead of M_{rc}. Let $Y_{rc}(M)$ be the potential outcome for unit (r, c) under treatment M. As pointed out by Rubin (1986), the potential outcome is well-defined only if the following assumption holds:

Assumption S1 *(Consistency Assumption, CA) For each unit (r, c), the observed outcome satisfies $Y_{rc} = Y_{rc}(M)$.*

For any unit (r, c), let $\mathcal{N}_{rc} \subset \{(r', c') \neq (r, c) : 1 \leq r' \leq R, 1 \leq c' \leq C\}$ be an index subset dependent on (r, c), and let $\mathcal{N}_{-rc}$ denote all units not in $\{(r, c)\} \cup \mathcal{N}_{rc}$. We impose the following assumption.

Assumption S2 *(Sparse Interference Assumption, SIA) For each unit (r, c), there exists a smallest index subset $\mathcal{N}_{rc}$ with cardinal $s_{rc} \ll RC$ such that for any $M_{\mathcal{N}_{-rc}}$ and $M'_{\mathcal{N}_{-rc}}$, the potential outcomes satisfy $Y_{rc}(M_{rc}, M_{\mathcal{N}_{rc}}, M_{\mathcal{N}_{-rc}}) = Y_{rc}(M_{rc}, M_{\mathcal{N}_{rc}}, M'_{\mathcal{N}_{-rc}})$.*

The set $\mathcal{N}_{rc}$ satisfying Assumption S2 is the interference neighbor set for unit (r, c) of interest. We consider three causal estimands of interest in this work. The first two are individual-level effects, following the definitions in Forastiere et al. (2021). Specifically, we define the direct effect (DE) for unit (r, c) by

$$\mathrm{DE}_{rc} = \sum_b \mathbb{E}\left\{Y_{rc}(M_{rc} = 1, M_{\mathcal{N}_{rc}} = b) - Y_{rc}(M_{rc} = -1, M_{\mathcal{N}_{rc}} = b)\right\} P(M_{\mathcal{N}_{rc}} = b)$$

where $\mathcal{N}_{rc}$ denotes the set of potential interference neighbors of unit (r, c), and b ranges over possible treatment assignments to $\mathcal{N}_{rc}$. This quantity captures the average effect of changing the treatment status of unit (r, c), holding the treatment assignments of its neighbors fixed. The interference effect (IE) for unit (r, c), given its own treatment status $a \in \{1, -1\}$, is defined as

$$\mathrm{IE}_{rc}(a) = \sum_b \mathbb{E}\left\{Y_{rc}(M_{rc} = a, M_{\mathcal{N}_{rc}} = b) - Y_{rc}(M_{rc} = a, M_{\mathcal{N}_{rc}} = -1)\right\} P(M_{\mathcal{N}_{rc}} = b),$$

which quantifies the average effect of changing the neighbors' treatments from a baseline (all-1) to b, while keeping unit (r, c)'s own treatment fixed at a. Finally, we define the average treatment effect (ATE), which reflects the global average effect of applying treatment versus control across the entire experimental area:

$$\text{ATE} = (RC)^{-1} \sum_{r,c} \mathbb{E} \left\{ Y_{rc}(M_{rc} = 1, M_{\mathcal{N}_{rc}} = 1) - Y_{rc}(M_{rc} = -1, M_{\mathcal{N}_{rc}} = -1) \right\},$$

where $M_{\mathcal{N}_{rc}} = a$ indicates that all elements of the vector $M_{\mathcal{N}_{rc}}$ are equal to a. This estimand captures the expected difference in outcomes when all units are assigned to treatment versus all assigned to control. We finally introduce the classical conditional independence assumption.

Assumption S3 *(Conditional Independence Assumption, CIA) Let X_{rc} denote the observed covariates for unit (r, c). Then the joint treatment assignment M is assumed to be independent of all potential outcomes, conditional on the covariates $\{X_{rc}\}_{r,c}$.*

This assumption, also referred to as the unconfoundedness assumption, is fundamental for the identification of causal effects from observational data. We assume the following structural model for the potential outcomes:

$$Y_{rc}(M) = X_{rc}^{\top} \beta_{rc} + M_{rc} L_{rc} + M_{\mathcal{N}_{rc}}^{\top} S_{rc} + \varepsilon_{rc}(M), \tag{S1}$$

where $\beta_{rc} \in \mathbb{R}^d$ is the coefficient vector for covariates, $L_{rc} \in \mathbb{R}$ represents the direct effect of the treatment at unit (r, c), and $S_{rc} \in \mathbb{R}^{|\mathcal{N}_{rc}|}$ encodes the interference effects from neighboring units. Here, $M_{\mathcal{N}_{rc}}$ denotes the vector of treatment assignments for the interference neighbors of unit (r, c). The error term $\varepsilon_{rc}(M)$ has mean zero and is independent of the treatment assignment, conditional on the covariates. This model reflects the assumption that potential outcomes are linearly affected by the treatment status of each unit and its interference neighbors, consistent with Assumption S2. Under Assumption S1–S3, the observed outcome for each sample i corresponds to the potential outcome under the realized treatment assignment M_i. Therefore, with repeated observations over time indexed by

$i = 1, \ldots, n$, the potential outcome model yields the observed regression model

$$Y_{i,rc} = X_{i,rc}^{\top}\beta_{rc} + M_{i,rc}L_{rc} + M_{i,\mathcal{N}_{rc}}^{\top}S_{rc} + \varepsilon_{i,rc},$$

where $\varepsilon_{i,rc}$ are independent mean-zero errors. Then we have the following conclusions.

Lemma S1 *Under CA, SIA and CIA, we have*

$$\mathrm{DE}_{rc} = \sum_{b} E_X[E\{Y_{rc}|M_{rc} = 1, M_{\mathcal{N}_{rc}} = b, X_{rc}) - E\{Y_{rc}|M_{rc} = -1, M_{\mathcal{N}_{rc}} = b, X_{rc})]P(M_{\mathcal{N}_{rc}} = b),$$

$$\mathrm{IE}_{rc}(a) = \sum_{b} E_X[E\{Y_{rc}|M_{rc} = a, M_{\mathcal{N}_{rc}} = b, X_{rc}) - E\{Y_{rc}|M_{rc} = a, M_{\mathcal{N}_{rc}} = -1, X_{rc})]P(M_{\mathcal{N}_{rc}} = b),$$

$$\mathrm{ATE} = (RC)^{-1}\sum_{r,c}\sum_{b} E_X[E\{Y_{rc}|M_{rc} = 1, M_{\mathcal{N}_{rc}} = 1, X_{rc}) - E\{Y_{rc}|M_{rc} = -1, M_{\mathcal{N}_{rc}} = -1, X_{rc})].$$

Proof. By Assumption S1, the potential outcome can be expressed by

$$E\{Y_{rc}(M_{rc} = a, M_{N_{rc}} = b)\} = E\{Y_{rc}|M_{rc} = a, M_{N_{rc}} = b\}.$$

By Assumptions S3,

$$E\{Y_{rc}(M_{rc} = 1, M_{\mathcal{N}_{rc}} = b) - Y_{rc}(M_{rc} = -1, M_{\mathcal{N}_{rc}} = b)|X_{rc}\}$$

$$= E\{Y_{rc}|M_{rc} = 1, M_{\mathcal{N}_{rc}} = b, X_{rc}) - E\{Y_{rc}|M_{rc} = -1, M_{\mathcal{N}_{rc}} = b, X_{rc}),$$

$$E\{Y_{rc}(M_{rc} = a, M_{\mathcal{N}_{rc}} = b) - Y_{rc}(M_{rc} = a, M_{\mathcal{N}_{rc}} = -1)|X_{rc}\}$$

$$= E\{Y_{rc}|M_{rc} = a, M_{\mathcal{N}_{rc}} = b, X_{rc}) - E\{Y_{rc}|M_{rc} = a, M_{\mathcal{N}_{rc}} = -1, X_{rc}).$$

The expression for ATE can be derived similarly, and we complete the proof.

■

Under model (S1), the causal estimands DE, IE, and ATE can be expressed in terms of model parameters L_{rc} and S_{rc} provided that identification assumptions hold.

Lemma S2 *Under CA, SIA, CIA and model (1), we have*

$$\mathrm{DE}_{rc} = 2L_{rc}, \quad \mathrm{IE}_{rc}(a) = 2\sum_{b}\left(\sum_{\{j:b_j=1\}} S_{rc,j}\right)P(M_{\mathcal{N}_{rc}} = b),$$

$$\textit{and } \mathrm{ATE} = \frac{2}{RC}\sum_{r,c}\left\{L_{rc} + \sum_{j\in\mathcal{N}_{rc}} S_{rc,j}\right\}.$$

Proof. We take the derivation of DE_{rc} as an example, and $\text{IE}_{rc}(a)$, ATE can be derived similarly. Under model (1), it holds that

$$E\{Y_{rc}|M_{rc}=1, M_{\mathcal{N}_{rc}}=1, X_{rc})\} = X_{rc}\beta_{rc} + L_{rc} + \mathbf{1}_{\mathcal{N}_{rc}}^{\top} S_{rc}$$

$$E\{Y_{rc}|M_{rc}=-1, M_{\mathcal{N}_{rc}}=1, X_{rc})\} = X_{rc}\beta_{rc} - L_{rc} + \mathbf{1}_{\mathcal{N}_{rc}}^{\top} S_{rc}.$$

Then it is straightforward to obtain the above conclusions.

■

Remark. We implicitly assume sufficient variability in treatment assignments across time so that all relevant configurations of $M_{\mathcal{N}_{rc}}$ occur with positive probability (a form of positivity). This is typically satisfied in dynamic environments like ride-hailing platforms.

Remark. Under model (1), DE and IE are separate and additive, and hence $\text{IE}_{rc}(a)$ is independent of a. Therefore, we can simply denote it as IE_{rc}. The probability distribution $P(M_{\mathcal{N}_{rc}} = b)$ is determined by the treatment assignment mechanism.

Remark. In the context of the sharing economy, a key question in causal inference is policy evaluation—for instance, whether implementing a new algorithm (such as a subsidy policy or a supply-demand matching mechanism) can improve total driver income. When the treatment assignment mechanism applies the new algorithm uniformly across all units (i.e., a global treatment), we have $P(M_{\mathcal{N}_{rc}} = 1) = 1$ for all units (r, c). In this case, the following relationship holds

$$\text{ATE} = \frac{1}{RC}\sum_{r,c}(\text{DE}_{rc} + \text{IE}_{rc}).$$

We are particularly interested in detecting the elements in $\mathcal{N}_{rc}$ for each unit (r, c) from the observational data, as this can provide insights into the mechanisms underlying treatment effects. Understanding these mechanisms can inform the development of more effective strategies. Assumption S2 covers many existing assumptions commonly used in the causal inference literature to address spatial interference, including the partial IE and local network interference (Sobel 2006, Liu et al. 2016, Sävje et al. 2021). However, Assumption S2 allows for variation in the interference structure across space. In scenarios where prior knowledge suggests high-dimensional interference, one may refer to Leung (2022) which assumes IE decays as the distance between two units increases. This assumption differs from our sparse interference assumption. Both assumptions have their respective advantages. While the decaying interference assumption potentially accommodates high-dimensional IE, our sparse assumption allows for distant interactions.

S2 More on the Model Assumptions

S2.1 Model justification

Two important assumptions of our model are the additivity of DE and IE, and the spatial independence of the noises. In this subsection, we verify these two assumptions in the case study.

For the additivity of DE and IE, we augment the model to include an interaction term between local and neighboring treatments and test its statistical significance. To do this, we apply the multi-split method proposed in Meinshausen et al. (2009), which is designed for high-dimensional linear models. The method aggregates p-values from inferences across multiple random splits to ensure the control of family-wise error rate. For each group of samples, an ANOVA based on the lasso estimators is conducted to obtain the corresponding p-value. The minimum p-value across the $R \times C$ locations is $p_{\min} = 0.677$, indicating no strong evidence of interaction effects. This result supports the validity of the additive specification in our application.

For the spatial independence of the noises, we applied the high-dimensional correlation test proposed by Cai & Jiang (2011) to the residuals in our case study. Let $q = R \times C$, and let $\Theta \in \mathbb{R}^{q \times q}$ be the correlation matrix of the residual vector $\varepsilon = \varepsilon_{rc}$. The null hypothesis $H_0 : \Theta = I_q$ is tested using the statistic

$$T_n = \max_{1 \leq j < k \leq q} \hat{\Theta}_{jk}^2,$$

with the asymptotic distribution given by

$$\mathbb{P}\left(nT_n - 4 \log q + \log \log q \leq x\right) \to e^{-(1/\sqrt{8\pi})e^{-x/2}}.$$

We reject H_0 if

$$T_n \geq n^{-1}\left\{4 \log q - \log \log q - \log(8\pi) - 2 \log \log(1-\alpha)^{-1}\right\},$$

where $\alpha = 0.05$. The result fails to reject H_0, indicating that spatial residual correlations in the data are negligible.

We also acknowledge that in domains where spatial dependence arises from unobserved or poorly measured factors, additional modeling (e.g., spatial error models or random effects) may be necessary. The main challenge lies in deriving an upper bound for the operator norm of element-correlated matrices in the low-rank estimation. One potential solution is to assume that the spatial dependence is local. Extending our framework to explicitly account for such spatial correlations is an important direction for future work. We have added these points in the discussion section to clarify the feasibility and limitations of this assumption. To further explore this issue empirically, we conducted a simulation study to evaluate the robustness of our method under spatially correlated errors. Specifically, we vectorized the spatial error terms ε_{rc} and imposed a Toeplitz covariance structure $\Sigma_{jk}(\rho) = \rho^{|j-k|}$ for $j, k = 1, \ldots, RC$, with correlation levels $\rho \in 0.5, 0.7, 0.9$. For reference, $\rho = 0$ corresponds to the original setting with spatially independent errors. As shown in Figure S1, the method maintains stable performance even under strong spatial correlation ($\rho = 0.9$), suggesting practical robustness.

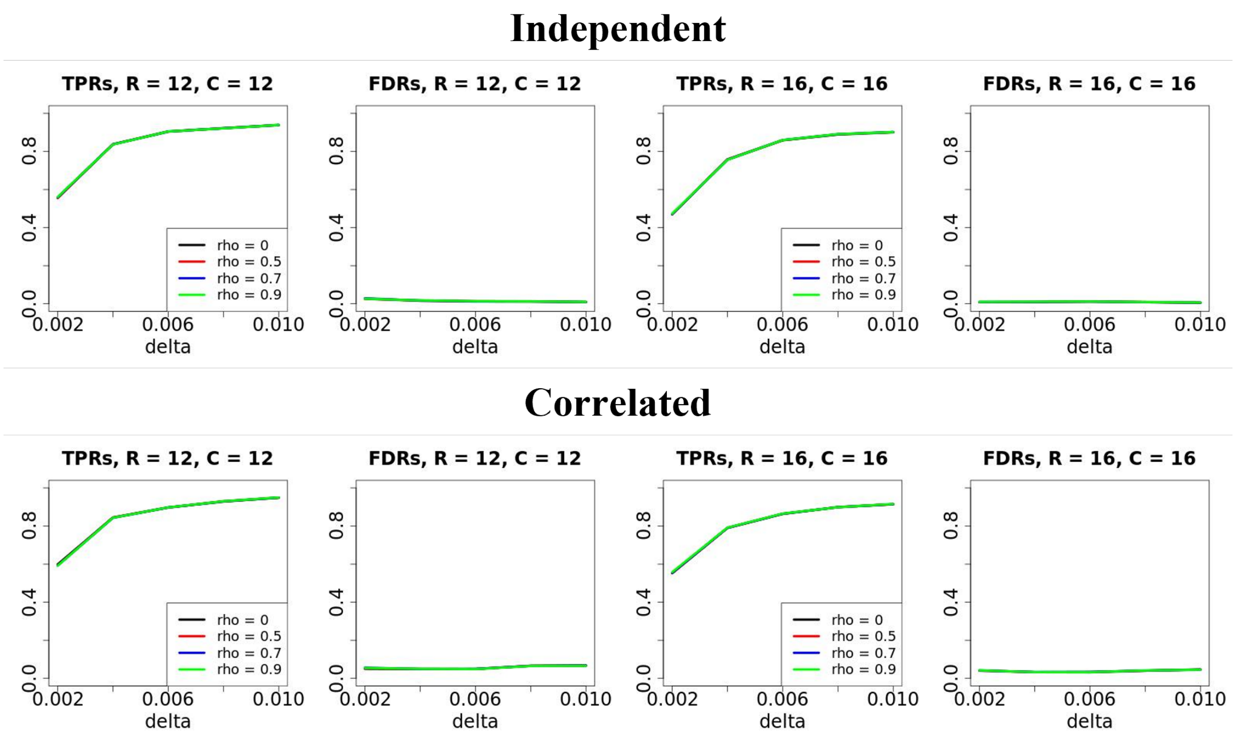

Figure S1: Performance of the detection method under varying levels of spatial error correlation.

S2.2 Robustness to model mis-specification

Though our model rely on the low-rank of DE matrix and temporal independence of noises to establish theoretical guarantees, the proposed method perform well when these assumptions are violated.

First, to examine the robustness of our method beyond the exact low-rank setting, we considered a scenario where the direct effect matrix L is full-rank. Specifically, we construct $L_{rc} = 0.01 \cdot F_{rc}$ where $F_{rc} = \sum_{i=1}^{n} X_{irc}\beta_{rc}/n$ represents the expected outcome of unit (r, c). Under this construction, the matrix $L = (L_{rc})$ no longer satisfies the low-rank assumption. Figure S2 compares the true positive rates (TPRs) and false discovery rates (FDRs) between the original low-rank setting and this full-rank alternative. The results show only minor differences in performance, suggesting that our method remains effective even when the low-rank assumption is violated.

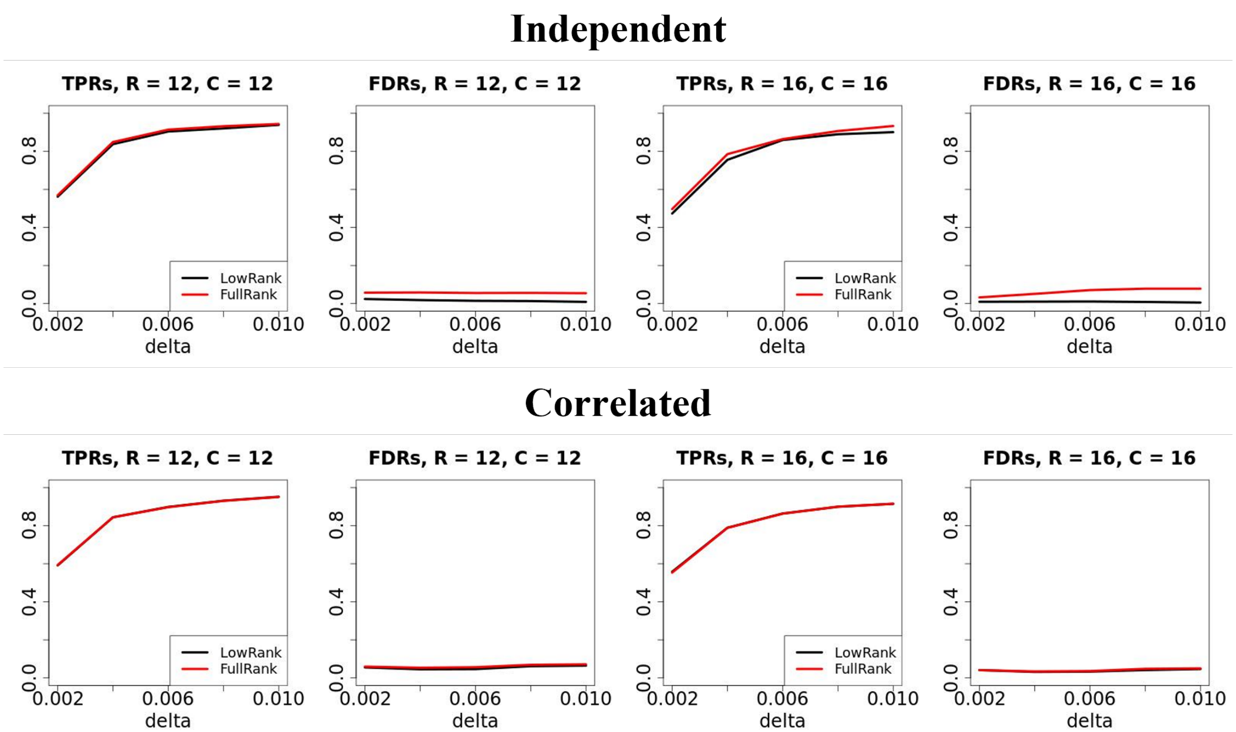

Figure S2: Detection results under a full-rank direct effect matrix.

Second, we evaluated robustness to temporal dependence in the error terms. Specifically, for each spatial location (r, c), we introduced temporal correlation across n repeated samples by assuming that the error vector ε_{irc}, $i = 1, \ldots, n$ follows a stationary Gaussian process with autocorrelation structure

$$\Sigma(\rho) = \left\{\rho^{|j-k|}\right\}_{j,k=1}^{n},$$

where $\rho \in \{0.5, 0.7, 0.9\}$. The case $\rho = 0$corresponds to the baseline scenario with temporally independent errors. As shown in Figure S3, the TPRs and FDRs under temporal correlation remain very similar to the independent case, even when the correlation is strong ($\rho = 0.9$). This indicates that the proposed method is robust to moderate-to-strong temporal dependence in the noise.

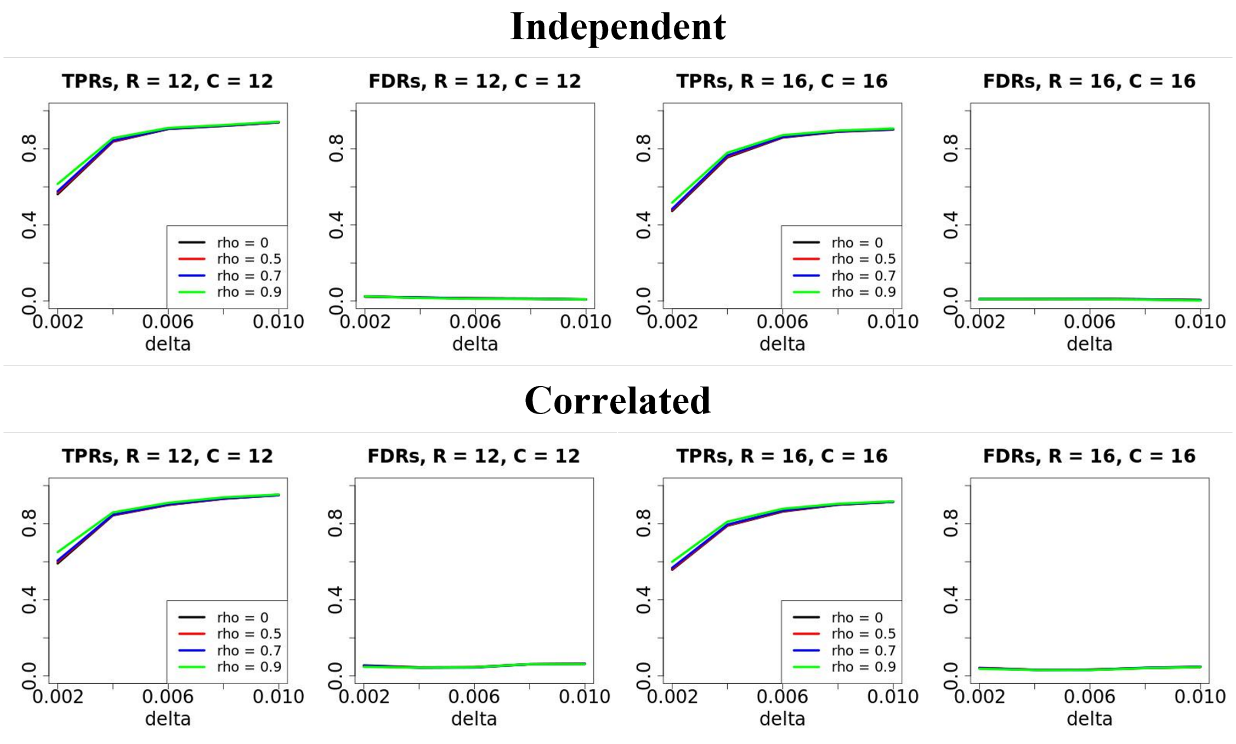

Figure S3: Detection results under temporally correlated error terms.

S3 More on Algorithms and Tuning Parameters

This section provides additional discussion and empirical evidence regarding the numerical performance of Algorithm 1, which performs joint estimation of the direct and interference effects via an iterative procedure. More explanations for the BiRS detection algorithm, and recommendations for the selection of tuning parameters

S3.1 Computation and convergence of Algorithm 1

Convergence Behavior: Algorithm 1 alternates between updating the direct effect matrix L using a low-rank regularized regression, and the interference structure S via Lasso-based sparse regression. While theoretical convergence guarantees for the full alternating optimization procedure under our model setting remain analytically challenging, we evaluate the convergence empirically across various grid sizes.

Specifically, we record the number of outer iterations until convergence under the correlated treatment design (with strength parameter $\delta = 0.1$) for different grid sizes ranging from 8×8 to 18×18. The results in Table S1 show that the algorithm typically converges within 6 to 7 iterations, and the number of iterations does not grow significantly with grid size. This supports the practical stability and efficiency of the proposed method.

Computational Cost: We also report the average computation time (in minutes) for the estimation step across the same grid scales. As expected, the cost increases with the number of grid points, since each iteration involves solving $R \times C$ Lasso problems and matrix updates of size proportional to RC. However, the computation time remains manageable across all tested settings (e.g., less than 0.3 minutes for a 324-cell grid), making the method feasible for moderately large spatial applications.

Table S1: The empirical average computing time (mins) for estimation, testing, and detection procedures, as well as the number of iterations of the proposed method in the correlated treatment scenario with the strength parameter $\delta = 0.1$. The scale of $R \times C$ ranges from 8×8 to 18×18.

(R, C)	$(8, 8)$	$(10, 10)$	$(12, 12)$	$(14, 14)$	$(16, 16)$	$(18, 18)$
Iterations	6.050	6.000	6.000	6.950	6.400	6.050
Estimation	0.025	0.037	0.059	0.111	0.160	0.225

S3.2 Estimation of σ_{rc}

The estimation of σ_{rc} follows the method proposed in Sun & Zhang (2012) and can be implemented by the R package `scalreg`. Specifically, let $Z_{rc} = (X_{rc}, M_{rc}, M_{-rc})$ and $\gamma_{rc} = \left(\beta_{rc}^\top, L_{rc}, S_{rc}^\top\right)^\top$, then σ_{rc} can be estimated by the following iterative algorithm:

$$\hat{\sigma}_{rc} \leftarrow n^{-1/2} \left\|Y_{rc} - Z_{rc}\hat{\gamma}_{rc}\right\|_F ,$$

$$\lambda_\sigma \leftarrow 2\hat{\sigma}_{rc}\lambda_\sigma,$$

$$\hat{\gamma}_{rc} \leftarrow \hat{\gamma}_{rc}^{\text{new}} \;\text{ if } L_{\lambda_\sigma, rc}\left(\hat{\gamma}_{rc}^{\text{new}}\right) \leq L_{\lambda_\sigma}\left(\hat{\gamma}_{rc}\right),$$

where

$$L_{\lambda_\sigma, rc}(\gamma_{rc}) = \frac{1}{n}\left\|Y_{rc} - Z_{rc}\gamma_{rc}\right\|_F^2 + \lambda_\sigma \left\|\gamma_{rc}\right\|_1 ,$$

λ_σ is a constant and termination criterion is the convergence of $\hat{\sigma}_{rc}$. In our problem, following the suggestion of Sun & Zhang (2013), we set the universal penalty level $\lambda_\sigma = \sqrt{2}\tilde{L}_n(k_0/RC)$ with $\tilde{L}_n(t) = n^{-1/2}\Phi^{-1}(1-t)$, where Φ is the cumulative distribution function for $\mathcal{N}(0, 1)$, and k_0 is the solution to $k = \tilde{L}_1^4(k/p) + 2\tilde{L}_1(k/p)$.

S3.3 Illustration of BiRS algorithm

For easy presentation, we give a flow chart in Figure S4 for clustered signals, where the signals reside in several separated signal regions, to illustrate how the BiRS algorithm works.

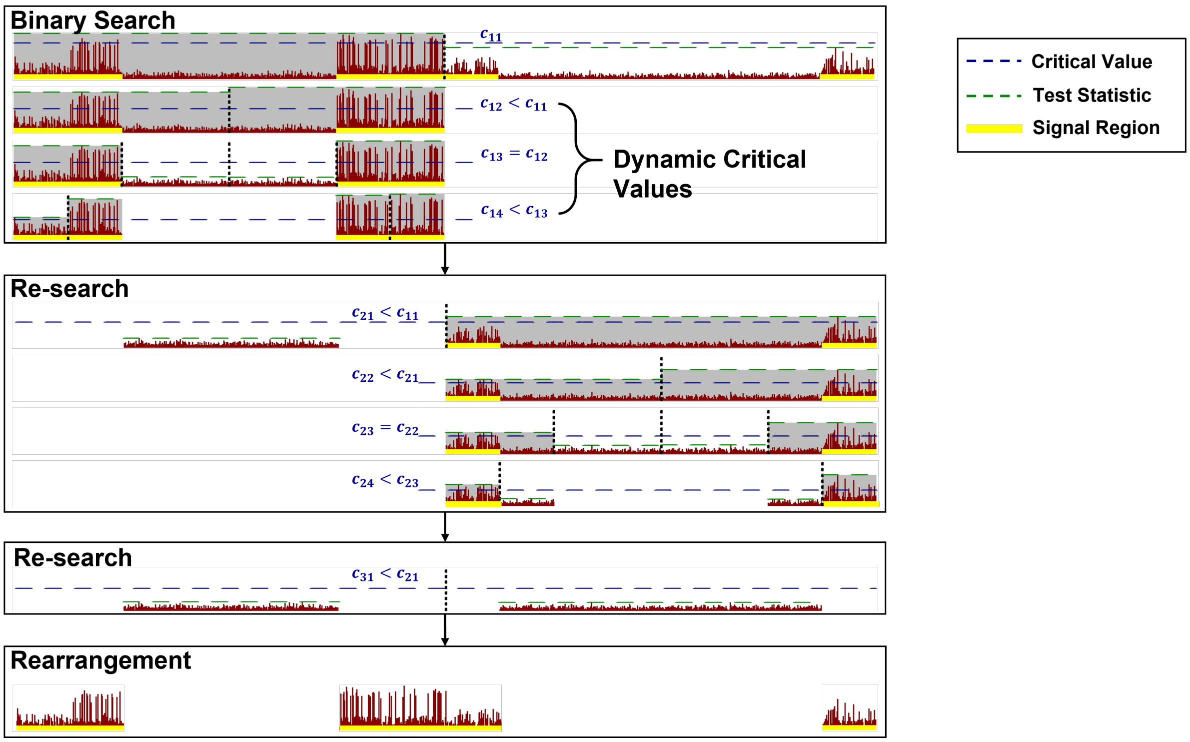

Figure S4: BiRS detection scheme. The red lines are the normalized estimation $\sqrt{n}\text{Vec}(\hat{S})$.

From this scheme, we can see that the BiRS algorithm identifies signals through an efficient binary search process that iteratively splits regions until a minimum size is reached (in our problem, the minimum size is set to be 1), enabling rapid detection without prior knowledge of segment boundaries. A key component is the use of dynamically adjusted critical values, which ensures that BiRS can achieve higher detection power while effectively controlling the family-wise error rate. After the binary search process, a re-search procedure is applied to remove detected significant signals and initiates a new search among the remaining regions. This iterative process continues until no further significant signals are found, after which selected segments are merged to adapt to the true signal structure.

S3.4 Tuning parameter selection

Finally, we further clarify the parameter selection involved in our method. Our proposed method involves two main types of tuning parameters:

1. Low-rank penalty parameter λ.

This parameter controls the nuclear norm penalty $\|L\|_*$, which encourages a low-rank structure in the direct effect matrix. In practice, we select λ using standard 5-fold cross-validation, a widely adopted approach that balances model fit and regularization.

2. Sparse penalty parameter λ_{rc}.

These control the ℓ_1-penalization in the estimation of interference effects $\mathbf{S}_{rc}$. Theoretically, we follow the guidance in high-dimensional literature and set

$$\lambda_{rc} \geq A\sigma_{rc}\sqrt{\log(RC)/n},$$

where $A > 2\sqrt{2}$ is a universal constant and σ_{rc} is the noise level at unit (r, c). In our implementation, we take $A = 2\sqrt{2.1}$, and estimate σ_{rc} using the scaled lasso method (Sun & Zhang 2012), as detailed in our response to Comment 7.

To assess the sensitivity of our method to the choice of A, we conducted a robustness study with $A \in \{2.1, 2.3, 2.5, 2.7\}$. The results, shown in Figure S5, indicate that our method remains stable across different values, with only minor variation in detection performance. This supports the robustness of our approach with respect to tuning parameters.

3. Other parameter settings.

The step size (or updating rate) in our algorithm is set to $\eta = 2$, following the recommendation of Beck & Teboulle (2009). The convergence tolerance is $\tau = 10^{-4}$, based on the

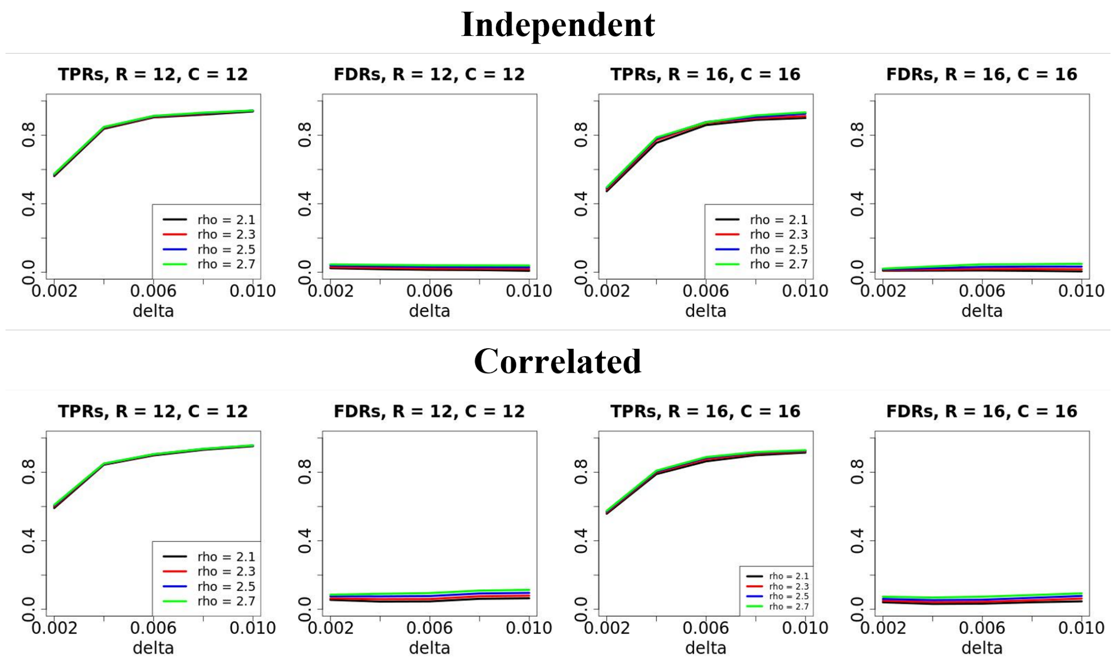

Figure S5: The detection results for different tuning parameters.

relative change of updates across iterations:

$$\delta^{(t+1)} = \frac{\|\beta^{(t+1)} - \beta^{(t)}\|_F}{\|\beta^{(t)}\|_F} + \frac{\|L^{(t+1)} - L^{(t)}\|_F}{\|L^{(t)}\|_F} + \frac{\|S^{(t+1)} - S^{(t)}\|_F}{\|S^{(t)}\|_F}.$$

For the global testing and detection procedures, we use $B = 1000$ bootstrap replications to ensure accuracy and stability of the inference.

S4 Key Lemmas

Lemma S3 *Let $(Z_1, \ldots, Z_p)'$ be a zero-mean multivariate normal random vector with covariance matrix $\Omega = (\omega_{jk})_{1\le j,k\le p}$ and diagonals $\omega_{j,j} = 1$ for $1 \le j \le p$. Suppose that $\max_{1\le j\ne k\le p} |\omega_{jk}| \le r < 1$ and $\max_j \sum_{k=1}^p \omega_{jk} \le C_0$. Then for any $x \in \mathbb{R}$, as $p \to \infty$,*

$$\mathbb{P}\left\{\max_{1\le j\le p} |Z_j| \le \sqrt{2\log p - \log\log p + x}\right\} \to \exp\left\{-\frac{1}{\sqrt{\pi}} \exp\left(-\frac{x}{2}\right)\right\}.$$

Lemma S4 *Assume the estimator $\hat{L}$ of L^* satisfies that*

$$\frac{K}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\|M_{rc}\Delta(L_{rc})\|_2^2 \leq \frac{2}{n}\left\langle \chi(M,\mathcal{P},\varepsilon),\Delta(L)\right\rangle + \lambda\left(\|L^*\|_* - \left\|\hat{L}\right\|_*\right) + \delta, \tag{S2}$$

where $\Delta(L) = \hat{L} - L^*$. Let $U^r \in \mathbb{R}^{R\times r}$ and $V^r \in \mathbb{R}^{C\times r}$ be matrices consisting of the top r left and right singular value vectors of L^*. Then there exists a matrix decomposition $\Delta(L) = \Delta'(L) + \Delta''(L)$ such that:

(a) the matrix $\Delta'(L)$ satisfies the constraint rank$(\Delta'(L)) \leq 2r$;

(b) if $\lambda > 2\|\chi(M,\mathcal{P},\varepsilon)\|_{\mathrm{op}}/n$, then the nuclear norm of $\Delta''(L)$ is bounded as

$$\|\Delta''(L)\|_* \leq 3\|\Delta'(L)\|_* + 4\sum_{j=r+1}^{\min(R,C)}\sigma_j(L^*) + \delta.$$

Lemma S5 *Let the random vector $U_n \in \mathbb{R}^p$ be the test statistic and $U_n^{(1)},\ldots,U_n^{(N)} \in \mathbb{R}^p$ be generated by the bootstrap procedure in the previous test. Denote J_0 to be the null set and J_1 to be the signal set. Suppose that for any index set $\Lambda_0 \subset J_0$, the random vector $U_n(\Lambda_0)$ satisfies that*

$$\mathbb{P}\left\{\|U_n(\Lambda_0)\|_\infty > c_B(\alpha,\Lambda_0)\right\} \lesssim \alpha,$$

where $c_B(\alpha,\Lambda_0)$ is the empirical threshold generated by $U_n^{(1)}(\Lambda_0),\ldots,U_n^{(N)}(\Lambda_0)$.

We further assume that U_n satisfies that there exists some $\beta > \alpha$ such that for some $H \to \infty$,

$$D(H) = \sup_{i_1,\ldots,i_H}\frac{\mathbb{P}\left(\cap_{j=1}^H A_{i_j}\right)}{\Pi_{j=1}^H \mathbb{P}\left(A_{i_j}\right)} \leq \frac{1}{(2\beta)^H},$$

where $A_i = \{|U_{n,i}| \geq t_i\}$ and t_i is the $1-\alpha$ quantile of $|U_{n,i}|$.

Let $\hat{J}_1^{BiRS}$ be the estimated signal set by BiRS algorithm, then there exists a constants c such that

$$\mathbb{P}\left\{\mathcal{J}\left(J_1,\hat{J}_1^{BiRS}\right) \geq 1-\frac{\eta}{s}\right\} \geq \left\{1-\frac{8\log(p+1)}{n^2}\right\}\left\{1-c\left(\frac{\alpha}{\beta}\right)^\eta\right\}, \tag{S3}$$

where $s = |J_1|$ and $\eta = \left|\hat{J}_1^{BiRS}\right|$.

Lemma S3 is Lemma 6 in Cai et al. (2014). One can refer to this paper for detailed proof.

S5 Theoretical Proofs

S5.1 Proof of Lemma S4

We write the singular value decomposition (SVD) as $L^* = UDV^\top$, where $U \in \mathbb{R}^{R\times R}$ and $V \in \mathbb{R}^{C\times C}$. Note that the matrices U^r and V^r are given by the first r columns of U and V respectively. We then define the matrix $\Gamma = U^\top \Delta(L) V$, and write it in block form as

$$\Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix},$$

where $\Gamma_{11} \in \mathbb{R}^r$ and $\Gamma_{22} \in \mathbb{R}^{(R-r)\times(C-r)}$. We now define the matrices

$$\Delta''(L) = U \begin{bmatrix} 0 & 0 \\ 0 & \Gamma_{22} \end{bmatrix} V^\top,$$

and $\Delta'(L) = \Delta(L) - \Delta''(L)$. For easy notation, we denote $\Delta = \Delta(L)$ and the same for $\Delta'(L), \Delta''(L)$. Note that we have

$$\operatorname{rank}(\Delta') = \operatorname{rank} \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & 0 \end{bmatrix} \le \operatorname{rank} \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ 0 & 0 \end{bmatrix} + \operatorname{rank} \begin{bmatrix} \Gamma_{11} & 0 \\ \Gamma_{21} & 0 \end{bmatrix} \le 2r,$$

which establishes Lemma S4(a). Moreover, we note for future reference that by construction of Δ'', the nuclear norm satisfies the decomposition

$$\|\Pi_{\mathcal{A}^r}(L^*) + \Delta''\|_* = \|\Pi_{\mathcal{A}^r}(L^*)\|_* + \|\Delta\|_*,$$

where $L* = \Pi_{\mathcal{A}^r}(L^*) + \Pi_{\mathcal{B}^r}(L^*)$ and $\Pi_{\mathcal{A}^r}(L^*) = {U^r}^\top \Gamma_{11} V^r$.

We now turn to the proof of Lemma S4(b). According to equation S2, the error Δ satisfies that

$$0 \leq \frac{2}{n} \langle \chi(M, \mathcal{P}, \varepsilon), \Delta(L) \rangle + \lambda \left(\|L^*\|_* - \left\| \hat{L} \right\|_* \right) + \delta, \tag{S4}$$

using the decomposition $L* = \Pi_{\mathcal{A}^r}(L^*) + \Pi_{\mathcal{B}^r}(L^*)$, we have

$$\begin{aligned} \left\| \hat{L} \right\|_* &= \|(\Pi_{\mathcal{A}^r}(L^*) + \Delta'') + (\Pi_{\mathcal{B}^r}(L^*) + \Delta')\|_* \\ &\geq \|(\Pi_{\mathcal{A}^r}(L^*) + \Delta'')\|_* - \|(\Pi_{\mathcal{B}^r}(L^*) + \Delta')\|_* \\ &\geq \|(\Pi_{\mathcal{A}^r}(L^*)\|_* + \|\Delta''\|_* - \|(\Pi_{\mathcal{B}^r}(L^*)\|_* - \|\Delta'\|_* . \end{aligned}$$

Consequently, we have

$$\|L^*\|_* - \left\| \hat{L} \right\|_* \leq 2 \|\Pi_{\mathcal{B}^r}(L^*)\|_* + \|\Delta'\|_* - \|\Delta''\|_* .$$

Substituting this inequality into the bound S4, we obtain,

$$0 \leq \frac{2}{n} \|\langle \chi(M, \mathcal{P}, \varepsilon), \Delta(L) \rangle\|_{\text{op}} \|\Delta\|_* + \lambda \{2 \|\Pi_{\mathcal{B}^r}(L^*)\|_* + \|\Delta'\|_* - \|\Delta''\|_*\} + \delta.$$

Finally, since $\frac{2}{n} \|\langle \chi(M, \mathcal{P}, \varepsilon), \Delta(L) \rangle\|_{\text{op}} \leq \lambda$ by assumption, we conclude that

$$0 \leq \lambda \left\{ 2 \|\Pi_{\mathcal{B}^r}(L^*)\|_* + \frac{3}{2} \|\Delta'\|_* - \frac{1}{2} \|\Delta''\|_* \right\} + \delta.$$

Since $\|\Pi_{\mathcal{B}^r}\|_* = \sum_{j=r+1}^{\min(R,C)} \sigma_j(L^*)$, we finish our proof. ■

S5.2 Proof of Lemma S5

In order to obtain the theoretical true positive rate, we need some algorithm based inequalities. We first look into the case that there exists two signal points $j_1.j_2$, let

$$A_1 = \{\text{The binary search step detected } j_1\},$$

and

$$A_2 | A_1 = \{\text{The first re-search detected } j_2\}$$

, according to the power analysis Theorem 3, we have that $\mathbb{P}(A_2|A_1) \geq 1 - 8/n^2(p-1)$ (a dimension reduction accurred after the first binary search). Then under the case that there exist only two signal points, we have

$$\mathbb{P}(A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1) \geq P_1 \left\{1 - \frac{8}{n^2 p}\right\},$$

with $\mathbb{P}(A_1) = 1 - \frac{8}{n^2 p}$. Then by induction, in general, when there exists s signal points, let $A = \{$all the signal points are detected$\}$, we have

$$\mathbb{P}(A) \geq \Pi_{r=1}^{s} \left\{1 - \frac{8}{n^2(p-r-1)}\right\} \geq 1 - \frac{8}{n^2} \sum_{i=1}^{p} \frac{1}{i} \geq 1 - \frac{8\log(p+1)}{n^2}.$$

Then we give the theoretical true positive rate tends to 1 with high probability and we prove the theoretical Jaccard index tends to 1 with high probability. The previous statements say that $\left|J_1 \cap \hat{J}_1\right| \to s$. Then the Jaccard index only depends on how many non-signal points are detected by the BiRS algorithm. Let

$$A_\zeta = \{\exists \zeta \text{non-signal points be detected under } s \text{ true signal points are removed}\}.$$

Since the consistent of detecting certain signal point has been proved. We just need to specify the value of $\mathbb{P}_{A_\zeta}$.

Assume that the BiRS detected ζ points after s true signal points be removed through M whole binary search, $1 \leq M \leq \zeta$, the corresponding probability is P_M. Let $K = p - s$ and

$$B(\zeta_i) = \left\{\exists \zeta_i \text{ points been detected among the remaining } K - \sum_{j=1}^{i-1} \zeta_j \text{ points}\right\}.$$

Denoting $\Lambda_i = \{$ the index of points remaining after i whole binary search under the true signal regions removed $\}$, $i = 1, \ldots, M$. Let

$$C_{ik} = \{\text{the points } k \text{ is selected as signal point at step } i\},$$

$i = 0, \ldots, M-1; k = 1, \ldots, K$. Then we have,

$$\begin{aligned} P_M &\leq \sum_{\zeta_1+\cdots+\zeta_M=\zeta} \prod_{i=1}^{M} \mathbb{P}\left[B(\zeta_i)\right] \\ &\leq \sum_{\zeta_1+\cdots+\zeta_M=\zeta} \prod_{i=1}^{M} \mathbb{P}\left\{\cup_{\lambda_{k1},\ldots,\lambda_{k\zeta_i}\in\Lambda_{i-1}} \left[C_{\lambda_{k1}i} \cap \cdots \cap C_{\lambda_{k\zeta_i}i}\right]\right\} \\ &\leq \sum_{\zeta_1+\cdots+\zeta_M=\zeta} \prod_{i=1}^{M} \left\{\sum_{\lambda_{k1},\ldots,\lambda_{k\zeta_i}\in\Lambda_{i-1}} \mathbb{P}\left[C_{\lambda_{k1}i} \cap \cdots \cap C_{\lambda_{k\zeta_i}i}\right]\right\} \end{aligned} \tag{S5}$$

For the probability of C_{ik}, by the first condition and Bonferroni correction, we have that $\mathbb{P}\left(C_{ik}\right) \leq \alpha / \left(K - \sum_{j=0}^{i-1} \zeta_j\right)$, continue with (S5), we have,

$$\begin{aligned} P_M &\leq \sum_{\zeta_1+\cdots+\zeta_M=\zeta} \binom{K}{\zeta_1} \cdots \binom{K - \sum_{j=1}^{M-1}\zeta_j}{\zeta_M} \prod_{i=1}^{M} \left(\frac{\alpha}{K - \sum_{j=1}^{i-1}\zeta_j}\right)^{\zeta_i} D\left(\zeta_i\right) \\ &\leq \sum_{\zeta_1+\cdots+\zeta_M=\zeta} \binom{K}{\zeta_1} \cdots \binom{K - \sum_{j=1}^{M-1}\zeta_j}{\zeta_M} \frac{(K-\zeta)!}{K!} \alpha^{\zeta} \prod_{i=1}^{M} D\left(\zeta_i\right). \end{aligned}$$

By the second condition, there exists constants H such that $\prod_{i=1}^{M} D(\zeta_i) = H \cdot (1/2\beta)^{\zeta_i}$ and recall that,

$$\begin{aligned} \sum_{\zeta_1+\cdots+\zeta_M=\zeta} \binom{K}{\zeta_1} \cdots \binom{K - \sum_{j=1}^{N-1}\zeta_j}{\zeta_M} &= \frac{K!}{(K-\zeta)!} \sum_{\zeta_1+\cdots+\zeta_M=\zeta} \frac{1}{\zeta_1! \cdots \zeta_M!} \\ &\leq \frac{K! \binom{\zeta-1}{M-1}}{(K-\zeta)!}. \end{aligned}$$

Then we have $P_M \leq H\binom{\zeta-1}{M-1}(\alpha/2\beta)^{\zeta}$, thus, we have that,

$$\mathbb{P}_{A_\zeta} \leq \sum_{i=\zeta}^{K} \sum_{j=1}^{i} H \binom{i-1}{j-1} \left(\frac{\alpha}{2\beta}\right)^{i} \leq H \sum_{i=\zeta}^{K} \left(\frac{\alpha}{\beta}\right)^{i} \leq H \cdot \left(\frac{\alpha}{\beta}\right)^{\zeta}.$$

Then we finish the proof of Lemma S5. ■

S5.3 Proof of Theorem 1

Assume that $(\hat{\beta}_{rc}^{\top}, \hat{L}_{rc}, \hat{S}_{rc}^{\top})_{r,c}^{\top}$ is the minimizer of

$$\frac{1}{n} \sum_{r=1}^{R} \sum_{c=1}^{C} \left\|Y_{rc} - X_{rc}\beta_{rc} - M_{rc}L_{rc} - M_{-rc}S_{rc}\right\|_2^2 + \lambda \left\|L\right\|_* + \sum_{r=1}^{R} \sum_{c=1}^{C} \lambda_{rc} \left\|S_{rc}\right\|_1,$$

and that the true model is

$$Y_{rc} = X_{rc}\beta^*_{rc} + M_{rc}L^*_{rc} + M_{-rc}S^*_{rc} + \varepsilon_{rc}.$$

By the profiling procedure, $\hat{\beta}_{rc}$ is the minimizer of $\left\|Y_{rc} - X_{rc}\beta_{rc} - M_{rc}\hat{L}_{rc} - M_{-rc}\hat{S}_{rc}\right\|$, then

$$\hat{\beta}_{rc} = \left(X_{rc}^\top X_{rc}\right)^{-1} X_{rc}^\top \left(X_{rc}\beta^*_{rc} - M_{rc}\Delta(L_{rc}) - M_{-rc}\Delta(S_{rc}) + \varepsilon_{rc}\right),$$

where $\Delta(\beta_{rc}) = \hat{\beta}_{rc} - \beta^*_{rc}$, $\Delta(L_{rc}) = \hat{L}_{rc} - L^*_{rc}$ and $\Delta(S_{rc}) = \hat{S}_{rc} - S^*_{rc}$. Thus,

$$X_{rc}\Delta(\beta_{rc}) = -\mathcal{P}_{X_{rc}}M_{rc}\Delta(L_{rc}) - \mathcal{P}_{X_{rc}}M_{-rc}\Delta(S_{rc}) + \mathcal{P}_{X_{rc}}\varepsilon_{rc}, \tag{S6}$$

where $\mathcal{P}_{X_{rc}} = X_{rc}\left(X_{rc}^\top X_{rc}\right)^{-1} X_{rc}^\top$ is the projection matrix.

Next, we note that $(\hat{S}, \hat{L}^\top)^\top$ is the minimizer of

$$\frac{1}{n}\sum_{r=1}^{R}\sum_{c}^{C}\left\|Y_{rc} - X_{rc}\hat{\beta}_{rc} - M_{rc}L_{rc} - M_{-rc}S_{rc}\right\|_2^2 + \lambda\|L\|_* + \sum_{r=1}^{R}\sum_{c=1}^{C}\lambda_{rc}\|S_{rc}\|_1.$$

Let $\bar{\varepsilon}_{rc} = \varepsilon_{rc} - X_{rc}\Delta(\beta_{rc})$, then we have

$$\begin{aligned}
&\frac{1}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\|M_{rc}\Delta(L_{rc}) + M_{-rc}\Delta(S_{rc})\|_2^2 \\
&-\frac{2}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\{M_{rc}\Delta(L_{rc}) + M_{-rc}\Delta(S_{rc})\}^\top \mathcal{P}_{X_{rc}}\{M_{rc}\Delta(L_{rc}) + M_{-rc}\Delta(S_{rc})\} \\
&\le \frac{2}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\varepsilon_{rc}^\top(I - \mathcal{P}_{X_{rc}})M_{rc}\Delta(L_{rc}) + \lambda(\|L^*\|_* - \left\|\hat{L}\right\|_*) \\
&+\sum_{r=1}^{R}\sum_{c=1}^{C}\left\{\frac{2}{n}\varepsilon_{rc}^\top(I - \mathcal{P}_{X_{rc}})M_{-rc}\Delta(S_{rc}) - \lambda_{rc}\|\Delta(S_{rc})\|_1\right\} \\
&\le \frac{2}{n}\langle\chi(M,\mathcal{P},\varepsilon),\Delta(L)\rangle + \lambda(\|L^*\|_* - \left\|\hat{L}\right\|_*) \\
&+\sum_{r=1}^{R}\sum_{c=1}^{C}\left\{\frac{2}{n}\varepsilon_{rc}^\top(I - \mathcal{P}_{X_{rc}})M_{-rc}\Delta(S_{rc}) - \lambda_{rc}\|\Delta(S_{rc})\|_1\right\},
\end{aligned}$$

where $\chi(M,\mathcal{P},\epsilon) = (\varepsilon_{rc}^\top(I - \mathcal{P}_{X_{rc}})M_{rc})_{r,c}$. Next, we consider the last term in the last line of the above inequality. Let

$$\mathcal{C} = \left\{(r,c) : \lambda_{rc} \le \frac{2}{n}\left\|\varepsilon_{rc}^\top(I - \mathcal{P}_{X_{rc}})M_{-rc}\right\|_\infty\right\}.$$

From Lemma S1, it holds that

$$\mathbb{P}\left\{ \frac{\left\| \varepsilon_{rc}^{\top}(I - \mathcal{P}_{X_{rc}})M_{-rc} \right\|_{\infty}}{\sqrt{n}\sigma_{rc}} \leq \sqrt{2\log RC - \log\log RC + x} \right\} \to \exp\left\{ -\frac{1}{\sqrt{\pi}} \exp\left(-\frac{x}{2} \right) \right\}.$$

Take $\lambda_{rc} = A\sigma_{rc}\sqrt{\log RC/n}$ with $A > 2\sqrt{2}$. Then with probability tending to 1, $|\mathcal{C}| \leq K_1 RC^{2-A^2/8}/\sqrt{\pi \log RC}$. Similarly, since $\hat{S}_{rc}$ is the minimizer of

$$\frac{1}{n}\left\| Y_{rc} - X_{rc}\hat{\beta}_{rc} - M_{rc}\hat{L}_{rc} - M_{-rc}S_{rc} \right\|_2^2 + \lambda_{rc}\left\| S_{rc} \right\|_1.$$

We have

$$\begin{aligned}
\frac{1}{n}\left\| M_{-rc}\Delta(S_{rc}) \right\|_2^2 \leq& \frac{2}{n}\tilde{\varepsilon}_{rc}^{\top} M_{-rc}\Delta(S_{rc}) - \lambda \left\| \Delta(S_{rc}) \right\|_1 \\
\leq& \frac{2}{n}\left\{ -(I - \mathcal{P}_{X_{rc}})M_{rc}\Delta(L_{rc}) + \mathcal{P}_{X_{rc}}M_{-rc}\Delta(S_{rc}) \right\}^{\top} M_{-rc}\Delta(S_{rc}) \\
&+ \left\{ \left\| \frac{2}{n}\varepsilon_{rc}^{\top}\left(I - \mathcal{P}_{X_{rc}} \right) M_{-rc} \right\|_{\infty} - \lambda_{rc} \right\} \left\| \Delta(S_{rc}) \right\|_1.
\end{aligned}$$

Then outside the set $\mathcal{C}$, by Assumption 2, there exists a constant K_2 such that

$$\left\| M_{-rc}\Delta(S_{rc}) \right\|_2 \leq K_2 \left\| M_{rc}\Delta(L_{rc}) \right\|,$$

and inside the set $\mathcal{C}$, it has the classical lasso error bound. Recalling the common error bound for lasso estimator, we have that

$$\begin{aligned}
\frac{K_3}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\left\| M_{rc}\Delta(L_{rc}) \right\|_2^2 \leq& \frac{2}{n}\left\langle \chi(M, \mathcal{P}, \varepsilon), \Delta(L) \right\rangle \\
&+ \lambda\left(\left\| L^* \right\|_* - \left\| \hat{L} \right\|_* \right) + \frac{K_4 RC^{2-A^2/8}\sqrt{\log RC}}{n}.
\end{aligned}$$

Next, we need to bound the nuclear norm and the Frobenius norm of the low-rank part.

By Lemma S2, we have that,

$$\begin{aligned}
&\sum_{r=1}^{R}\sum_{c=1}^{C}\frac{K_3}{n}\left\|M_{rc}\Delta(L_{rc})\right\|_2^2\\
&\le \frac{2}{n}\left\langle \chi(M,\mathcal{P},\varepsilon),\Delta(L)\right\rangle+\lambda\left\|\Delta(L)\right\|_*+\frac{K_4RC^{2-A^2/8}\sqrt{\log RC}}{n}\\
&\le \frac{2}{n}\left\|\chi(M,\mathcal{P},\varepsilon)\right\|_{\text{op}}\left\|\Delta(L)\right\|_*+\lambda\left\|\Delta(L)\right\|_*+\frac{K_4RC^{2-A^2/8}\sqrt{\log RC}}{n}\\
&\le 2\lambda\left\|\Delta(L)\right\|_*+\frac{K_4RC^{2-A^2/8}\sqrt{\log RC}}{n}\\
&\le 2\lambda\left(\left\|\Delta'(L)\right\|+\left\|\Delta''(L)\right\|\right)+\frac{K_4RC^{2-A^2/8}\sqrt{\log RC}}{n}\\
&\le 8\lambda\left\|\Delta'(L)\right\|_*+8\lambda\sum_{j=r+1}^{\min(R,C)}\sigma_j(L^*)+\frac{K_4RC^{2-A^2/8}\sqrt{\log RC}}{n}\\
&\le 8\sqrt{2r}\lambda\left\|\Delta(L)\right\|_2+8\lambda\sum_{j=r+1}^{\min(R,C)}\sigma_j(L^*)+\frac{K_4RC^{2-A^2/8}\sqrt{\log RC}}{n}.
\end{aligned}$$

Since $K_3\left\|M_{rc}\Delta(L_{rc})\right\|_2^2/n=K_5\left\|\Delta(L)\right\|_2^2$, it holds that

$$\left\|\Delta(L)\right\|_2\le\max\left\{\frac{8\sqrt{2r}\lambda}{K_5},\sqrt{\frac{8\lambda\sum_{j=r+1}^{\min(R,C)}\sigma_j(L^*)}{K_5}},\sqrt{\frac{K_4}{K_5n}}RC^{1-A^2/16}\log^{1/4}(RC)\right\}.$$

Since $\lambda=\frac{2}{n}\left\|\chi(M,\mathcal{P},\varepsilon)\right\|_{\text{op}}$, then in the set

$$\mathcal{B}_q(v_q)=\left\{A\in\mathbb{R}^{R\times C}:\sum_{j=1}^{\min(R,C)}\theta_j^q(A)\le v_q\right\},$$

$\lambda\le K_6\sqrt{\frac{R+C}{n}}$ with probability at least $1-\exp\{-(R+C)\}$. Thus, with probability at least $1-\exp\{-(R+C)\}$,

$$\left\|\Delta(L)\right\|_2\le\max\left\{K_7v_q^{1/2}\left(\frac{R+C}{n}\right)^{1/2-q/4},\sqrt{\frac{K_4}{K_6n}}RC^{1-A^2/16}\log^{1/4}(RC)\right\}.$$

Then the result in Theorem 1 is a simple extension of this result when $q=0$.

Under the alternative hypothesis, since $(\hat{L},\hat{S})$ is the minimizer of

$$\frac{1}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\left\|Y_{rc}-X_{rc}\hat{\beta}_{rc}-M_{rc}L_{rc}-M_{-rc}S_{rc}\right\|_2^2+\lambda\left\|L\right\|_*+\sum_{r=1}^{R}\sum_{c=1}^{C}\lambda_{rc}\left\|S_{rc}\right\|_1.$$

Similarly, we have

$$\frac{1}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\|M_{rc}\Delta(L_{rc})+M_{-rc}\Delta(S)_{rc}\|_2^2 \leq \frac{2}{n}\langle\chi(M,\mathcal{P},\varepsilon),\Delta(L)\rangle+\lambda\left(\|L^*\|_*-\left\|\hat{L}\right\|_*\right)$$

$$+\sum_{r=1}^{R}\sum_{c=1}^{C}\lambda_{rc}\left(\|S^*_{rc}\|_1-\left\|\hat{S}_{rc}\right\|_1\right).$$

In this case, when $\lambda \geq 2\|\chi(M,\mathcal{P},\varepsilon)\|_{\mathrm{op}}$, there exists a decomposition

$$\Delta(L)=\Delta'(L)+\Delta''(L),$$

where $\mathrm{rank}(\Delta'(L)) \leq 2r$ and

$$\|\Delta''(L)\|_* \leq 3\|\Delta'(L)\|_*+4\sum_{j=r+1}^{\min(R,C)}\sigma_j(L^*)+2\sum_{r=1}^{R}\sum_{c=1}^{C}\lambda_{rc}\|\Delta(S_{rc})\|_1/\lambda.$$

Since $(\hat{\beta},\hat{L},\hat{S})$ is the minimizer of

$$\frac{1}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\|Y_{rc}-X_{rc}\beta_{rc}-M_{rc}L_{rc}-M_{-rc}S_{rc}\|_2^2+\lambda\|L\|_*+\sum_{r=1}^{R}\sum_{c=1}^{C}\lambda_{rc}\|S_{rc}\|_1.$$

Then we have that

$$\begin{aligned}
&\frac{1}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\|X_{rc}\Delta(\beta_{rc})+M_{rc}\Delta(L_{rc})+M_{-rc}\Delta(S_{rc})\|_2^2\\
&\leq \frac{2}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\left\|\varepsilon_{rc}^{\top}X_{rc}\right\|_2\|\Delta(\beta_{rc})\|_2+\frac{2}{n}\|\chi(M,\varepsilon)\|_{\mathrm{op}}\|\Delta(L)\|_*\\
&+\frac{2}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\left\|\varepsilon_{rc}^{\top}M_{-rc}\right\|_\infty\|\Delta(S_{rc})\|_1+\lambda\|\Delta(L)\|_*+\sum_{r=1}^{R}\sum_{c=1}^{C}\lambda_{rc}\|\Delta(S_{rc})\|_1.
\end{aligned}$$

Recalling the selection of tuning parameter, it holds that

$$
\begin{aligned}
&\frac{1}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\|X_{rc}\Delta(\beta_{rc})+M_{rc}\Delta(L_{rc})+M_{-rc}\Delta(S_{rc})\|_2^2\\
&\leq \frac{2}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\left\|\varepsilon_{rc}^{\top}X_{rc}\right\|_2\|\Delta(\beta_{rc})\|_2+2\lambda\|\Delta(L)\|_*+4\sum_{r=1}^{R}\sum_{c=1}^{C}\lambda_{rc}\|\Delta(S_{rc})\|_1\\
&\leq \frac{2}{\sqrt{n}}\sum_{r=1}^{R}\sum_{c=1}^{C}\left\|\frac{\varepsilon_{rc}^{\top}X_{rc}}{n}\right\|_2\|\Delta(\beta_{rc})\|_2+8\sqrt{2r}\lambda\|\Delta(L)\|_2+8\lambda\sum_{j=r+1}^{\min(R,C)}\sigma_j(L^*)\\
&+12\sum_{r=1}^{R}\sum_{c=1}^{C}\lambda_{rc}\left(\|\Delta(S_{rc})_{\delta_1}\|_1+\|\Delta(S_{rc})_{\delta_0}\|_1\right)\\
&\leq \frac{2}{\sqrt{n}}\sum_{r=1}^{R}\sum_{c=1}^{C}\left\|\frac{\varepsilon_{rc}^{\top}X_{rc}}{n}\right\|_2\|\Delta(\beta_{rc})\|_2+8\sqrt{2r}\lambda\|\Delta(L)\|_2+8\lambda\sum_{j=r+1}^{\min(R,C)}\sigma_j(L^*)\\
&+24\sum_{r=1}^{R}\sum_{c=1}^{C}\lambda_{rc}\|\Delta(S_{rc})_{\delta_1}\|_1\\
&\leq \frac{2}{\sqrt{n}}\sum_{r=1}^{R}\sum_{c=1}^{C}\left\|\frac{\varepsilon_{rc}^{\top}X_{rc}}{n}\right\|_2\|\Delta(\beta_{rc})\|_2+8\sqrt{2r}\lambda\|\Delta(L)\|_2+8\lambda\sum_{j=r+1}^{\min(R,C)}\sigma_j(L^*)\\
&+24\sum_{r=1}^{R}\sum_{c=1}^{C}\lambda_{rc}\sqrt{s_{rc}}\|\Delta(S_{rc})\|_2\\
&\leq \frac{2}{\sqrt{n}}\left\{\max_{r,c}\left\|\frac{\varepsilon_{rc}^{\top}X_{rc}}{n}\right\|_2\right\}\sqrt{RC}\|\Delta(\beta)\|_2+8\sqrt{2r}\lambda\|\Delta(L)\|_2+8\lambda\sum_{j=r+1}^{\min(R,C)}\sigma_j(L^*)\\
&+24\max_{r,c}\{\lambda_{rc}\sqrt{s_{rc}}\}\sqrt{RC}\|\Delta(S)\|_2.
\end{aligned}
$$

According to Assumption 3, let $K_8=\kappa_{rc}^2(s_{rc},1)$, then we have

$$
K_8\|\Delta(\beta,L,S)\|_2^2\leq\frac{1}{n}\sum_{r=1}^{R}\sum_{c=1}^{C}\|X_{rc}\Delta(\beta_{rc})+M_{rc}\Delta(L_{rc})+M_{-rc}\Delta(S_{rc})\|_2^2,
$$

and

$$
\|\Delta(\beta,L,S)\|\leq\max\left\{\max_{r,c}\left\|\frac{\varepsilon_{rc}^{\top}X_{rc}}{n}\right\|_2\frac{2\sqrt{3RC}}{K_8\sqrt{n}},\frac{8\sqrt{6r}\lambda}{K_8},\max_{r,c}\{\lambda_{rc}\sqrt{s_{rc}}\}\frac{24\sqrt{3RC}}{K_8}\right\}.
$$

Then, it is easy to prove that with probability at least $1-\exp\{-(R+C)\}$,

$$
\|\Delta(S,L)\|_2\leq K_9\max_{r,c}\left\{\sigma_{rc}\sqrt{s_{rc}\log(RC/n)}\right\}\sqrt{RC},
$$

with $K_9=32A/\min_{r,c}\kappa_{r,c}^2(s_{rc},1)$. This completes the proof of Theorem 1. ■

S5.4 Proof of Theorem 2

For size analysis, consider

$$\begin{aligned}\tilde{\varepsilon}_{rc} &= \varepsilon_{rc} - X_{rc}\Delta(\beta_{rc}) - M_{rc}\Delta(L_{rc}) \\ &= (I - \mathcal{P}_{X_{rc}})\varepsilon - (I - \mathcal{P}_{X_{rc}})M_{rc}\Delta(L_{rc}) + \mathcal{P}_{X_{rc}}M_{-rc}\Delta(S_{rc}) \\ &= \bar{\varepsilon}_{rc} + \bar{\Delta}_{rc}.\end{aligned}$$

By the error bound of L in Theorem 1 and Assumption 4, 6, with probability at least $1 - \exp\{-(R+C)\}$, there exists a fixed constant K_{10}, K_{11} such that

$$\left|\bar{\Delta}_{irc}\right| \leq K_{10}\frac{v_0^{1/2}\left(R+C\right)^{1/2}}{(nRC)^{1/2}},$$

for $i = 1, \ldots, n$.

Since $\sqrt{n}\hat{S}_{rc}$ is the minimizer of

$$V_n^{rc}(u_{rc}) = -2u_{rc}^\top\left(\frac{1}{n}M_{-rc}^\top\tilde{\varepsilon}_{rc}\right) + u_{rc}^\top\left(\frac{1}{n}M_{-rc}^\top M_{-rc}\right)u_{rc} + \frac{\lambda_{rc}}{\sqrt{n}}\left\|u_{rc}\right\|_1,$$

then $\sqrt{n}\hat{S}$ is the minimizer of

$$V_n(u) = -2u^\top\left(\frac{1}{\sqrt{n}}G^\top\tilde{\varepsilon}\right) + u^\top\left(\frac{1}{n}G^\top G\right)u + \frac{1}{\sqrt{n}}\left\|\Lambda u\right\|_1,$$

where $u = (u_{11}^\top, \ldots, u_{RC}^\top)^\top$, $\tilde{\varepsilon} = (\tilde{\varepsilon}_{11}^\top, \ldots, \tilde{\varepsilon}_{RC}^\top)^\top$ and $G = \operatorname{diag}\{M_{-11}, \ldots, M_{-RC}\}$. Let

$$f(Z) = \arg\min_u -2u^\top Z + u^\top\left(\frac{1}{n}G^\top G\right)u + \left\|\Lambda_0 u\right\|_1,$$

then we have that for a rectangle A, there exists a rectangle B such that

$$\mathbb{P}\left(\sqrt{n}\hat{S} \in A\right) = \mathbb{P}\left(1/\sqrt{n}G^\top\tilde{\varepsilon} \in B\right).$$

Similarly, $\sqrt{n}\bar{S}$ is the minimizer of

$$\bar{V}_n(u) = -2u^\top\left(\frac{1}{\sqrt{n}}G^\top\bar{\varepsilon}\right) + u^\top\left(\frac{1}{n}G^\top G\right)u + \frac{1}{\sqrt{n}}\left\|\Lambda u\right\|_1,$$

then $\mathbb{P}(\sqrt{n}\bar{S} \in A) = \mathbb{P}(1/\sqrt{n}G^\top \bar{\varepsilon} \in B)$. At the same time,

$$\left|\mathbb{P}\left(\frac{1}{\sqrt{n}}G^\top \tilde{\varepsilon} \in B\right) - \mathbb{P}\left(\frac{1}{\sqrt{n}}G^\top \bar{\varepsilon} \in B\right)\right| \leq \max_j \mathbb{P}\left[\left|\left(\frac{1}{\sqrt{n}}G^\top \bar{\varepsilon}\right)_j\right| \leq K_{11}\frac{v_0^{1/2}(R+C)^{1/2}}{(RC)^{1/2}}\right] \to 0$$

holds with probability at least $1 - \exp\{-(R+C)\}$.

Thus, we have that with probability at least $1 - \exp\{-(R+C)\}$,

$$\sup_{B\in\mathcal{B}}\left|\mathbb{P}\left(\frac{1}{\sqrt{n}}G^\top \tilde{\varepsilon} \in B\right) - \mathbb{P}\left(S_n^F \in B\right)\right| \leq 2K_{12}B_n^2\left\{\log^7\left(R^2C^2n\right)/n\right\}^{1/6}.$$

Next, we consider the bootstrap procedure. Let $\varepsilon_{rc}^{(b)} = (I - \mathcal{P}_{X_{rc}})\, e_{rc}^{(b)}$, where $e_{rc}^{(b)} \sim \mathcal{N}(0, \hat{\sigma}_{rc}^2 I)$, and $\hat{\sigma}_{rc}$ is estimated by scale lasso procedure under the lasso problem

$$\frac{1}{n}\left\|Y_{rc} - \alpha_{rc} - Z_{rc}\gamma_{rc}\right\|_F^2 + \lambda\left\|\gamma_{rc}\right\|_1$$

with $Z_{rc} = (X_{rc}, M_{rc}, A_{rc})$ and $\gamma_{rc} = (\beta_{rc}, L_{rc}, S_{rc})$. Then similarly, $\sqrt{n}\hat{S}^{(b)}$ is the minimizer of

$$V_n^{(b)}(u) = -2u^\top\left(\frac{1}{\sqrt{n}}G^\top \varepsilon^{(b)}\right) + u^\top\left(\frac{1}{n}G^\top G\right)u + \frac{1}{\sqrt{n}}\left\|\Lambda u\right\|_1,$$

and $\mathbb{P}(\sqrt{n}\hat{S}^{(b)} \in A) = \mathbb{P}(1/\sqrt{n}G^\top \varepsilon^{(b)} \in B)$. Next we derive the bound of

$$\rho_n^{MB} = \sup_{B\in\mathcal{B}}\left|\mathbb{P}\left(\frac{1}{\sqrt{n}}G^\top \varepsilon^{(b)} \in B\right) - \mathbb{P}(S_n^F \in B)\right|.$$

Since $\mathbb{E}\left\{1/\sqrt{n}G^\top \varepsilon^{(b)} \in B\right\} = \mathbb{E}\left\{S_n^F\right\} = 0$, we only need to calculate there covariance matrix. Let $D = 1/nG^\top\left(I - \mathcal{P}_X\right)G$, $\Sigma = \operatorname{diag}\left\{\sigma_{11}^2 I, \ldots, \sigma_{RC}^2 I\right\}$ and $\hat{\Sigma} = \operatorname{diag}\left\{\hat{\sigma}_{11}^2 I, \ldots, \hat{\sigma}_{RC}^2 I\right\}$. We have that

$$\frac{1}{\sqrt{n}}G^\top \varepsilon^{(b)} \sim \mathcal{N}\left(0, \hat{\Sigma}D\right), S_n^F \sim \mathcal{N}\left(0, \Sigma D\right)$$

Let $\hat{\Delta}_n = \left\|\hat{\Sigma}D - \Sigma D\right\|_\infty$. We present a lemma here, which is Lemma 5 of Xue & Yao (2020). One can refer to Xue & Yao (2020) for a detailed proof of Lemma S6.

Lemma S6 *If* $\min_j \mathbb{E}\left(S_{nj}^{\mathcal{G}}\right)^2 \geq b > 0$, *then for any sequence of constants* $\bar{\Delta}_n > 0$, *on the event* $\left\{\hat{\Delta}_n \leq \bar{\Delta}_n\right\}$, *we have the following,*

$$\rho_n^{MB} \lesssim \left(\bar{\Delta}_n\right)^{1/3}\left(\log R^2C^2\right)^{2/3}.$$

Since $\hat{\Delta}_n = \left\|\hat{\Sigma} D - \Sigma D\right\|_\infty = \left\|\hat{\Sigma} - \Sigma\right\|_\infty \|D\|_\infty = \max|\hat{\sigma}^2_{rc} - \sigma^2| \|D\|_\infty$, by Theorem 2 in Sun & Zhang (2013) and Assumption 5,

$$\sqrt{n}\left(\hat{\sigma}_{rc} - \sigma_{rc}\right) \to_d \mathcal{N}(0, 1/2\sigma^2_{rc}).$$

Noting that $\hat{\sigma}_{rc}$ is independent across r, c, we have that with probability at least $1 - \exp\left\{-\sqrt{\log RC}\right\}$,

$$\begin{aligned}
\hat{\Delta}_n &= \max\left|\hat{\sigma}^2_{rc} - \sigma^2\right| \|D\|_\infty = K_{13} \max\left|\hat{\sigma}^2_{rc} - \sigma^2\right| \\
&\le 2K_{15} \max \sigma_{rc} \max|\hat{\sigma}_{rc} - \sigma_{rc}| \\
&\le 2K_{15} \max \sigma_{rc} \sqrt{\frac{2\log(RC)}{n}}.
\end{aligned}$$

Let $\bar{\Delta}_n = 2K_{15} \max \sigma_{rc} \sqrt{\frac{2\log(RC)}{n}}$, then with probability 1,

$$\rho_n^{MB} \lesssim \left\{2K_{15} \max \sigma_{rc} \sqrt{\frac{2\log(RC)}{n}}\right\}^{1/3} \left(\log R^2C^2\right)^{2/3} \le K_{16} (\max \sigma_{rc})^{1/3} \left\{\frac{\log^5 RC}{n^2}\right\}^{1/6}.$$

Finally, we have that with probability at least $1 - \exp\{-(R+C)\} - \exp\left\{-\sqrt{\log RC}\right\}$,

$$\begin{aligned}
&\sup_{t\ge 0}\left|\mathbb{P}\left(\left\|\sqrt{n}\hat{S}^{(b)}\right\|_\infty \le t\right) - \mathbb{P}\left(\left\|\sqrt{n}\hat{S}\right\|_\infty \le t\right)\right| \\
&\le \sup_{A\in\mathcal{A}}\left|\mathbb{P}\left(\sqrt{n}\hat{S}^{(b)} \in A\right) - \mathbb{P}\left(\sqrt{n}\hat{S} \in A\right)\right| \\
&= \sup_{B\in\mathcal{B}}\left|\mathbb{P}\left(\frac{1}{\sqrt{n}} G^\top \varepsilon^{(b)} \in B\right) - \mathbb{P}\left(\frac{1}{\sqrt{n}} G^\top \tilde{\varepsilon} \in B\right)\right| \\
&= \sup_{B\in\mathcal{B}}\left|\mathbb{P}\left(\frac{1}{\sqrt{n}} G^\top \varepsilon^{(b)} \in B\right) - \mathbb{P}\left(S_n^F \in B\right)\right| + \sup_{B\in\mathcal{B}}\left|\mathbb{P}\left(S_n^F \in B\right) - \mathbb{P}\left(\frac{1}{\sqrt{n}} G^\top \tilde{\varepsilon} \in B\right)\right| \\
&\lesssim 2K_{15} B_n^2 \left\{\log^7\left(R^2C^2n\right)/n\right\}^{1/6} + K_{16} (\max \sigma_{rc})^{1/3} \left\{\frac{\log^5 RC}{n^2}\right\}^{1/6}.
\end{aligned}$$

The proof of Theorem 2 is completed. ■

S5.5 Proof of Theorem 3

From the error bound in Theorem 1, with probability at least $1 - \exp\{-(R+C)\}$,

$$\begin{aligned}
&\mathbb{P}\left(\left\|\sqrt{n}\hat{S}\right\|_{\infty} \geq c_B(\alpha)\right) \geq \mathbb{P}\left\{\left\|\sqrt{n}\left(\hat{S}-S^*\right)\right\|_{\infty} \leq \|\sqrt{n}S^*\|_{\infty} - c_B(\alpha)\right\} \\
&\geq \mathbb{P}\left\{\left\|\sqrt{n}\left(\hat{S}-S^*\right)\right\|_{\infty} \leq \|\sqrt{n}S^*\|_{\infty} - c_B(\alpha) - \left\|\sqrt{n}\hat{S}^{(b)}\right\|_{\infty}\right\} \\
&= \mathbb{P}\left\{\left\|\sqrt{n}\hat{S}^{(b)}\right\| \leq \|\sqrt{n}S^*\|_{\infty} - c_B(\alpha) - \left\|\sqrt{n}\left(\hat{S}-S^*\right)\right\|_{\infty}\right\} \\
&\geq \mathbb{P}\left\{\left\|\sqrt{n}\hat{S}^{(b)}\right\|_{\infty} \leq \|\sqrt{n}S^*\|_{\infty} - c_B(\alpha) - K_9 \max_{r,c}\left\{\sigma_{rc}\sqrt{s_{rc}\log(RC)}\right\}\right\}.
\end{aligned}$$

Then we need identify the quantity of $\left\|\sqrt{n}\hat{S}^{(b)}\right\|_{\infty}$. For region (r,c) and $\lambda_{rc} = A\sigma_{rc}\sqrt{\frac{\log(RC)}{n}}$ with $A > 2\sqrt{2}$, it holds that

$$\begin{aligned}
\mathbb{P}\left(\hat{S}^{(b)} = 0\right) &= \left\{\mathbb{P}\left(\left\|\frac{2}{n}\varepsilon_{rc}^{\top}M_{-rc}\right\|_{\infty} \geq \lambda_{rc}\right)\right\}^{RC} \to \left\{1 - \sqrt{\frac{1}{\pi\log RC}}\left(\frac{1}{RC}\right)^{A^2/8-1}\right\}^{RC} \\
&\to \exp\left\{-\frac{1}{\sqrt{\pi\log RC}\,(RC)^{A^2/8-2}}\right\}.
\end{aligned}$$

Then for $2\sqrt{2} < A < 4$, under which $\mathbb{P}\left(\hat{S}^{(b)} = 0\right) \to 0$ and if $\hat{S}^{(b)} \neq 0$, let J_1 be the index set of the nonzero elements of $\hat{S}^{(b)}$, then we have that

$$\frac{1}{n}G_{J_1}^{\top}\left(\varepsilon^{(b)} - G_{J_1}\hat{S}_{J_1}^{(b)}\right) = \Lambda_{J_1}\text{sign}\left(\hat{S}_{J_1}^{(b)}\right)/2.$$

Then we have that

$$\frac{1}{n}\left\|G_{J_1}^{\top}G_{J_1}\hat{S}_{J_1}^{(b)}\right\|_{\infty} \leq \frac{1}{n}\left\|G_{J_1}^{\top}\varepsilon^{(b)}\right\|_{\infty} + \max_{r,c}\lambda_{rc}/2,$$

by Assumption 3 and $\kappa_{rc}(1,3) = 1$, we have that,

$$\sqrt{n}\left\|\hat{S}^{(b)}\right\|_{\infty} \cdot \min_{r,c}\kappa(s_{rc},3) \leq \left\|G_{J_1}^{\top}\varepsilon^{(b)}/\sqrt{n}\right\|_{\infty} + \sqrt{n}\max_{r,c}\lambda_{rc}/2.$$

Then by Lemma S1 and the elements of G come from $\{-1,1\}$, we have that

$$c_B(\alpha) \leq \frac{\sigma_{\max}}{\min_{r,c}\kappa_{rc}(s_{rc},3)}\left\{\sqrt{4\log RC - \log\log RC - \log\pi - \log(1/1-\alpha)} + \sqrt{2\log RC}\right\},$$

Similarily, by

$$\frac{1}{n}\left\|G_{J_1}^\top G_{J_1}\hat{S}_{J_1}^{(b)}\right\|_\infty \geq \frac{1}{n}\left\|G_{J_1}^\top \varepsilon^{(b)}\right\|_\infty - \max_{r,c}\lambda_{rc}/2,$$

we have that

$$c_B(\alpha) \geq \frac{\sigma_{\max}}{\min_{r,c}\kappa_{rc}(s_{rc},3)}\left\{\sqrt{4\log RC - \log\log RC - \log\pi - \log(1/1-\alpha)} - \sqrt{2\log RC}\right\},$$

Let $\|S^*\|_\infty \geq K_9\sigma_{\max}s_{\max}\sqrt{\log(R^2C^2n)/n}$ (Assumption 7). Then we have that

$$\left\|\sqrt{n}S^*\right\|_\infty - c_B(\alpha) - K_9\max_{r,c}\left\{\sigma_{rc}\sqrt{s_{rc}\log(RC)}\right\} \geq 2\max_{r,c}\sigma_{rc}\log^{1/2}(R^2C^2n).$$

Thus, it holds that,

$$\begin{aligned}
&\mathbb{P}\left(\left\|\sqrt{n}\hat{S}\right\|_\infty \geq c_B(\alpha)\right)\\
&\geq \mathbb{P}\left\{\left\|\sqrt{n}\hat{S}^{(b)}\right\|_\infty \leq 2\max_{r,c}\sigma_{rc}\log^{1/2}(R^2C^2n)\right\}\\
&\geq 1 - 2R^2C^2\exp\left\{-\frac{4\max_{r,c}\sigma_{rc}^2\log(R^2C^2n)}{2\max_{r,c}\sigma_{rc}^2}\right\}\\
&\geq 1 - \frac{8}{(nRC)^2} \to 1.
\end{aligned}$$

The proof of Theorem 3 is completed. ■

S5.6 Proof of Theorem 4

We first prove equation (7) in the theorem. Recall that

$$\begin{aligned}
\tilde{\varepsilon}_{rc} &= \varepsilon_{rc} - X_{rc}\Delta(\beta_{rc}) - M_{rc}\Delta(L_{rc})\\
&= (I-\mathcal{P}_{X_{rc}})\varepsilon - (I-\mathcal{P}_{X_{rc}})M_{rc}\Delta(L_{rc}) + \mathcal{P}_{X_{rc}}M_{-rc}\Delta(S_{rc})\\
&= \bar{\varepsilon}_{rc} + \bar{\Delta}_{rc}.
\end{aligned}$$

By the profiling procedure, $\bar{\Delta}_{rc}$ can be seen as a bias term and let J_0 be the index set of zero elements of S_{rc}, similarily, we have that

$$\left\|\frac{1}{n}M_{-rc}^\top\hat{S}_{rc}^{J_0}\right\|_\infty \leq \frac{1}{n}\left\|M_{-rc}^\top\bar{\varepsilon}_{rc}\right\|_\infty + \frac{1}{n}\left\|M_{-rc}^\top\bar{\Delta}_{rc}\right\|_\infty + \lambda_{rc}/2,$$

it has been proved that $\max\{|\Delta(L_{rc})|, \|\Delta(S_{rc}^{J_1})\|_1\} \leq K_9\sigma_{rc}s_{rc}\sqrt{\log(RC)/n}$. By Assumption 8(a), we have $\|M_{-rc}^{\top}\bar{\Delta}_{rc}\|_\infty \leq \sigma_{rc}\sqrt{\log(RC)/n}$. Recall that

$$\left\|\hat{S}_{rc}^{J_0}\right\|_\infty \leq \left\|\hat{S}_{rc}^{J_0}\right\|_1 \leq 4\left\|\hat{S}_{rc}\right\|_1,$$

then we have that

$$4\left\|\sqrt{n}\hat{S}_{rc}^{J_0}\right\|_\infty \cdot \kappa_{rc}(s_{rc}, 3) \leq \left\|M_{-rc}^{T}\bar{\varepsilon}_{rc}\right\|_\infty + \sigma_{rc}\sqrt{\log(RC)} + \sigma_{rc}\sqrt{2\log(RC)}.$$

By simple calculation,

$$\mathbb{P}\left(\frac{\left\|\sqrt{n}\hat{S}_{rc}^{J_0}\cdot\kappa_{rc}\right\|_\infty}{\sigma_{rc}} \leq \sqrt{\log RC} + \sqrt{2\log RC - \log\log RC + x}\right) \geq \exp\left\{-\frac{1}{\sqrt{\pi}}\exp\left(-\frac{x}{2}\right)\right\}.$$

This gives

$$\mathbb{P}\left(\left\|\sqrt{n}\hat{S}_{rc}^{J_0}\right\|_\infty \geq c_B(\alpha)\right) \leq \frac{1}{\sqrt{\pi RC\log(RC)}}.$$

In each unit (r, c), the stepdown or BiRS algorithm tends to select the signal points first. By Lemma S3 and Assumption 9, the probability that this two method select all the signal points is just $1 - 8\log(RC + 1)/n^2$. For the number of false detected signal points, since $\varepsilon_{rc}, r = 1, \ldots, R; c = 1, \ldots, C$ is independent across units, then by the profiling procedure, conditional on the estimates of L and β, we have for some constant K_{17},

$$\mathbb{P}(\hat{J}_1^{step} > \eta) \leq 1 - \sum_{k=\eta}^{RC}\binom{RC}{k}\left(1 - \frac{1}{\sqrt{\pi RC\log RC}}\right)^k\left(1 - \frac{1}{\sqrt{\pi RC\log RC}}\right)^{RC-k}$$

where $\hat{J}_1^{step}$ is the estimated signal sets by stepdown. Thus, we finish the proof of the first part of Theorem 4.

Next, we prove equation (8) of Theorem 4, we only need to validate the two conditions in Lemma S3.

For the first condition, recall that the lasso solution satisfies that if $\hat{S}_{rcj} \neq 0$ for $j \in J_0$, then

$$\frac{1}{n}M_{-rc,j}^{\top}\left(\tilde{Y}_{rc} - M_{-rc}^{J_0}\hat{S}_{rc}^{J_0}\right) = \lambda_{rc}\mathrm{sign}(\hat{S}_{rcj}),$$

where $\tilde{Y}_{rc} = Y_{rc} - X_{rc}\hat{\beta}_{rc} - M_{rc}\hat{L}_{rc} - M_{-rc}^{J_1}\hat{S}^{J_1}$, i.e.,

$$\frac{1}{n}M_{-rc,j}^\top\left(\dot{Y}_{rc} - M_{-rc}^{J_0}\hat{S}_{rc}\right) = \lambda_{rc}\text{sign}(\hat{S}_{rcj}) + \frac{1}{n}\Delta(L_{rc})M_{rc}^\top(I - \mathcal{P}_{X_{rc}})M_{-rc,j}$$
$$+ \frac{1}{n}\Delta^\top(S_{rc}^{J_0})M_{-rc}^{J_0\top}\mathcal{P}_{X_{rc}}M_{-rc,j} + \frac{1}{n}\Delta^\top(S_{rc}^{J_1})M_{-rc}^{J_1\top}(I - \mathcal{P}_{X_{rc}})M_{-rc,j},$$

where $\dot{Y}_{rc} = Y_{rc} - X_{rc}\beta_{rc}^* - M_{rc}L_{rc}^* - M_{-rc}^{J_1}S^{*J_1}$. By Assumption 8(b) and Assumption 6, we have that

$$\frac{1}{n}\Delta(L_{rc})M_{rc}^\top(I - \mathcal{P}_{X_{rc}})M_{-rc,j} + \frac{1}{n}\Delta^\top(S_{rc}^{J_0})M_{-rc}^{J_0\top}\mathcal{P}_{X_{rc}}M_{-rc,j}$$
$$+ \frac{1}{n}\Delta^\top(S_{rc}^{J_1})M_{-rc}^{J_1\top}(I - \mathcal{P}_{X_{rc}})M_{-rc,j} = o(\lambda_{rc}),$$

then the distributions of the estimator $\hat{S}_{rc}^{J_0}$ and $\hat{S}_{rc}^{(b)J_0}$ are asymptotically the same. Then we validate the first condition of Lemma S3.

For the second condition of Lemma S3, from the proof of Theorem 2, we have that

$$\mathbb{P}\left(\sqrt{n}\hat{S} \in A\right) = \mathbb{P}\left(1/\sqrt{n}G^\top\tilde{\varepsilon} \in B\right),$$

since G is block diagonal, in order to check the third condition, we only need to check the third condition in certain location (r, c). With out loss of generality, let

$$A_i = \left\{\left|\frac{1}{\sqrt{n}}M_{-rc,i}^\top\tilde{\varepsilon}\right| \geq t_i\right\},$$

with t_i is the $1-\alpha$ quantile of $\left|\frac{1}{\sqrt{n}}M_{-rc,i}^\top\tilde{\varepsilon}\right|$. By the central limit theorem and w.l.o.g, set $\sigma_{rc} = 1$, we have

$$Z_i = \frac{1}{\sqrt{n}}M_{-rc,i}^\top\tilde{\varepsilon} \sim \mathcal{N}(0, \Omega_{rc}),$$

with $H_{rc} = M_{-rc}^\top M_{-rc}/n$. We only need to prove the inequality of $D(H)$ with $H = RC - 1$. Denote $Z = (Z_1, \ldots, Z_H)^\top$, let x be the $1-\alpha$ quantile of standard normal. Then we have, let $d = K_3 s_{rc}$

$$
\begin{aligned}
\mathbb{P}\left(A_1 \cap \cdot \cap A_H\right) &= \frac{1}{(2\pi)^{H/2}\left|\Omega_{rc}\right|^{1/2}} \int_{|z|_{\min} \geq x} \exp\left(-\frac{1}{2} z^T \Omega_{rc}^{-1} z\right) \mathrm{d}z \\
&= \frac{1}{(2\pi)^{H/2}\left|\Omega_{rc}\right|^{1/2}} \int_{|z|_{\min} \geq x, \|z\|_2 > \log H} \exp\left(-\frac{1}{2} z^T \Omega_{rc}^{-1} z\right) \mathrm{d}z \\
&+ \frac{1}{(2\pi)^{H/2}\left|\Omega_{rc}\right|^{1/2}} \int_{|z|_{\min} \geq x, \|z\|_2 \leq \log H} \exp\left(-\frac{1}{2} z^T \Omega_{rc}^{-1} z\right) \mathrm{d}z \\
&\leq c \exp\left\{-\frac{\log^2 H}{2H}\right\} \\
&+ \frac{1}{(2\pi)^{H/2}\left|\Omega_{rc}\right|^{1/2}} \int_{|z|_{\min} \geq x, \|z\|_2 \leq \log H} \exp\left(-\frac{1}{2} z^T \Omega_{rc}^{-1} z\right) \mathrm{d}z \\
&\leq c d^{-H} + \frac{1}{(2\pi)^{H/2}\left|\Omega_{rc}\right|^{1/2}} \int_{|z|_{\min} \geq x, \|z\|_2 \leq \log H} \exp\left(-\frac{1}{2} z^T \Omega_{rc}^{-1} z\right) \mathrm{d}z.
\end{aligned}
$$

Since $\mathrm{Cov}\left(Z_i, Z_j\right) = o(1/d)$, we have

$$
\left\|\Omega_{rc}^{-1} - I\right\|_2 \leq \left\|\Omega_{rc}^{-1}\right\|_2 \left\|\Omega_{rc} - I\right\|_2 = o(1/d).
$$

Thus,

$$
\mathbb{P}\left(A_1 \cap \cdot \cap A_H\right) \leq \{1 + o(1)\}\left(\alpha^H + (1/d)^H\right).
$$

Then we have,

$$
D(H) \leq \left(\frac{\alpha^H + (1/d)^H}{\alpha^H}\right) \leq \frac{1}{(2\beta)^H},
$$

with $\beta = \frac{\alpha}{2(\alpha + 1/d)}$. Then we validate the second condition, then we finish the proof of Theorem 4.

■

References

Beck, A. & Teboulle, M. (2009), 'A fast iterative shrinkage-thresholding algorithm for linear inverse problems', *SIAM journal on imaging sciences* **2**(1), 183–202.

Cai, T. T. & Jiang, T. (2011), 'Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices', *The Annals of Statistics* **39**(3), 1496 – 1525.

Cai, T. T., Liu, W. & Xia, Y. (2014), 'Two-sample test of high dimensional means under dependence', *Joural of the Royal Statistical Society* **76**(2), 349–372.

Forastiere, L., Airoldi, E. M. & Mealli, F. (2021), 'Identification and estimation of treatment and interference effects in observational studies on networks', *Journal of the American Statistical Association* **116**(534), 901–918.

Leung, M. P. (2022), 'Causal inference under approximate neighborhood interference', *Econometrica* **90**(1), 267–293.

Liu, L., Hudgens, M. G. & Becker-Dreps, S. (2016), 'On inverse probability-weighted estimators in the presence of interference', *Biometrika* **103**(4), 829–842.

Meinshausen, N., Meier, L. & Bühlmann, P. (2009), 'p-values for high-dimensional regression', *Journal of the American Statistical Association* **104**(488), 1671–1681.

Rubin, D. B. (1986), 'Comment: Which ifs have causal answers', *Journal of the American statistical association* **81**(396), 961–962.

Sobel, M. E. (2006), 'What do randomized studies of housing mobility demonstrate? causal inference in the face of interference', *Journal of the American Statistical Association* **101**(476), 1398–1407.

Sun, T. & Zhang, C.-H. (2012), 'Scaled sparse linear regression', *Biometrika* **99**(4), 879–898.

Sun, T. & Zhang, C.-H. (2013), 'Sparse matrix inversion with scaled lasso', *The Journal of Machine Learning Research* **14**(1), 3385–3418.

Sävje, F., Aronow, P. & Hudgens, M. (2021), 'Average treatment effects in the presence of unknown interference', *Annals of statistics* **49**(2), 673.

Xue, K. & Yao, F. (2020), 'Distribution and correlation-free two-sample test of high-dimensional means', *The Annals of Statistics* **48**(3), 1304 – 1328.