Biometrika (2025), **103**, 1, *p.* 1
Printed in Great Britain

Inference for spatial random effects model under dependence

By S. Bonnerjee
Department of Statistics, University of Chicago
sohambonnerjee@uchicago.edu

S. Deb
Decision Sciences, Indian Institute of Management Bangalore
soudeep@iimb.ac.in

and W. B. Wu
Department of Statistics, University of Chicago
wbwu@uchicago.edu

Summary

We study statistical inference for spatial random effects models under general dependence structures. Unlike classical spatial econometric models that assume parametric autoregressive dependence or Gaussian random fields, our framework accommodates broad classes of weakly dependent spatial processes defined through functional dependence measures. Building on Wu's (2005) theory of physical dependence, we establish central limit theorems for least-squares estimators in both fixed- and random-design settings, under mild stability and leverage conditions. The asymptotic covariance structure is characterized nonparametrically, and a consistent HAC estimator of the asymptotic variance is developed. The proposed theory relaxes Gaussianity and parametric assumptions commonly imposed in spatial panel data analysis, thereby extending asymptotic inference to a wide family of nonlinear and non-Gaussian spatial random fields.

Some key words: spatial field, random effects, central limit theory, HAC

1. Introduction

Spatial interactions in real-life datasets are a common cause of heterogeneity, often arising in form of spill-over effects between cross-sectional units, or regional effects in panel or longitudinal spatial datasets. In this context, random effects or mixed effects modeling to account for spatial correlations is a well-studied topic of interest in econometrics and statistics. These endeavors have found wide applications in analysis of various public expenditure and policy affects, population and industrial growth e.g. Case (1991); LeSage (1999); Dharshing (2017); Imran et al. (2023); various production parameters e.g. Audretsch & Feldman (1996); Druska & Horrace (2004); Dasgupta et al. (2018); Romão & Nijkamp (2019); epidemiology and disease incidence modeling e.g. Reece & Hulse (2020); modelling housing prices e.g. Helpman (1998); Hanson (2005) etc. In this context of panel data, Swamy & Arora (1972) pioneered the use of random effects model to account for acrossindividuals interactions. However, they did not use it to model spatial interactions. On the other hand, classical texts on spatial modeling assumed a particular parametric set-up in order to use likelihood-based methods such as those in Cliff & Ord (1981), or generalized moments estimators, as in Kelejian & Prucha (2010). Related important research by Anselin & Bera (1998); Baltagi & Li (2006); Baltagi et al. (2007); Kapoor et al. (2007); Corrado & Fingleton (2011) consider uncorrelated region-specific random effects, and model the spatial interaction in lieu of a first order autoregressive model. We do not attempt to summarize such a huge literature in the small space of this paper; instead, we would point the readers to Anselin (2013); Baltagi (2021) for a detailed overview.

Moving beyond the spatial autoregressive structure, more general covariance structures for Gaussian random fields were explored in the seminal work by Stein (1999), Section 2.7. However, Gaussianity is an idealized assumption that often cannot be verified. Central limit theory can be established in general stationary spatial random fields El Machkouri et al. (2013); Deb et al. (2017), but their extension to applications geared towards spatial random effect modeling is yet to be explored. The fundamental challenge in this direction is that the regression estimator can no longer be expressed as a sample mean of a stationary random field; therefore, the arguments by the previously-mentioned papers do not carry forward automatically. Moreover, in absence of a well-defined covariance structure, it remains to provide a valid non-parametric estimate of the variance of the regression estimate.

In spatial analysis, most often one comes across irregularly spaced datasets. Concretely, suppose $(\Gamma_n)_{n\in\mathbb{N}} \subset \mathbb{Z}^d$ be the set of finite sampling locations. We assume that $\Gamma_n := \{L_{n,1}, \dots, L_{n,r_n}\}$ satisfies $r_n := |\Gamma_n| \to \infty$ as $n \to \infty$. For a streamlined presentation, we can ignore these intermediary parameters by letting $r_n = n$; moreover, we will ignore the subscripts n in the locations $L_{n,k}$ to write $L_k = L_{n,k}$. For each location L_i, $i \in [n]$, suppose observations are available for ℓ_i many replicates. Mathematically speaking, we have access to $(Y_{ij}, \boldsymbol{X}_{ij}) \in \mathbb{R} \otimes \mathbb{R}^p$, $j \in [\ell_i], i \in [n]$. In this article, we focus on the random affect model

$$Y_{ij} = \boldsymbol{X}_{ij}^\top \boldsymbol{\beta} + U_{L_i} + \varepsilon_{ij},\ \boldsymbol{\beta} \in \mathbb{R}^p,\ \boldsymbol{X}_{ij} \in \mathbb{R}^p,\ j \in [\ell_i],\ i \in [n], \tag{1}$$

where, U_{L_i} denotes the spatial random effect corresponding to the location L_i. Further, for ease of exposition, let $\mathbb{X} = (\boldsymbol{X}_{11} : \boldsymbol{X}_{12} : \dots : \boldsymbol{X}_{n\ell_n})^\top \in \mathbb{R}^{\sum_i \ell_i \times p}$ be the concatenated form of the design matrix, and $\boldsymbol{Y} = (Y_{11} : Y_{12} : \dots : Y_{n\ell_n})^\top \in \mathbb{R}^{\sum_i l_i}$ be the response vector. Note that in the model (1), only $(Y_{ij}, \boldsymbol{X}_{ij})$ are observed. Assuming $S = \mathbb{X}^\top \mathbb{X}$ to be invertible, we analyze the least square estimate, defined as

$$\hat{\boldsymbol{\beta}}_{LS} = S^{-1} \mathbb{X}^\top \boldsymbol{Y}. \tag{2}$$

1.1. *Organization of the paper*

1.2. *Notations*

For $i = (i_1, \cdots, i_p) \in \mathbb{Z}^d$, let $|i| = i_1 \cdots i_p$. We denote the set $\{1, \dots, n\}$ by $[n]$. The d-dimensional Euclidean space is $\mathbb{R}^d$. For a vector $a \in \mathbb{R}^d$, $|a|$ denotes its Euclidean norm. For a matrix $M \in \mathbb{R}^{d\times m}$, $\rho^\star(A)$ denotes its largest singular value. For a random vector $X \in \mathbb{R}^d$ and $p \ge 1$, we denote $\|X\|_p := (\mathbb{E}[|X|^p])^{1/p}$; in particular, $\|X\| := \|X\|_2$. We also denote in-probability convergence, and stochastic boundedness by $o_{\mathbb{P}}$ and $O_{\mathbb{P}}$ respectively. We write $a_n \lesssim b_n$ if $a_n \le C b_n$ for some constant $C > 0$, and $a_n \asymp b_n$ if $C_1 b_n \le a_n \le C_2 b_n$ for some constants $C_1, C_2 > 0$. Often we denote $a_n \lesssim b_n$ by $a_n = O(b_n)$. Additionally, if $a_n/b_n \to 0$, we write $a_n = o(b_n)$.

2. Spatial dependency structure

In this section, we describe our approach to capturing the dependency between different spatial locations. In contrast to a direct Bayesian approach by encoding the dependence into a joint across-location prior information, we provide a more general framework of dependent random fields by building on Wu (2005). Note that this approach has also appeared previously in El Machkouri et al. (2013); Deb et al. (2017).

Let $(e_k)_{k\in\mathbb{Z}^d}$ be i.i.d. random variables independent of ε's, and for $k \in \mathbb{Z}^d$, let $\mathcal{F}_k := \sigma(e_{k-j} : j \in \mathbb{Z}^d)$. Then we will model

$$U_k = g(\mathcal{F}_k),\ k \in \mathbb{Z}^d, \tag{3}$$

for a measurable function $g : \otimes_{s\in\mathbb{Z}^d}\mathbb{R} \to \mathbb{R}$. The characterization (3) is quite general, and arises naturally out of writing out the joint distribution of $(U_k)_{k\in\mathbb{Z}^d}$ in terms of compositions of conditional quantile functions of i.i.d. uniform random variables. Note that $\{U_k\}_{k\in\mathbb{Z}^d}$ can be viewed as a *spatial stationary process*, in that for any sequence $\{t_1, t_2, \ldots, t_m\} \subset \mathbb{Z}^d$ and any $k \in \mathbb{Z}$, it follows that $(U_{t_1}, U_{t_2}, \ldots, U_{t_m}) \overset{d}{=} (U_{t_1+k}, U_{t_2+k}, \ldots, U_{t_m+k})$.

Assume that $U_k \in L_q$, where L_q is the set of all random variables with finite q-th moment. In order to quantify the dependence across spatial locations, we use the idea of coupling (Wu (2005)) to define dependence measures. Let $e'_i, e_j, i, j \in \mathbb{Z}^d$ be i.i.d. The *functional dependence measure* (FDM) can be defined as follows.

Definition 1 (Functional Dependency Measure and Stability). *Let $U_i \in L_q, q \geqslant 1,\ i \in \mathbb{Z}^d$. Define*

$$\delta_{i,q} = \left\|U_i - U_{i,\{0\}}\right\|_q,\ \textit{where } U_{i,\{0\}} = g\left(e^*_{i-s}; s \in \mathbb{Z}^d\right),$$

and $e^*_j = e_j$ if $j \neq 0$ and $e^*_0 = e'_0$. Due to stationarity, we will let $\delta_{-i,q} := \delta_{i,q}$. Also for $m \geqslant 0$, let

$$\Theta_{m,q} = \sum_{|j|\geqslant m} \delta_{j,q}$$

be the Dependency-Adjusted Norm (DAN). The random field (U_i) defined in (3) is said to be q-stable if $\Delta_q := \Theta_{0,q} < \infty$.

The FDM $\delta_{i,p}$ at $i \in \mathbb{Z}^d$ encapsulates, on average, the effect of U_k on U_{k+i} for any $k \in \mathbb{Z}^d$. For example, in a *linear random field* defined by $U_k = \sum_{s\in\mathbb{Z}^d} a_s e_{k-s}$, $\delta_{i,q} = |a_i|\|e_0 - e^\star_0\|_q$. On the other hand, The DAN, $\Theta_{m,q}$ aggregates these effects over $|i| \geq m$, to characterize the tail decay of the dependence as the locations are further away from each other. Finally, the concept of *stability* enshrines a notion of weak-dependence, and acts as a version of long-run variance of the spatial random field. This set-up is quite general, accommodating widely ranging classes of dependent random fields. We illustrate with two examples of q-stable process.

Example 1 (Random fields with Lipschitz Bernoulli shifts). Consider a sequence of non-negative coefficients $\{b_i\}_{i\in\mathbb{Z}^d}$ such that $B := \sum_{i\in\mathbb{Z}^d} b_i < \infty$. A general class of spatially dependent random fields of the form (1) is described by a coordinate-wise Lipschitz function

$$|g(\boldsymbol{x}) - g(\boldsymbol{y})| \leq \sum_{i\in\mathbb{Z}^d} b_i|x_i - y_i|,\ \boldsymbol{x} = (x_i)_{i\in\mathbb{Z}^d}, \boldsymbol{y} = (y_i)_{i\in\mathbb{Z}^d} \in \otimes_{i\in\mathbb{Z}^d}\mathbb{R}. \tag{4}$$

For example, the spatial autoregressive process, defined as

$$U_i = F((U_{i-j})_{j\in\mathcal{N}}; e_i),\ \mathcal{N} \subset \mathbb{Z}^d \text{ is finite and } 0 \notin \mathcal{N},$$

satisfies (4). Observe that from (4) it also follows that

$$\delta_{i,p} \leq \sum_{i\in\mathbb{Z}^d} b_i \|e_i - e_i^\star\|_q = b_i\|e_0 - e_0'\|_q,$$

and consequently, $\Delta_q \leq B$. Therefore, the processes satisfying (4) are q-stable. In particular, such processes are known to satisfy *Geometric moment contraction* if $b_i = O(\rho^i)$ for some $\rho \in (0,1)$; see Example 2 of Deb et al. (2017). Another interesting example of this class is the spatial *max-stable* process $U_i = \max_{s\in\mathbb{Z}^d} a_s e_{i-s}, i \in \mathbb{Z}^d$, used to model spatial extreme events (Buishand et al., 2008; Ribatet & Sedki, 2013). Since it holds

$$|U_i - U_i^\star| = |\max\{a_i e_0, \max_{s\neq i} a_s e_{i-s}\} - \max\{a_i e_0^\star, \max_{s\neq i} a_s e_{i-s}\}| \leq |a_i||e_0 - e_0^\star|,$$

it follows from (4) that this process is also q-stable if $\{a_i\}_{i\in\mathbb{Z}^d}$ are absolutely summable.

Example 2 (Logarithmic stochastic volatility models). Recently, *rough stochastic volatility* models have been well explored in asset price models and as a good fit to option prices Bayer et al. (2020, 2021); Wu et al. (2022). Translated to our notation, a simplified version of this model can be represented as

$$U_i = \zeta_i \exp(\sum_{s\in\mathbb{Z}^d} \alpha_s z_{i-s}), i \in \mathbb{Z}^d,$$

where $z_k \overset{i.i.d.}{\sim} N(0,1)$, and ζ_i are i.i.d. independent of z_i with $\|\zeta_0\|_2 < \infty$. Here $e_i = (\zeta_i, z_i)$. Note that

$$\begin{aligned}\delta_{i,2} &\leq 2\|\zeta_0\|_2 \|\exp(\sum_{s\in\mathbb{Z}^d} \alpha_s z_{i-s}) - \exp(\sum_{s\in\mathbb{Z}^d} \alpha_s z^\star_{i-s})\|_2 \\ &\lesssim \|\exp(\sum_{s\in\mathbb{Z}^d, s\neq i} \alpha_s z_{i-s})\|_2 \|\exp(\alpha_i z_0) - \exp(\alpha_i z_0')\|_2 \\ &\lesssim \exp(c\sum_{s\neq i}\alpha_s^2)|\alpha_i|\exp(c\alpha_i^2),, \qquad (5)\end{aligned}$$

where the final inequality follows from $\|\exp(\alpha_i z_0) - \exp(\alpha_i z_0')\|_2^2 \leq \mathrm{Var}(\exp(\alpha_i z_0))$. Clearly, if $\sum_s |\alpha_s| < \infty$, then we must have

$$\Delta_2 \lesssim \sum_{s\in\mathbb{Z}^d} |\alpha_s| \exp(c\sum_{s\in\mathbb{Z}^d}\alpha_s^2) < \infty.$$

3. CENTRAL LIMIT THEORY FOR REGRESSION WITH RANDOM EFFECTS

In this section, we systematically develop an asymptotic theory for estimating $\boldsymbol{\beta}$ in (1) by $\hat{\boldsymbol{\beta}}_{LS}$. Our subsequent discussions are divided into two subsections. In Section 3.1, we describe the regularity conditions imposed upon which central limit theory. Moving on, in Section 3.2, we present the central limit theorem for $\boldsymbol{\beta}$ itself, and discuss its connotations. In particular, we also present an accompanying result for random design, that also highlights the practicality of our regularity conditions for the fixed-design scenario.

3.1. *Regularity conditions on* (1) *and* $\mathbb{X}$

Before we describe the critical assumptions underpinning our theoretical results, it is instrumental to introduce some notations. Let the auto-covariance function of the dependent spatial random effects U_{L_i}'s be defined as $\gamma_{i-j} := \mathrm{Cov}(U_i, U_j)$ for $i, j \in \mathbb{Z}^d$. For generic $k, l \geqslant 1$, denote by $\mathbf{1}_k = (1, \ldots, 1)^\top \in \mathbb{R}^k$. Define the matrix $\Sigma_U = J\Gamma J^\top$, where,

$$\Gamma = \begin{pmatrix} \gamma_0 & \gamma_{L_2-L_1} & \cdots & \gamma_{L_n-L_1} \\ \vdots & \vdots & \ldots & \vdots \\ \gamma_{L_1-L_n} & \cdots & \cdots & \gamma_0 \end{pmatrix} \in \mathbb{R}^{n\times n}, \; J = \begin{pmatrix} \mathbf{1}_{l_1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{l_2} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0} & \ldots & \ldots & \mathbf{1}_{l_n} \end{pmatrix} \in \mathbb{R}^{\sum_i \ell_i \times n}. \quad (6)$$

and $T = \mathbb{X}^\top \Sigma_U X \in \mathbb{R}^{p\times p}$, and

$$A = S^{-1} T S^{-1} \in \mathbb{R}^{p\times p}. \quad (7)$$

The following regularity assumptions are necessitated to ensure a valid Gaussian approximation.

Assumption 1. **(L)**: S is invertible, and it holds that $\max_{1\leqslant k\leqslant \sum_i \ell_i} (XS^{-1}\mathbb{X}^\top)_{kk} \to 0$ as $n \to \infty$.

(W): T is invertible, and there exists a constant $c > 0$ such that $\rho^\star(T^{-1}) \sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2 \leqslant c$.

(D): It holds that $\frac{\rho^\star(S)}{\sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2} \to 0$ as $n \to \infty$.

Remark 1. Condition **(L)** is also known as the *Huber's condition* Huber (1973). It controls the leverage, facilitating the application of Lindeberg condition in order to deduce the asymptotic normality of regression coefficients. In multivariate linear regression analysis, such assumptions have been ever-present (Arnold, 1980; Bickel & Freedman, 1983; Lei & Ding, 2021; Jochmans, 2022). In contrast, Condition **(W)** can be thought of as a weak-dependence condition. In particular, **(W)** ensures that the stationary processes $(U_i)_{i\in\mathbb{Z}^d}$ can be well-approximated by m-dependent processes for a certain choice of m. Finally, condition **(D)** ensures that it is the variance contributed by U_i's that dominates in the covariance structure of $\hat{\boldsymbol{\beta}}$. It is possible to derive central limit theorem in absence of **(D)**; however in these cases, the variance may become intractable and consequently, difficult to estimate directly.

It is illuminating to further gain perspective on Conditions **(W)** and **(D)** by exploring them for the much simpler intercept-only model

$$Y_{i,j} = \boldsymbol{\beta} + U_i + \varepsilon_{i,j}, \boldsymbol{\beta} \in \mathbb{R}, \quad (8)$$

with $(U_i)_{i\in\mathbb{Z}^d}$ are as in (3). For model (8), the least-square estimator of $\boldsymbol{\beta}$ simplifies to $\hat{\boldsymbol{\beta}} = \sum_{i,j} Y_{i,j} / \sum_i \ell_i$. For (8), **(W)** and **(D)** condense into

Assumption 2. **(W')** There exists a constant $c > 0$ such that $\frac{\sum_{i,k} \ell_i \ell_k \gamma_{|L_i - L_k|}}{\sum_i \ell_i^2} \geqslant c$.

(D') It holds that $\frac{\sum_i \ell_i}{\sum_i \ell_i^2} \to 0$ as $n \to \infty$.

Note that, under the model (8), **(L)** becomes trivial, since $(\sum_i \ell_i)^{-1} \leqslant |\Gamma_n|^{-1} \to 0$. Note that **(W')** effectively constrains the dependence of the spatial random effect to have a faster decay. On the other hand, the interpretation of **(D')** can be understood to be ensuring that ℓ_i's are large enough. Mathematically speaking, it implies either $\max_i \ell_i >> \sqrt{n}$, or in general, at least $\asymp n^c$ number of ℓ_i's are of the order at least $n^{(1-c)/2}$, $c \in (0,1)$, $d > 0$. Even though in principle, a Gaussian approximation result do not require this condition, in order to obtain an identifiable asymptotic variance, we would be required to capture the auto-correlations between random effects $(U_i)_{i\in\mathbb{Z}}$ separately to the variability due to the noise $\varepsilon_{i,j}$'s, which restricts us to the conditions such as **(D')** and **(D)**.

3.2. *Central limit theory*

In light of the Conditions **(L)**, **(W)**, **(D)**, we present our first main result: a central limit theorem for the estimate $\hat{\boldsymbol{\beta}}$.

THEOREM 1. *Grant assumptions **(L)**, **(W)**, **(D)** for sampling locations $(\ell_i)_{i=1}^n$, the random effects $U_k, k \in \mathbb{Z}^d$ and the design matrix X respectively, along with $\Delta_2 < \infty$. Recall A from (7). Then, assuming the model (1) , it holds that*

$$A^{-1/2}(\hat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta}) \xrightarrow{d} N(0, I_p). \tag{9}$$

*In particular, for model (8), **(W')**, **(D')** along with $\Delta_2 < \infty$ ensures that*

$$\sqrt{\frac{(\sum_{i=1}^n \ell_i)^2}{\sum_{i,k} \ell_i \ell_k \gamma_{|L_i - L_k|}}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0,1). \tag{10}$$

The relationship between **(W)**, **(D)** and **(W')**, **(D')** can be further analyzed through the lens of random design matrices.

THEOREM 2. *Consider the model (1). For the design matrix $\mathbb{X}$, and $i \in [n]$, suppose $\boldsymbol{X}_{ij} = (1 : \boldsymbol{Z}_{ij})^\top$ with the random variables $\boldsymbol{Z}_{ij} \in \mathbb{R}^{p-1}$ satisfying $\mathbb{E}[\boldsymbol{Z}_{ij}] = \boldsymbol{\mu}_i$, and $\boldsymbol{Z}_{ij} - \boldsymbol{\mu}_i \overset{i.i.d.}{\sim} \mathrm{subG}(\sigma_i^2)$ for all $j \in [\ell_i]$. Let $\sup_i(|\boldsymbol{\mu}_i| \vee \sigma_i) = O(1)$ and*

$$\lambda_{\min}\Big(\sum_{i,k} \ell_i \ell_k \gamma_{|L_i-L_k|} \tilde{\boldsymbol{\mu}}_i \tilde{\boldsymbol{\mu}}_k^\top\Big) \geq c_0 \sum_{i,k} \ell_i \ell_k \gamma_{|L_i-L_k|}, \text{ where } \tilde{\boldsymbol{\mu}}_i = (1 : \boldsymbol{\mu}_i^\top)^\top. \tag{11}$$

*Further grant $\Delta_2 < \infty$. Then, under the conditions **(W')** and **(D')**, (9) holds.*

4. ESTIMATION OF VARIANCE

Observe that from (9), we need to estimate

$$A = S^{-1}\mathbb{X}^\top \Sigma_U \mathbb{X} S^{-1},$$

where we recall that $S = \mathbb{X}^\top \mathbb{X}$. For $k \in \mathbb{Z}^d$, let

$$\mathcal{A}_k := \{(i,j) : L_i - L_j = k\}, \text{ and } \mathcal{L} := \{k \in \mathbb{Z}^d : \mathcal{A}_k \text{ is non-empty}\}.$$

Let $K : \mathbb{R} \to \mathbb{R}$ be a symmetric kernel with bounded support $[-\omega, \omega]$, with $K \in \mathcal{C}^1$, and $\sup_x |K'(x)| \leqslant C$. With a slight abuse of notation, for $v \in \mathbb{R}^d$, let $K(v) := K(v_1) \cdots K(v_d)$. Since the spatial stationary field $(U_i)_{i\in\mathbb{Z}}$ is unobserved, we will employ an estimate based on $R_{\ell_i} := \bar{Y}_i - \bar{\boldsymbol{X}}_{i\cdot}^\top \hat{\boldsymbol{\beta}}$. For a bandwidth $B_n \to \infty$, define

$$\hat{\Gamma} := \begin{pmatrix} R_{L_1}^2 & R_{L_1}R_{L_2} & \dots & R_{L_1}R_{L_n} \\ \vdots & \vdots & \dots & \vdots \\ R_{L_n}R_{L_1} & R_{L_n}R_{L_2} & \dots & R_{L_n}^2 \end{pmatrix},$$

and let

$$\hat{\Sigma}_U(B_n) = J(\hat{\Gamma} \circ \mathcal{K}_n)J^\top; \ \hat{A}(B_n) = S^{-1}\mathbb{X}^\top \hat{\Sigma}_U(B_n)\mathbb{X}S^{-1}, \tag{12}$$

with $(\mathcal{K}_n)_{i,j} = K(B_n^{-1}(L_i - L_j))$, and $\circ$ denotes matrix Hadamard product. In order to infer the consistency of $\hat{A}$, we will impose the following two additional conditions on the design.

Assumption 3. **(V1)** It holds that $\frac{\max_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2} \to 0$ as $n \to \infty$.
(V2) It holds that $\rho^\star(S^{-1})\sqrt{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^4} \to 0$ as $n \to \infty$.

Condition **(V1)** is the standard "balance" condition, which ensures the sums of the design matrix at each of the locations, are not widely-varying. In other words, **(V1)** establishes a restriction on between-group variability. In contrast, **(V2)** can be viewed through the lens of within group variability. Indeed, if one assumes a control on the within-group variability that for each location i, $\max_{1\leqslant j \leqslant \ell_i} |\boldsymbol{X}_{ij}|^2 \leqslant C|\bar{\boldsymbol{X}}_{i\cdot}|^2$ for some fixed constant $C > 0$, then it is easy to deduce

$$\max_{1\leqslant k \leqslant \sum_i \ell_i} (\mathbb{X}S^{-1}\mathbb{X}^\top)_{kk} \leqslant C\rho^\star(S^{-1}) \max_i |\bar{\boldsymbol{X}}_{i\cdot}|^2 \leqslant \frac{C}{\min_i \ell_i} \rho^\star(S^{-1}) \sqrt{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^4} \to 0, \ \text{as } n \to \infty,$$

where the limiting assertion follows from **(V2)** and $\min_i \ell_i \geqslant 1$. In this sense, Condition **(V2)** can also be be viewed as a stronger version of **(L)**. Note that $\max_{1\leqslant j\leqslant \ell_i} |\boldsymbol{X}_{ij}|^2 \leqslant C|\bar{\boldsymbol{X}}_{i\cdot}|^2$ will usually be satisfied as long as ℓ_i's are not too large. However, we do not keep ourselves restricted to this condition, and resort to a more general assumption **(V2)**. It is useful to also look that the versions of **(V1)** and **(V2)** for the much simpler model (8), whereupon the necessity of these additional restrictions will become clearer. Indeed, for the intercept-only model, **(V1)** and **(V2)** condenses into

Assumption 4. **(V')** It holds that $\frac{\max_i \ell_i^2}{\sum_i \ell_i^2} = o(1)$,

since $\frac{\sum_i \ell_i^2}{(\sum_i \ell_i)^2} \leqslant \frac{\max_i \ell_i}{\sum_i \ell_i} \leqslant \frac{\max_i \ell_i^2}{\sum_i \ell_i^2}$. Therefore, for the intercept-only model, **(V')** alone suffices in establishing the necessary control on the magnitudes of ℓ_i. For the general case of a fixed design, note that in Assumption 3, the presence of $|\bar{\boldsymbol{X}}_{i\cdot}|$ precludes the possibility of **V1** implying **V2**. This is to be expected, since the magnitudes of the design matrix X is also needed to be controlled to ensure a consistent estimation of variance, thereby necessitating two separate, slightly convoluted conditions. The following result establishes the consistency of our variance estimate under these conditions.

THEOREM 3. *For the model* (1)*, suppose the sampling locations* $(\ell_i)_{i=1}^n$*, the random effects* $U_k, k \in \mathbb{Z}^d$ *and the design matrix* X *satisfy the assumptions* ***(W)**, **(D)** of Assumption* 1*, and* ***(V1)**, **(V2)** in Assumption* 3*. Moreover, in* (12)*, choose a* $B_n \to \infty$ *satisfying*

$$B_n^d \Psi \to 0, \textit{as } n \to \infty, \textit{ where } \Psi = \max\left\{\frac{\max_i \ell_i |\bar{\boldsymbol{X}}_{i\cdot}|}{\sqrt{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}}, \rho^\star(S^{-1})^2 \sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^4, \frac{\rho^\star(S)}{\sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}\right\}. \tag{13}$$

If $\kappa(S) \leqslant C$ for a constant C independent of n, and $\Delta_q < \infty$ for some $q > 4$, then, it holds that

$$\rho^\star(A^{-1}(\hat{A} - A)) = O_{\mathbb{P}}(\Theta_{B_n^d,q} + B_n^{-1} \sum_{m=1}^{B_n^d} \Theta_{m,q} + B_n^d \Psi) = o_{\mathbb{P}}(1),$$

where we recall $\Theta_{m,q}$ from Definition 1.

Remark 2 (Choice of the bandwidth). An optimal choice of the bandwidth will crucially depend on (i) the (unknown) dependency structure of the spatial random effects, and (ii) the (known) design matrix X, through Ψ. Often, especially in the context of time series, it is assumed that $\Theta_{m,q}$ decays polynomially, i.e. $\Theta_{m,q} = O(m^{-\gamma})$ for some $\gamma > 1$; see Wu (2005); Berkes et al. (2014); Karmakar & Wu (2020). This is also known as weakly dependent setting. Then, the optimal choice of B_n can be obtained as $B_n \asymp \Psi^{-\frac{1}{d+1}}$, which is independent of γ.

Corresponding to Theorem 2, we provide a result tackling the estimation of asymptotic variance for the random-design case.

THEOREM 4. *Consider the assumptions of Theorem 2, and further assume that*

$$\lambda_{\min}(\sum_i \ell_i \tilde{\boldsymbol{\mu}}_i \tilde{\boldsymbol{\mu}}_i^\top) \geq c_0 \sum_i \ell_i. \tag{14}$$

*Additionally, grant **(V')**. Then, in (12), as long as $B_n \to \infty$ satisfies*

$$B_n^d \boldsymbol{\Psi} = o(1), \ \textit{with } \boldsymbol{\Psi} = \max\left\{\frac{\max_i \ell_i^2}{\sum_i \ell_i^2}, \frac{\sum_i \ell_i}{\sum_i \ell_i^2}\right\}, \tag{15}$$

it holds that $\rho^\star(A^{-1}(\hat{A} - A)) = o_{\mathbb{P}}(1)$.

REFERENCES

ANSELIN, L. (2013). *Spatial econometrics: methods and models*, vol. 4. Springer Science & Business Media.

ANSELIN, L. & BERA, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics textbooks and monographs* **155**, 237–290.

ARNOLD, S. F. (1980). Asymptotic validity of F tests for the ordinary linear model and the multiple correlation model. *J. Amer. Statist. Assoc.* **75**, 890–894.

AUDRETSCH, D. B. & FELDMAN, M. P. (1996). Rd spillovers and the geography of innovation and production. *The American Economic Review* **86**, 630–640.

BALTAGI, B. H. (2021). *Econometric analysis of panel data*, vol. 4. Springer.

BALTAGI, B. H. & LI, D. (2006). Prediction in the panel data model with spatial correlation: the case of liquor. *Spatial Economic Analysis* **1**, 175–185.

BALTAGI, B. H., SONG, S. H., JUNG, B. C. & KOH, W. (2007). Testing for serial correlation, spatial autocorrelation and random effects using panel data. *J. Econometrics* **140**, 5–51.

BAYER, C., FRIZ, P. K., GASSIAT, P., MARTIN, J. & STEMPER, B. (2020). A regularity structure for rough volatility. *Mathematical Finance* **30**, 782–832.

BAYER, C., HARANG, F. A. & PIGATO, P. (2021). Log-modulated rough stochastic volatility models. *SIAM Journal on Financial Mathematics* **12**, 1257–1284.

BERKES, I., LIU, W. & WU, W. B. (2014). Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.* **42**, 794–817.

BICKEL, P. J. & FREEDMAN, D. A. (1983). Bootstrapping regression models with many parameters. In *A Festschrift for Erich L. Lehmann*, Wadsworth Statist./Probab. Ser. Wadsworth, Belmont, CA, pp. 28–48.

BUISHAND, T. A., DE HAAN, L. & ZHOU, C. (2008). On spatial extremes: with application to a rainfall problem. *Ann. Appl. Stat.* **2**, 624–642.

Burkholder, D. L. (1973). Distribution function inequalities for martingales. *The Annals of Probability* **1**, 19–42.

Case, A. C. (1991). Spatial patterns in household demand. *Econometrica: Journal of the Econometric Society* , 953–965.

Cliff, A. D. & Ord, J. K. (1981). *Spatial processes: models & applications*. Pion Ltd., London.

Cook, R. D. & Weisberg, S. (1982). *Residuals and influence in regression*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Corrado, L. & Fingleton, B. (2011). Multilevel modelling with spatial effects .

Dasgupta, S., Hossain, M. M., Huq, M. & Wheeler, D. (2018). Climate change, salinization and high-yield rice production in coastal bangladesh. *Agricultural and Resource Economics Review* **47**, 66–89.

Deb, S., Pourahmadi, M. & Wu, W. B. (2017). An asymptotic theory for spectral analysis of random fields. *Electronic Journal of Statistics* **11**, 4297 – 4322.

Dharshing, S. (2017). Household dynamics of technology adoption: A spatial econometric analysis of residential solar photovoltaic (pv) systems in germany. *Energy research & social science* **23**, 113–124.

Druska, V. & Horrace, W. C. (2004). Generalized moments estimation for spatial panel data: Indonesian rice farming. *American Journal of Agricultural Economics* **86**, 185–198.

El Machkouri, M., Volnỳ, D. & Wu, W. B. (2013). A central limit theorem for stationary random fields. *Stochastic Processes and their Applications* **123**, 1–14.

Hanson, G. H. (2005). Market potential, increasing returns and geographic concentration. *Journal of international economics* **67**, 1–24.

Heinrich, L. (1988). Asymptotic behaviour of an empirical nearest-neighbour distance function for stationary poisson cluster processes. *Mathematische Nachrichten* **136**, 131–148.

Helpman, E. (1998). The size of regions. topics in public economics. edited by d. pines, e. sadka, i. zilcha.

Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1**, 799–821.

Imran, M., Hayat, N., Saeed, M. A., Sattar, A. & Wahab, S. (2023). Spatial green growth in china: exploring the positive role of investment in the treatment of industrial pollution. *Environmental Science and Pollution Research* **30**, 10272–10285.

Jochmans, K. (2022). Heteroscedasticity-robust inference in linear regression models with many covariates. *J. Amer. Statist. Assoc.* **117**, 887–896.

Kapoor, M., Kelejian, H. H. & Prucha, I. R. (2007). Panel data models with spatially correlated error components. *J. Econometrics* **140**, 97–130.

Karmakar, S. & Wu, W.-B. (2020). Optimal gaussian approximation for multiple time series. *Statistica Sinica* **30**, 1399 – 1417.

Kelejian, H. H. & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *J. Econometrics* **157**, 53–67.

Lei, L. & Ding, P. (2021). Regression adjustment in completely randomized experiments with a diverging number of covariates. *Biometrika* **108**, 815–828.

LeSage, J. P. (1999). A spatial econometric examination of china's economic growth. *Geographic Information Sciences* **5**, 143–153.

Reece, A. S. & Hulse, G. K. (2020). Canadian cannabis consumption and patterns of congenital anomalies: an ecological geospatial analysis. *Journal of addiction medicine* **14**, e195–e210.

Ribatet, M. & Sedki, M. (2013). Extreme value copulas and max-stable processes. *Journal de la société française de statistique* **154**, 138–150.

Romão, J. & Nijkamp, P. (2019). Impacts of innovation, productivity and specialization on tourism competitiveness–a spatial econometric analysis on european regions. *Current Issues in Tourism* **22**, 1150–1169.

Stein, M. L. (1999). *Interpolation of spatial data*. Springer Series in Statistics. Springer-Verlag, New York. Some theory for Kriging.

Swamy, P. A. V. B. & Arora, S. S. (1972). The exact finite sample properties of the estimators of coefficients in the error components regression models. *Econometrica* **40**, 261–275.

Wu, P., Muzy, J.-F. & Bacry, E. (2022). From rough to multifractal volatility: The log s-fbm model. *Physica A: Statistical Mechanics and its Applications* **604**, 127919.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences* **102**, 14150–14154.

5. Appendix A

In this section we provide a proof of Theorem 1. The proof essentially follow the line of argument in El Machkouri et al. (2013); however we provide the detailed proof for the sake of completeness.

Proof of Theorem 1. Write the model (1) as

$$Y = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{U} + \boldsymbol{\varepsilon}, \tag{16}$$

where $\boldsymbol{U} = (U_{L_1}\mathbf{1}_{l_1}^\top : \ldots : U_{L_n}\mathbf{1}_{l_n}^\top)^\top$, and $\boldsymbol{\varepsilon} = (\varepsilon_{1,1} : \ldots, \varepsilon_{n,l_n})^\top$. Note that from (16), it can be written

$$\hat{\boldsymbol{\beta}}_{LS} = \boldsymbol{\beta} + S^{-1}\mathbb{X}^\top\boldsymbol{U} + S^{-1}\mathbb{X}^\top\boldsymbol{\varepsilon}. \tag{17}$$

Condition **(L)** ensures applicability of Lindeberg-Levy central limit theorem on $S^{-1}\mathbb{X}^\top\varepsilon$, and dictates that

$$(\sigma_\varepsilon^2 S^{-1})^{-1/2} S^{-1}\mathbb{X}^\top\boldsymbol{\varepsilon} \xrightarrow{d} Z_\varepsilon \stackrel{d}{=} N(0, I_p) \tag{18}$$

Next, we aim to show that

$$A^{-1/2}S^{-1}\mathbb{X}^\top\boldsymbol{U} \xrightarrow{d} Z \stackrel{d}{=} N(0, I_p). \tag{19}$$

Fix $m_n \in \mathbb{N}$ such that $m_n \to \infty$ and $m_n/|\Gamma_n| \to 0$, as $n \to \infty$. Subsequently we will omit the subscript n. Consider the σ field $\mathcal{F}_k^m := \sigma(e_{k-j} : |j| \leqslant m)$, and define the m-dependent random variables as

$$U_k^{(m)} := \mathbb{E}[U_k|\mathcal{F}_k^m], \ k \in \mathbb{Z}^d.$$

As a stepping stone towards proving (19), let us first show that

$$\mathbb{E}[VV^\top] \to 0 \text{ as } n \to \infty, \tag{20}$$

where $V = A^{-1/2}S^{-1}\mathbb{X}^\top(\boldsymbol{U} - \boldsymbol{U}^{(m)}) \in \mathbb{R}^p$. By Cauchy-Schwarz inequality, it is sufficient to show $\mathbb{E}[V^\top V] \to 0$ as $n \to \infty$, which is equivalent to

$$\mathbb{E}[(\boldsymbol{U} - \boldsymbol{U}^{(m)})^\top\mathbb{X}T^{-1}\mathbb{X}(\boldsymbol{U} - \boldsymbol{U}^{(m)})] \to 0 \text{ as } n \to \infty. \tag{21}$$

We proceed as follows.

$$\begin{aligned}\mathbb{E}[(\boldsymbol{U} - \boldsymbol{U}^{(m)})^\top S(\boldsymbol{U} - \boldsymbol{U}^{(m)})] &= \sum_{k=1}^p \left\| \sum_{i=1}^n (U_{L_i} - U_{L_i}^{(m)})\ell_i \bar{X}_{i\cdot,k} \right\|^2 \\ &\leqslant \left(\sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2 \right) \Delta_p^{(m)}\end{aligned}$$

where the inequality is due to Proposition 1 of El Machkouri et al. (2013). Therefore, the left hand side of (21) can be bounded as

$$\mathbb{E}[(\boldsymbol{U} - \boldsymbol{U}^{(m)})^\top\mathbb{X}T^{-1}\mathbb{X}(\boldsymbol{U} - \boldsymbol{U}^{(m)})] \leqslant \rho^\star(T^{-1})\left(\sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2 \right)\Delta_p^{(m)} \to 0, \tag{22}$$

by virtue of Condition **(W)** and Lemma 2 of El Machkouri et al. (2013). This shows (20), which, along with $\mathbb{E}[V] = \mathbf{0}$, immediately implies that

$$V \xrightarrow{p} \mathbf{0}.$$

Henceforth, we focus on a central limit theory $A^{-1/2}S^{-1}\mathbb{X}^\top \boldsymbol{U}^{(m)}$. We will apply Theorem 5. For ease of notations, let

$$M_n(i) = W_i U^m_{L_i},$$

where $W_i = A^{-1/2}S^{-1}\mathbb{X}^\top \mathsf{R}_i \in \mathbb{R}^p$ with $\mathsf{R}_i = (0 : \ldots : \mathbf{1}^\top_{\ell_i} : 0 \ldots : 0)^\top \in \mathbb{R}^{\sum_i \ell_i}$. Clearly, $\sum_{i=1}^n M_n(i) = A^{-1/2}S^{-1}\mathbb{X}^\top \boldsymbol{U}^{(m)}$, and similarly $\sum_{i=1}^n W_i U_{L_i} = A^{-1/2}S^{-1}\mathbb{X}^\top \boldsymbol{U}$. Therefore,

$$\begin{aligned}
&\|(\sum_{i=1}^n M_n(i))(\sum_{i=1}^n M_n(i))^\top - (\sum_{i=1}^n W_i U_{L_i})(\sum_{i=1}^n W_i U_{L_i})^\top\| \\
\leqslant &\|A^{-1/2}S^{-1}\mathbb{X}^\top(\boldsymbol{U} - \boldsymbol{U}^{(m)})\|(\|A^{-1/2}S^{-1}\mathbb{X}^\top \boldsymbol{U}\| + \|A^{-1/2}S^{-1}\mathbb{X}^\top \boldsymbol{U}^{(m)}\|) \qquad \to 0
\end{aligned}$$

similar to (22). In view of $\mathbb{E}[(\sum_{i=1}^n W_i U_{L_i})(\sum_{i=1}^n W_i U_{L_i})^\top] = I$, one obtains

$$\mathbb{E}[(\sum_{i=1}^n M_n(i))(\sum_{i=1}^n M_n(i))^\top] \to I, \text{ as } n \to \infty. \tag{23}$$

Moreover, Condition **(W)** dictates that

$$\begin{aligned}
\sum_{i\in\Gamma_n} \mathbb{E}[\|M_n(i)\|^2] &= \sum_{i\in\Gamma_n} \mathbb{E}[(U^m_{L_i})^2]\mathsf{R}_i^\top \mathbb{X} T^{-1}\mathbb{X}^\top \mathsf{R}_i \\
&\leqslant \operatorname{Var}(U_0)\rho^\star(T^{-1}) \sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2 < \infty.
\end{aligned} \tag{24}$$

For the Lindeberg Condition, fix $\varepsilon > 0$. Define

$$L_n(\varepsilon) = m_n^{2d} \sum_{i\in\Gamma_n} \mathbb{E}\Big(\|M_n(i)\|^2 \mathbb{I}\{\|M_n(i)\| \geqslant \varepsilon m_n^{-2d}\}\Big).$$

Define $\psi(x) := \mathbb{E}[U_0^2 \mathbb{I}\{|U_0| \geqslant x\}]$. Choose

$$m_n := \min\left\{ \left\lfloor \rho^\star(T^{-1})^{-\frac{1}{12d}} (\max_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2)^{-\frac{1}{6d}} \right\rfloor, \left\lfloor (\psi(\rho^\star(T^{-1})))^{-\frac{1}{4d}} \right\rfloor \right\}.$$

Clearly, (23) and (24) are valid for this choice. Proceeding as in Lemma 3 of El Machkouri et al. (2013), one can derive that $\lim_{n\to\infty} L_n(\varepsilon) = 0$. Therefore, an application of Theorem 5 implies a central limit theory for $U^m_{L_i}$, in turn showing (19).

Finally, in view of **(D)** and **(W)**, one can deduce $A^{-1/2}S^{-1/2} \to O$, which enables us to conclude the proof. □

Theorem 5 (Generalization of Theorem 2, Heinrich (1988)). *Let $(\Gamma_n)_{n\geqslant 1}$ be a sequence of finite subsets of $\mathbb{Z}^d$ with $|\Gamma_n| \to \infty$ as $n \to \infty$ and let $(m_n)_{n\geqslant 1}$ be a sequence of positive integers. For each $n \geqslant 1$, let $\{U_n(j), j \in \mathbb{Z}^d, U_n(j) \in \mathbb{R}^p\}$ be an m_n-dependent random field with $\mathbb{E}U_n(j) = 0$ for all j in $\mathbb{Z}^d$. Assume that $\mathbb{E}[(\sum_{j\in\Gamma_n} U_n(j))(\sum_{j\in\Gamma_n} U_n(j))^\top] \to \Sigma$ as $n \to \infty$ with $\rho^\star(\Sigma) > c > 0$. Further assume that*

$$\sum_{j\in\Gamma_n} \mathbb{E}[\|U_n(j)\|^2] \leqslant C, \tag{25}$$

and for any $\varepsilon > 0$ it holds that

$$\lim_{n\to\infty} L_n(\varepsilon) := m_n^{2d} \sum_{j\in\Gamma_n} \mathbb{E}\left(\|U_n(j)\|^2 \mathbb{I}_{\|U_n(j)\|\geqslant \varepsilon m_n^{-2d}}\right) = 0. \tag{26}$$

Then it can be deduced that $\sum_{j\in\Gamma_n} U_n(j) \xrightarrow{d} N(0,\Sigma)$.

Proof of Theorem 3. The main idea of the proof is to accurately estimate the matrix Σ_U. To that end, we begin with an oracle estimate. Consider the kernel K as in the statement of the theorem, and define

$$\tilde{\Sigma}_U := \begin{pmatrix} U_{L_1}^2 J_{\ell_1,\ell_1} & U_{L_1}U_{L_2}K(\frac{L_2-L_1}{B_n})J_{\ell_1,\ell_2} & \dots & U_{L_1}U_{L_n}K(\frac{L_n-L_1}{B_n})J_{\ell_1,\ell_n} \\ \vdots & \vdots & \dots & \vdots \\ U_{L_1}U_{L_n}K(\frac{L_n-L_1}{B_n})J_{\ell_1,\ell_n} & U_{L_2}U_{L_n}K(\frac{L_n-L_2}{B_n})J_{\ell_2,\ell_n} & \dots & U_{L_n}^2 J_{\ell_n,\ell_n} \end{pmatrix}. \tag{27}$$

With the help of $\tilde{\Sigma}_U$, define the oracle estimate $\tilde{A} = S^{-1}\mathbb{X}^\top \tilde{\Sigma}_U \mathbb{X} S^{-1}$. We first quantify the approximation error of $\tilde{\Sigma}_U$.

The key ingredient in our proof is the argument via projection. Let $\tau : \mathbb{Z} \to \mathbb{Z}^d$ be a bijection, and for $s \in \mathbb{Z}$, let $\mathcal{G}_s = \sigma\left(e_{\tau(l)} : l \leqslant s\right)$ be a σ-field. The corresponding projection operator is defined by $\mathcal{P}_s(\cdot) := \mathbb{E}[\cdot \mid \mathcal{G}_s] - \mathbb{E}[\cdot \mid \mathcal{G}_{s-1}]$. Such operators are also used in El Machkouri et al. (2013). For convenience, let $D = XS^{-1} \in \mathbb{R}^{\sum_i \ell_i \times p}$. Clearly,

$$\tilde{A} - \mathbb{E}[\tilde{A}] = D^\top \sum_{u\in\mathbb{Z}} \mathcal{P}_u(\tilde{\Sigma}_U) D = \sum_{k\in\mathcal{L}} K(k/B_n) \sum_{u\in\mathbb{Z}} \mathcal{P}_u(D^\top \tilde{\Sigma}_k D), \tag{28}$$

where, for each $k \in \mathcal{L}$,

$$\tilde{\Sigma}_k := \begin{pmatrix} U_{L_1}^2 J_{\ell_1,\ell_1}\mathbf{I}_{k=0} & U_{L_1}U_{L_2}J_{\ell_1,\ell_2}\mathbf{I}_{k=L_1-L_2} & \dots & U_{L_1}U_{L_n}J_{\ell_1,\ell_n}\mathbf{I}_{k=L_1-L_n} \\ \vdots & \vdots & \dots & \vdots \\ U_{L_1}U_{L_n}J_{\ell_1,\ell_n}\mathbf{I}_{k=L_n-L_1} & U_{L_2}U_{L_n}J_{\ell_2,\ell_n}\mathbf{I}_{k=L_n-L_2} & \dots & U_{L_n}^2 J_{\ell_n,\ell_n}\mathbf{I}_{k=0} \end{pmatrix}.$$

Note that

$$\rho^\star\Big(\sum_{u\in\mathbb{Z}} \mathcal{P}_u(D^\top \tilde{\Sigma}_k D)\Big) \leqslant \sum_{s,t}^{p} \sum_{u\in\mathbb{Z}} \mathcal{P}_u\Big(\sum_{(i,j)\in\mathcal{A}_k} w_i^{(s)} w_j^{(t)} U_{L_i} U_{L_j}\Big),$$

where $w_i^{(s)} = \mathsf{R}_i^\top D_{\cdot s}$, with $D_{\cdot s} \in \mathbb{R}^{\sum_i \ell_i}$ being the s-th column of D, $1 \leqslant s \leqslant p$. Subsequently, fix $1 \leqslant s,t \leqslant d$. For each $k \in \mathcal{L}$, $\mathcal{P}_u(\sum_{(i,j)\in\mathcal{A}_k} w_i^{(s)} w_j^{(t)} U_{L_i} U_{L_j})$ are martingale differences with respect to the upwards filtration $\mathcal{G}_s = \sigma(e_{\tau(l)} : l \leqslant s)$. We will follow a slightly different treatment for the two regimes $q \in (2,4]$, ands $q > 4$. For the latter one, since $\|\cdot\|_{q/4}$ is a norm, an application of Burkholder's inequality Burkholder (1973) yields,

$$
\begin{aligned}
&\left\|\sum_{u\in\mathbb{Z}}\sum_{(i,j)\in\mathcal{A}_k} w_i^{(s)}w_j^{(t)}\mathcal{P}_u\left(U_{L_i}U_{L_j}\right)\right\|_{q/2}^2 \\
&= \left\|\left|\sum_{u\in\mathbb{Z}}\sum_{(i,j)\in\mathcal{A}_k} w_i^{(s)}w_j^{(t)}\mathcal{P}_u\left(U_{L_i}U_{L_j}\right)\right|^2\right\|_{q/4} \\
&\leq \left\|\sum_{u\in\mathbb{Z}}\left|\sum_{(i,j)\in\mathcal{A}_k} w_i^{(s)}w_j^{(t)}\mathcal{P}_u\left(U_{L_i}U_{L_j}\right)\right|^2\right\|_{q/4} \\
&\leqslant \sum_{u\in\mathbb{Z}}\left\|\sum_{(i,j)\in\mathcal{A}_k} w_i^{(s)}w_j^{(t)}\mathcal{P}_u\left(U_{L_i}U_{L_j}\right)\right\|_{q/2}^2 \\
&\leqslant C_q\sum_{u\in\mathbb{Z}}\left(\sum_{(i,j)\in\mathcal{A}_k} w_i^{(s)}w_j^{(t)}\left\|\mathcal{P}_u\left(U_{L_i}U_{L_j}\right)\right\|_{q/2}\right)^2 \\
&\leqslant C_q\left\|U_0\right\|_q^2\sum_{u\in\mathbb{Z}}\left(\sum_{(i,j)\in\mathcal{A}_k} w_i^{(s)}w_j^{(t)}\left(\delta_{\ell_i-\tau(u),q}+\delta_{L_j-\tau(u),q}\right)\right)^2 \\
&\leqslant 2C_q\left\|U_0\right\|_q^2\sum_{u\in\mathbb{Z}}\left[\left(\sum_i w_i^{(s)}w_{i_k}^{(t)}\delta_{\ell_i-\tau(u),q}\right)^2+\left(\sum_i w_i^{(s)}w_{i_k}^{(t)}\delta_{L_{i_k}-\tau(u),q}\right)^2\right]. \qquad (29)
\end{aligned}
$$

Now, in light of Cauchy-Schwarz inequality,

$$
\begin{aligned}
\sum_{u\in\mathbb{Z}}\left(\sum_i w_i^{(s)}w_{i_k}^{(t)}\delta_{\ell_i-\tau(u),q}\right)^2 &\leqslant \max_i w_i^{(s)^2}\sum_{u\in\mathbb{Z}}\left(\sum_i w_i^{(t)}\delta_{\ell_i-\tau(u),q}\right)^2 \\
&\leqslant \max_i w_i^{(s)^2}\sum_{u\in\mathbb{Z}}\left(\sum_i w_i^{(t)^2}\delta_{\ell_i-\tau(u),q}\right)\left(\sum_i \delta_{\ell_i-\tau(u),q}\right) \\
&\leqslant \max_i w_i^{(s)^2}\Delta_q^2\sum_{i\in\Gamma_n} w_i^{(t)^2}.
\end{aligned}
$$

Observing that $\sum_{k\in\mathcal{L}}K\left(k/B_n\right)=O\left(B_n^d\right)$, for $q>4$, from (28) we obtain

$$
\begin{aligned}
\left\|\rho^\star(\tilde{A}-\mathbb{E}[\tilde{A}])\right\|_{q/2} &\leqslant C_q B_n^d\Delta_q^2\left(\sum_{s=1}^p\max_i w_i^{(s)}\right)\left(\sum_{t=1}^p\sqrt{\sum_{i=1}^{|\Gamma_n|}w_i^{(t)^2}}\right) \\
&\leqslant C_{p,q}\Delta_q^2 B_n^d\left(\sum_{s=1}^p\max_i D_{\cdot s}^\top \mathsf{R}_i\right)\left(\sum_{t=1}^p\sqrt{D_{\cdot t}^\top\sum_{i=1}^{|\Gamma_n|}\mathsf{R}_i\mathsf{R}_i^T\ D_{\cdot t}}\right),
\end{aligned}
$$

which immediately yields, in light of **(W)** and some elementary matrix manipulations,

$$\|\rho^\star(A^{-1}(\tilde{A}-\mathbb{E}[\tilde{A}]))\|_{q/2} \leqslant C_{p,q}\Delta_q^2 B_n^d \kappa(S)^2 \frac{\max_i \ell_i |\bar{\boldsymbol{X}}_{i\cdot}|}{\sqrt{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}}. \tag{30}$$

For $q \in (2,4]$, the above argument fails since $\|\cdot\|_{q/4}$ is no longer a norm. Therefore, we will follow a slightly different argument which uses $(|a_1|+|a_2|+\ldots)^{q/4} \leq |a_1|^{q/4}+|a_2|^{q/4}+\ldots$ The following series of inequalities hold :

$$\begin{aligned}
&\left\|\sum_{u\in\mathbb{Z}}\sum_{(i,j)\in\mathcal{A}_k} w_i^{(s)}w_j^{(t)}\mathcal{P}_u\left(U_{L_i}U_{L_j}\right)\right\|_{q/2}^{q/2} \\
&\leq \left\|\sum_{u\in\mathbb{Z}}\left|\sum_{(i,j)\in\mathcal{A}_k} w_i^{(s)}w_j^{(t)}\mathcal{P}_u\left(U_{L_i}U_{L_j}\right)\right|^2\right\|_{q/4}^{q/4} \\
&\leqslant \sum_{u\in\mathbb{Z}}\left\|\sum_{(i,j)\in\mathcal{A}_k} w_i^{(s)}w_j^{(t)}\mathcal{P}_u\left(U_{L_i}U_{L_j}\right)\right\|_{q/2}^{q/2} \\
&\leqslant C_q\,\|U_0\|_q^{q/2}\sum_{u\in\mathbb{Z}}\left[\left(\sum_i w_i^{(s)}w_{i_k}^{(t)}\delta_{\ell_i-\tau(u),q}\right)^{q/2}+\left(\sum_i w_i^{(s)}w_{i_k}^{(t)}\delta_{L_{i_k}-\tau(u),q}\right)^{q/2}\right],
\end{aligned}$$

where the last inequality follows similar to (29). Now, applying Holder's inequality, one obtains

$$\begin{aligned}
\sum_{u\in\mathbb{Z}}\left(\sum_i w_i^{(s)}w_{i_k}^{(t)}\delta_{\ell_i-\tau(u),q}\right)^{q/2} &\leqslant \max_i w_i^{(s)^{q/2}}\sum_{u\in\mathbb{Z}}\left(\sum_i w_i^{(s)}\delta_{\ell_i-\tau(u),q}\right)^{q/2} \\
&\leqslant \max_i w_i^{(s)^{q/2}}\sum_{u\in\mathbb{Z}}\left(\sum_i w_i^{(s)^{q/2}}\delta_{\ell_i-\tau(u),q}\right)\left(\delta_{\ell_i-\tau(u),q}\right)^{q/2-1} \\
&\leqslant \max_i w_i^{(s)^{q/2}}\Delta_q^{q/2}\sum_{i=1}^{|\Gamma_n|}\left(w_i^{(s)^{q/2}}\right)
\end{aligned}$$

Therefore, for a general $q>2$, (30) can be generalized to

$$\|\rho^\star(A^{-1}(\tilde{A}-\mathbb{E}[\tilde{A}]))\|_{q/2} \leqslant C_{p,q}\Delta_q^2 B_n^d \kappa(S)^2 \frac{\max_i \ell_i |\bar{\boldsymbol{X}}_{i\cdot}|\cdot(\sum_i(\ell_i|\bar{\boldsymbol{X}}_{i\cdot}|)^{q'/2})^{2/q'}}{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2},\ q'=q\wedge 4. \tag{31}$$

Now, we focus on estimating the bias. Let $R_n = \rho^\star(\mathbb{E}[\tilde{A}]-A) \leqslant \rho^\star(S^{-1})^2\rho^\star(X^\top(\mathbb{E}[\tilde{\Sigma}_U]-\Sigma_U)X)$. Let us define the matrix

$$E_k := \begin{pmatrix} J_{l_1,l_1}\mathbf{I}_{k=0} & J_{l_1,l_2}\mathbf{I}_{k=L_1-L_2} & \ldots & J_{l_1,l_n}\mathbf{I}_{k=L_1-L_n} \\ \vdots & \vdots & \ldots & \vdots \\ J_{l_1,l_n}\mathbf{I}_{k=L_n-L_1} & J_{l_2,l_n}\mathbf{I}_{k=L_n-L_2} & \ldots & J_{l_n,l_n}\mathbf{I}_{k=0}\end{pmatrix}, k\in\mathbb{Z}^d.$$

This matrix is useful, since $\mathbb{E}[\tilde{\Sigma}_k] = \gamma_k E_k$. On the other hand, $\mathbb{E}[\tilde{\Sigma}_U] - \Sigma_U = \sum_{k\in\mathbb{Z}^d}(K(k/B_n)-1)\mathbb{E}[\tilde{\Sigma}_k]$. Moreover, a routine application of Cauchy-Schwarz inequality dictates that $\max_k \rho^\star(\mathbb{X}^T E_k \mathbb{X}) \leqslant \sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2$. Therefore, another application of **(W)** yields,

$$\rho^\star(A^{-1}(\mathbb{E}[\tilde{A}] - A)) \leqslant C_{p,q}\kappa(S)^2 \Big|\sum_{k\in\mathbb{Z}^d}(K(k/B_n)-1)\gamma_k\Big|. \tag{32}$$

Clearly,

$$\begin{aligned}
\left|\sum_{k\in Z^d}\gamma_k\left(K\left(k/B_n\right)-1\right)\right| &\leqslant \left|\sum_{k\notin[-B_n,B_n]^d}\gamma_k\right| + \sum_{k\in[-B_n,B_n]^d}|\gamma_k|\,|K\left(k/B_n\right)-1| \\
&\leqslant \sum_{k\notin[-B_n,B_n]^d}\sum_{s\in\mathbb{Z}^d}\delta_{s,q}\delta_{s+k,q} + C\sum_{k\in[-B_n,B_n]^d}\frac{1}{B_n^d}|k|\sum_{s\in\mathbb{Z}^d}\delta_{s,q}\delta_{s+k,q} \\
&\leqslant \Delta_q\Theta_{B_n^d,q} + C\frac{1}{B_n}\left(\sum_{m=1}^{B_n^d}\Theta_{m,p}\right)
\end{aligned}$$

where the last inequality from q-stability and the definition of $\Theta_{m,p}$, and the second term comes from $\sup_x |K'(x)| \leqslant C$. Combining this with (31) and (32), and noting that $\Delta_q < \infty$ due to q-stability, we finally arrive at

$$\rho^\star(A^{-1}(\tilde{A}-A)) \leqslant C_{p,q}\kappa(S)^2\left(\Theta_{B_n^d,q} + \frac{1}{B_n}\sum_{m=1}^{B_n^d}\Theta_{m,p} + B_n^d\frac{\max_i \ell_i|\bar{\boldsymbol{X}}_{i\cdot}|\cdot(\sum_i(\ell_i|\bar{\boldsymbol{X}}_{i\cdot}|)^{q'/2})^{2/q'}}{\sum_i \ell_i^2|\bar{\boldsymbol{X}}_{i\cdot}|^2}\right),\ q' = q\wedge 4. \tag{33}$$

Now we will show that $\tilde{A}$ is well-approximated by $\hat{A}$. Observe that $R_{\ell_i} = V_{\ell_i} + S_{\ell_i}$, where $V_{\ell_i} := U_{L_i} - \bar{\boldsymbol{X}}_{i\cdot}^\top S^{-1}\mathbb{X}^\top U$, and $S_{\ell_i} := \bar{\varepsilon}_{i\cdot} - \bar{\boldsymbol{X}}_{i\cdot}^\top S^{-1}\mathbb{X}^\top\varepsilon$. For $1 \leqslant i,j \leqslant |\Gamma_n|$, define the matrices

$$\begin{aligned}
B_{ij}^{(1)} &= V_{\ell_i}V_{L_j}K(\frac{L_i - L_j}{B_n})J_{\ell_i,\ell_j} \\
B_{ij}^{(2)} &= V_{\ell_i}S_{L_j}K(\frac{L_i - L_j}{B_n})J_{\ell_i,\ell_j} \\
B_{ij}^{(3)} &= S_{\ell_i}S_{L_j}K(\frac{L_i - L_j}{B_n})J_{\ell_i,\ell_j}.
\end{aligned}$$

We will use this matrices to define a set of intermediate block-partitioned matrices to go from $\hat{A}$ to $\tilde{A}$. Let

$$\Sigma^{(k)} = \begin{pmatrix} B_{(11)}^{(k)} & B_{(12)}^{(k)} & \cdots & B_{(1|\Gamma_n|)}^{(k)} \\ \vdots & \vdots & \cdots & \vdots \\ B_{(|\Gamma_n|1)}^{(k)} & B_{(|\Gamma_n|2)}^{(k)} & \cdots & B_{(|\Gamma_n||\Gamma_n|)}^{(k)} \end{pmatrix} \in \mathbb{R}^{\sum_i \ell_i \times \sum_i \ell_i},\ k = 1,2,3,$$

be the corresponding intermediate Σ matrices. Obviously, $\hat{\Sigma}_U - \tilde{\Sigma}_U = (\Sigma^{(1)} - \tilde{\Sigma}_U) + \Sigma^{(2)} + \Sigma^{(2)^\top} + \Sigma^{(3)}$. Therefore, we start off by showing that $\rho^\star(A^{-1}S^{-1}\mathbb{X}^\top\Sigma^{(2)}\mathbb{X}S^{-1})$ is small. To break $\Sigma^{(2)}$ further, note that

$$\Sigma^{(2)} = \Sigma_{T^{(2)}} - \Sigma_{W^{(2)}} + \Sigma_{H^{(2)}} + \Sigma_{Z^{(2)}},$$

where

$$V_{\ell_i}S_{\ell_k} = U_{L_i}\bar{\varepsilon}_k \cdot -\bar{\varepsilon}_{k\cdot}\bar{\boldsymbol{X}}_{i\cdot}^\top D^\top \boldsymbol{U} - U_{L_i}\bar{\boldsymbol{X}}_{k\cdot}^\top D^\top \varepsilon + U^\top D\bar{\boldsymbol{X}}_{i\cdot}\bar{\boldsymbol{X}}_{k\cdot}D^\top \varepsilon$$
$$= T_{ik}^{(2)} - W_{ik}^{(2)} + H_{ik}^{(2)} + Z_{ik}^{(2)},$$

and the $\mathbb{R}^{n\times n}$ matrices $T^{(2)}, W^{(2)}, H^{(2)}$ and $Z^{(2)}$ are formed with entries $T_{ij}^{(2)}$, $W_{ij}^{(2)}$, $H_{ij}^{(2)}$ and $Z_{ij}^{(2)}$ respectively; $\Sigma_{T^{(2)}} = J(T^{(2)} \circ \mathcal{K})J^T \in \mathbb{R}^{\sum_i \ell_i \times \sum_i \ell_i}$, and rest of the matrices are similarly defined. Here $\mathcal{K}_{ij} = K(\frac{L_i - L_j}{B_n})$ defines the matrix $\mathcal{K}$, and $\circ$ denotes the Hadamard product. Note that, invoking **(W)**,

$$\begin{aligned}
&\|\rho^\star(A^{-1}S^{-1}\mathbb{X}^\top \Sigma_{T^{(2)}}\mathbb{X}S^{-1})\|_{q/2} \\
\leqslant& C_{p,q}\frac{\kappa(S)}{\sum_i \ell_i|\bar{\boldsymbol{X}}_{i\cdot}|^2}\sum_{s,t}^p \sum_k K(k/B_n)\|\sum_{(i,j)\in\mathcal{A}_k}\ell_i l_j U_{L_i}\bar{\varepsilon}_j \bar{X}_{i\cdot,s}\bar{X}_{j\cdot,t}\|_{q/2} \\
\leqslant& C_{p,q}\Delta_q B_n^d \kappa(S)\frac{\max_i \ell_i|\bar{\boldsymbol{X}}_{i\cdot}|}{\sum_i \ell_i^2|\bar{\boldsymbol{X}}_{i\cdot}|^2},
\end{aligned} \tag{34}$$

where the final inequality employs a derivation similar to (31). For the analysis of $\Sigma_W^{(2)}$, $\Sigma_H^{(2)}$ and $\Sigma_Z^{(2)}$, we will employ Hölder's inequality $\|X^\top Y\|_{q/2} \lesssim_{d,q} \|X\|_q\|Y\|_q$. Clearly,

$$\begin{aligned}
&\|\rho^\star(A^{-1}S^{-1}X^\top \Sigma_{W^{(2)}}\mathbb{X}S^{-1})\|_{q/2} \\
\leqslant& C_{p,q}\frac{\kappa(S)}{\sum_i \ell_i|\bar{\boldsymbol{X}}_{i\cdot}|^2}\sum_k K(k/B_n)\|\rho^\star(\sum_{(i,j)\in\mathcal{A}_k}\ell_i l_j(\bar{\boldsymbol{X}}_{i\cdot}^\top D^\top \boldsymbol{U})\bar{\varepsilon}_{j\cdot}|\bar{\boldsymbol{X}}_{i\cdot}|\cdot|\bar{\boldsymbol{X}}_{j\cdot}|)\|_{q/2} \\
\leqslant& C_{p,q}B_n^d\kappa(S)\|D^\top \boldsymbol{U}\|_q\frac{\sqrt{\sum_i \ell_i^3|\bar{\boldsymbol{X}}_{i\cdot}|^6}}{\sum_i \ell_i^2|\bar{\boldsymbol{X}}_{i\cdot}|^2} \\
\leqslant& C_{p,q}\Delta_q B_n^d\kappa(S)\rho^\star(S^{-1})\sqrt{\frac{\sum_i \ell_i^3|\bar{\boldsymbol{X}}_{i\cdot}|^6}{\sum_i \ell_i^2|\bar{\boldsymbol{X}}_{i\cdot}|^2}} \\
\leqslant& C_{p,q}\Delta_q B_n^d\kappa(S)\rho^\star(S^{-1})\sqrt{\sum_i \ell_i|\bar{\boldsymbol{X}}_{i\cdot}|^4}\frac{\max_i \ell_i|\bar{\boldsymbol{X}}_{i\cdot}|}{\sqrt{\sum_i \ell_i^2|\bar{\boldsymbol{X}}_{i\cdot}|^2}},
\end{aligned} \tag{35}$$

where in the last line we have used $D^\top \boldsymbol{U} = S^{-1}\sum_i \ell_i U_{L_i}\bar{\boldsymbol{X}}_{i\cdot}$, along with Proposition 1 of El Machkouri et al. (2013). Moreover, an exact same treatment yields

$$\|\rho^\star(A^{-1}S^{-1}X^\top \Sigma_{H^{(2)}}\mathbb{X}S^{-1})\|_{q/2} \leqslant C_{p,q}\Delta_q B_n^d\kappa(S)\rho^\star(S^{-1})\sqrt{\sum_i \ell_i^2|\bar{\boldsymbol{X}}_{i\cdot}|^4}\frac{\max_i \ell_i|\bar{\boldsymbol{X}}_{i\cdot}|}{\sqrt{\sum_i \ell_i^2|\bar{\boldsymbol{X}}_{i\cdot}|^2}}, \tag{36}$$

$$\|\rho^\star(A^{-1}S^{-1}X^\top \Sigma_{Z^{(2)}}\mathbb{X}S^{-1})\|_{q/2} \leqslant C_{p,q}\Delta_q B_n^d\kappa(S)\rho^\star(S^{-1})^2\sum_i \ell_i^2|\bar{\boldsymbol{X}}_{i\cdot}|^4 \tag{37}$$

Moving on, we will show that $\rho^\star(A^{-1}S^{-1}X^\top \Sigma^{(3)}\mathbb{X}S^{-1})$ is small. Write

$$S_{\ell_i}S_{L_j} = \bar{\varepsilon}_{i\cdot}\bar{\varepsilon}_{j\cdot} - \bar{\varepsilon}_{i\cdot}\bar{\boldsymbol{X}}_{j\cdot}^\top D^\top \varepsilon - \bar{\varepsilon}_{j\cdot}\bar{\boldsymbol{X}}_{i\cdot}^\top D^\top \varepsilon + \bar{\boldsymbol{X}}_{i\cdot}^\top(D^\top \varepsilon\varepsilon^\top D)\bar{\boldsymbol{X}}_{j\cdot}$$
$$= T_{ij}^{(3)} - W_{ij}^{(3)} - W_{ji}^{(3)} + Z_{ij}^{(3)}.$$

We form the matrices $\Sigma_{T^{(3)}}, \Sigma_{W^{(3)}}$ and $\Sigma_{Z^{(3)}}$ similar to $\Sigma_{T^{(2)}}, \Sigma_{W^{(2)}}$ and $\Sigma_{Z^{(3)}}$ respectively. We immediately have, following (35),

$$\|\rho^\star(A^{-1}S^{-1}X^\top \Sigma_{T^{(3)}}\mathbb{X}S^{-1})\|_{q/2} \leqslant C_{p,q}\Delta_q B_n^d \kappa(S)\frac{\sum_i \ell_i |\bar{\boldsymbol{X}}_{i\cdot}|^2}{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2} \leqslant C_{p,q}\Delta_q B_n^d \kappa(S)\frac{\rho^\star(S)}{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2} \tag{38}$$

$$\begin{aligned}\|\rho^\star(A^{-1}S^{-1}X^\top \Sigma_{W^{(3)}}\mathbb{X}S^{-1})\|_{q/2} &\leqslant C_{p,q}\Delta_q B_n^d \kappa(S)\|D^\top \varepsilon\|_q \frac{\sqrt{\sum_i \ell_i^3 |\bar{\boldsymbol{X}}_{i\cdot}|^6}}{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}\\ &\leqslant C_{p,q}\Delta_q B_n^d \kappa(S)\rho^\star(S^{-1})\sqrt{\sum_i \ell_i |\bar{\boldsymbol{X}}_{i\cdot}|^4}\frac{\max_i \ell_i |\bar{\boldsymbol{X}}_{i\cdot}|}{\sqrt{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}}\end{aligned} \tag{39}$$

$$\|\rho^\star(A^{-1}S^{-1}X^\top \Sigma_{Z^{(3)}}\mathbb{X}S^{-1})\|_{q/2} \leqslant C_{p,q}\Delta_q B_n^d \kappa(S)\rho^\star(S^{-1})^2 \sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^4. \tag{40}$$

Here, (38) employs Cauchy-Schwarz inequality. Finally, we will prove an upper bound on $\rho^\star(A^{-1}D^\top(\Sigma^{(1)} - \tilde{\Sigma}_U)D)$, where we recall $\tilde{\Sigma}_U$ from (27). Writing

$$\begin{aligned}V_{\ell_i}V_{L_j} =& U_{L_i}U_{L_j} - U_{L_i}\bar{\boldsymbol{X}}_{j\cdot}^\top D^\top \boldsymbol{U} - U_{L_j}\bar{\boldsymbol{X}}_{i\cdot}^\top D^\top \boldsymbol{U} + \bar{\boldsymbol{X}}_{i\cdot}^\top (D^\top \boldsymbol{U}\boldsymbol{U} D^\top)\bar{\boldsymbol{X}}_{j\cdot}\\ :=& U_{L_i}U_{L_j} - W_{ij}^{(1)} - W_{ji}^{(1)} + Z_{ij}^{(1)},\end{aligned}$$

where we define $W^{(1)}$, $Z^{(1)}$, $\Sigma_{W^{(1)}}$ and $\Sigma_{Z^{(1)}}$ as before. From the definition of $\tilde{\Sigma}_U$, it is enough to upper bound the quantities $\rho^\star(A^{-1}D^\top \Sigma_{W^{(1)}}D)$ and $\rho^\star(A^{-1}D^\top \Sigma_{Z^{(1)}}D)$. In fact, in light of the definitions of the matrices $W^{(3)}$ and $Z^{(3)}$, we recover the bounds exactly same as (39) and (40):

$$\|\rho^\star(A^{-1}S^{-1}X^\top \Sigma_{W^{(1)}}\mathbb{X}S^{-1})\|_{q/2} \leqslant C_{p,q}\Delta_q B_n^d \kappa(S)\rho^\star(S^{-1})\sqrt{\sum_i \ell_i |\bar{\boldsymbol{X}}_{i\cdot}|^4}\frac{\max_i \ell_i |\bar{\boldsymbol{X}}_{i\cdot}|}{\sqrt{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}} \tag{41}$$

$$\|\rho^\star(A^{-1}S^{-1}X^\top \Sigma_{Z^{(1)}}\mathbb{X}S^{-1})\|_{q/2} \leqslant C_{p,q}\Delta_q B_n^d \kappa(S)\rho^\star(S^{-1})^2 \sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^4. \tag{42}$$

In light of $\kappa(S) \wedge \Delta_q = O(1)$, and (13), the proof concludes itself by applying **(D)** on (38), **(V2)** on (37), (42) and (40), and **(V1)** on (33), (34)-(36), (39) and (41). □

Proof of Theorem 2. The proof proceeds by establishing corresponding stochastic versions of Assumptions **(L)**, **(W)** and **(D)**, before invoking the Theorem 1 to infer the weak convergence. This argument is organized over the following three steps.

5.1. *Stochastic analogue of* ***(L)***

Write $\mathbb{X} = (\mathbf{1}_{\sum_i \ell_i} : Z)$, where $Z := (\mathbf{z}_1 : \cdots \mathbf{z}_{\sum_i \ell_i})^\top \in \mathbb{R}^{\sum l_i \times (p-1)}$ is a matrix with i.i.d. $N(\mu, \sigma^2)$ entries, and $\mathbf{z}_k = (Z_{k,2} : \cdots : Z_{k,p})^\top$. Let $h_{kk} = (\mathbb{X}S^{-1}\mathbb{X}^\top)_{kk}$. From equation (2.1.8) of Cook & Weisberg (1982), it follows that

$$h_{kk} = \frac{1}{\sum_i \ell_i} + z_k(\mathcal{Z}^\top \mathcal{Z})^{-1}z_k, k \in [\sum_i \ell_i], \tag{43}$$

where

$$z_k = \mathbf{z}_k - (\sum_i \ell_i)^{-1} \sum_k \mathbf{z}_k \in \mathbb{R}^{p-1}, k \in [\sum_i \ell_i], \text{ and } \mathcal{Z} = (z_1 : \cdots : z_{\sum_i \ell_i})^\top \in \mathbb{R}^{\sum_i \ell_i \times (p-1)}.$$

It is trivial to see that $(\sum_i \ell_i)^{-1} \mathcal{Z}^T \mathcal{Z} \xrightarrow{\mathbb{P}} \sigma^2 I_{p-1}$. Therefore,

$$\max_k z_k (\mathcal{Z}^\top \mathcal{Z})^{-1} z_k \leq \|(\mathcal{Z}^\top \mathcal{Z})^{-1}\|_{\text{op}} \max_k |z_k|^2 = O_{\mathbb{P}}(\frac{\log \sum_i \ell_i}{\sum_i \ell_i}),$$

which shows

$$\max_{1 \leq k \leq \sum_i \ell_i} (\mathbb{X} S^{-1} \mathbb{X}^\top)_{kk}) \xrightarrow{\mathbb{P}} 0 \text{ as } n \to \infty. \tag{44}$$

5.2. *Stochastic analogue of* ***(W)***

Let $\mathcal{M} = \sum_{i,k} \ell_i \ell_k \gamma_{|L_i - L_k|} \tilde{\boldsymbol{\mu}}_i \tilde{\boldsymbol{\mu}}_k^\top$. Let $\boldsymbol{\mu}_i = (\mu_{i2}, \dots, \mu_{ip})^\top$. Observe that

$$T = \sum_{i,k=1}^n \ell_i \ell_k \gamma_{|L_i - L_k|} \bar{\boldsymbol{X}}_{i\cdot} \bar{\boldsymbol{X}}_{k\cdot}^\top = \mathcal{M} + O_{\mathbb{P}}(\sum_{i,k=1}^n \sqrt{\ell_i} \ell_k \gamma_{|L_i - L_k|}), \tag{45}$$

where the $O_{\mathbb{P}}$ acts entry-wise. Let $G = \sum_{i,k} \ell_i \ell_k \gamma_{|L_i - L_k|}$. In view of **(W')** it holds that

$$\frac{\sum_{i,k=1}^n \sqrt{\ell_i} \ell_k \gamma_{|L_i - L_k|}}{\sum_{i,k=1}^n \ell_i \ell_k \gamma_{|L_i - L_k|}} \leq c\gamma_0 \frac{\sum_i \ell_i^{3/2}}{\sum_i \ell_i^2} = o(1).$$

Therefore,

$$G^{-1} |\lambda_{\min}(T) - \lambda_{\min}(M)| = o_{\mathbb{P}}(1),$$

which, in light of (11), ensures that $G^{-1} \lambda_{\min}(T) \geq c_0 (1 + o_{\mathbb{P}}(1))$. Consequently,

$$\rho^\star(GT^{-1}) = G(\lambda_{\min}(T))^{-1} \leq c_0^{-1}(1 + o_{\mathbb{P}}(1)). \tag{46}$$

On the other hand,

$$\sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2 = \sum_i \ell_i^2 + \sum_{k=2}^p \sum_i \ell_i^2 \bar{X}_{ik}^2 = \sum_i \ell_i^2 (1 + \sum_{k=2}^p \mu_{ik}^2) + O_{\mathbb{P}}\left(\sum_i \ell_i^{3/2} \sum_{k=2}^p \mu_{ik}\right) + O_{\mathbb{P}}(\sum_i \ell_i). \tag{47}$$

Recall **(D')**. In fact, $\frac{\sum_i l_i}{\sum_i l_i^2} = o(1)$ also implies that $\frac{\sum_i l_i^{3/2}}{\sum_i l_i^2} = o(1)$ via Cauchy-Schwarz inequality. Therefore, there exists a constant M such that

$$\frac{\sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}{\sum_i \ell_i^2} \leq M(1 + o_{\mathbb{P}}(1)),$$

which, in view of **(W')** and (46), implies that

$$\rho^\star(T^{-1}) \sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2 \leq \frac{M}{cc_0}(1 + o_{\mathbb{P}}(1)). \tag{48}$$

5.3. *Stochastic analogue of* **(D)**

Finally we shift attention to **(D)**. Clearly,

$$\rho^\star(S) \lesssim \sum_{i=1}^{n}\sum_{j=1}^{\ell_i} |\boldsymbol{X}_{ij}|^2 = \sum_i \ell_i + \sum_{k=2}^{p}\sum_{i=1}^{n}\sum_{j=1}^{\ell_i} Z_{ij,k}^2 = O_{\mathbb{P}}(\sum_i \ell_i). \tag{49}$$

On the other hand, from (47),

$$\frac{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}{\sum_i \ell_i^2} \geq 1 + o_{\mathbb{P}}(1), \tag{50}$$

and therefore, $(\sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2)^{-1} \leq (\sum_i \ell_i^2)^{-1}(1 + o_{\mathbb{P}}(1))$. Invoking **(D')**, (47) and (49) jointly provide that

$$\frac{\rho^\star(S)}{\sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2} = o_{\mathbb{P}}(1). \tag{51}$$

5.4. *Combining the pieces together*

In the following, we re-label all the relevant quantities with a subscript to emphasize its explicit dependence on the number of locations based on which the corresponding quantity is being computed. For example, if the design matrix $\mathbb{X}$ is computed based on some m locations and accompanying sequence of replications $\{l_i(m)\}_{i=1}^m$, the corresponding maximum leverage $\max_{1\leq k\leq \sum_i \ell_i(m)} (\mathbb{X}S^{-1}\mathbb{X}^\top)_{kk}$ may be denoted as $\big(\max_{1\leq k\leq \sum_i \ell_i} (\mathbb{X}S^{-1}\mathbb{X}^\top)_{kk}\big)_m$. Moreover, $\hat{\boldsymbol{\beta}}_{LS}(m)$ denote the corresponding least square estimate.

Now we hark back to the proof of Theorem 2. Observe that, to obtain (9), it is enough to show, given any increasing sequence of positive integers $\{t_k\}_{k\in\mathbb{N}}$, there exists a further subsequence $\{t_{n_k}\}$, $n_k \in \mathbb{N}$, such that

$$A_{t_{n_k}}^{-1/2}(\hat{\boldsymbol{\beta}}_{LS}(t_{n_k}) - \boldsymbol{\beta}) \xrightarrow{d} N(0, I_p). \tag{52}$$

We proceed as follows. From (44) and (51), there exists a sub-subsequence $\{t_{n_k^{(1)}}\} \subseteq \{t_n\}$ such that

$$\big(\max_k (\mathbb{X}S^{-1}\mathbb{X}^\top)_{kk}\big)_{t_{n_k^{(1)}}} \xrightarrow{a.s.} 0, \text{ and } \left(\frac{\rho^\star(S)}{\sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}\right)_{t_{n_k^{(1)}}} \xrightarrow{a.s.} 0. \tag{53}$$

Moreover, (48), there exists a subsequence $\{t_{n_k^{(2)}}\} \subseteq \{t_{n_k^{(1)}}\}$ such that

$$\left(\rho^\star(T^{-1}) \sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2\right)_{t_{n_k^{(2)}}} \leq 2M(cc_0)^{-1} \text{ almost surely.} \tag{54}$$

We choose $n_k = n_k^{(2)}$. In view of (53)-(54), an immediate application of Theorem 1 yields (52), completing our proof. □

Proof of Theorem 4. We proceed similar to Theorem 2, establishing stochastic counterparts of **(V1)** and **(V2)** before invoking Theorem 3.

5.5. *Stochastic analogue of* ***(V1)***

Observe that similar to (47), it holds that

$$\max_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2 \leq \max_i \ell_i^2 + \sum_{k=2}^{p} \max_i \left(\ell_i^2 \mu_{ik}^2 + \ell_i^{3/2} O_{\mathbb{P}}(\mu_{ik})\right) = O_{\mathbb{P}}(\max_i \ell_i^2), \tag{55}$$

where the $O_{\mathbb{P}}$ assertion follows from $\sup_i |\boldsymbol{\mu}_i| = O(1)$. Therefore, from (50), we have

$$\frac{\max_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2} = O_{\mathbb{P}}(\frac{\max_i \ell_i^2}{\sum_i \ell_i^2}) = o_{\mathbb{P}}(1). \tag{56}$$

5.6. *Stochastic analogue of* ***(V2)***

Again, similar to (47) it follows that $\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^4 = O_{\mathbb{P}}(\sum_i \ell_i^2)$. On the other hand, observe that $\sup_i \sigma_i = O(1)$ instructs

$$S = \sum_i \ell_i \tilde{\boldsymbol{\mu}}_i \tilde{\boldsymbol{\mu}}_i^\top + O_{\mathbb{P}}(\sqrt{\sum_i \ell_i}),$$

which immediately implies that

$$|\lambda_{\min}(S) - \lambda_{\min}(\sum_i \ell_i \tilde{\boldsymbol{\mu}}_i \tilde{\boldsymbol{\mu}}_i^\top)| = o_{\mathbb{P}}(\sum_i \ell_i).$$

Invoking (14), we have

$$\lambda_{\min} \geq \frac{c_0}{2}(1 + o_{\mathbb{P}}(1)) \sum_i \ell_i,$$

and consequently, $\rho^\star(S^{-1}) = O_{\mathbb{P}}((\sum_i \ell_i)^{-1})$. It follows

$$\rho^\star(S^{-2}) \sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^4 = O_{\mathbb{P}}(\frac{\sum_i \ell_i^2}{(\sum_i \ell_i)^2}) = O_{\mathbb{P}}(\frac{\max_i \ell_i^2}{\sum_i \ell_i^2}) = o_{\mathbb{P}}(1). \tag{57}$$

5.7. *Combining the pieces together*

Noting that a sequence of random variable $W_n \xrightarrow{\mathbb{P}} 0$ iff for every sub-sequence n_k, there exists a further sub-sequence n_{k_l} such that $W_{n_{k_l}} \xrightarrow{\mathbb{P}} 0$, we proceed similar to Section 5.4. Given any subsequence n_k, from (56), (57), and (51), there exists a subsequence n_{k_l} such that

$$\left(\frac{\max_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}\right)_{n_{k_l}} \xrightarrow{a.s.} 0, \ (\rho^\star(S^{-1})\sqrt{\sum_i \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^4})_{n_{k_l}} \xrightarrow{a.s.} 0, \text{ and } \left(\frac{\rho^\star(S)}{\sum_{i=1}^n \ell_i^2 |\bar{\boldsymbol{X}}_{i\cdot}|^2}\right)_{n_{k_l}} \xrightarrow{a.s.} 0.$$

Moreover, recall Ψ from Theorem 3. Clearly from (56) and (57) and by (15), along the sequence n_{k_l}, $B_{n_{k_l}} \Psi \xrightarrow{a.s.} 0$. Therefore, invoking Theorem 3, we obtain

$$\left(\rho^\star(A^{-1}(\hat{A} - A))\right)_{n_{k_l}} = o_{\mathbb{P}}(1),$$

which completes the proof via the double-subsequence argument as stated in the beginning of Section 5.4. $\square$