Spatial Heteroskedasticity and Autocorrelation Consistent Estimation of Covariance Matrix

Min Seong Kim and Yixiao Sun*

Department of Economics, UC San Diego

This version: April 2009

Abstract

This paper considers spatial heteroskedasticity and autocorrelation consistent (spatial HAC) estimation of covariance matrices of parameter estimators. We generalize the spatial HAC estimators introduced by Kelejian and Prucha (2007) to apply to linear and nonlinear spatial models with moment conditions. We establish its consistency, rate of convergence and asymptotic truncated mean squared error (MSE). Based on the asymptotic truncated MSE criterion, we derive the optimal bandwidth parameter and suggest its data dependent estimation procedure using a parametric plug-in method. The finite sample performances of the spatial HAC estimator are evaluated via Monte Carlo simulation.

Keywords: Asymptotic mean squared error, Heteroskedasticity and autocorrelation, Covariance matrix estimator, Optimal bandwidth choice, Robust standard error, Spatial dependence.

JEL Classification Number: C13, C14, C21

*Email: msk003@ucsd.edu and yisun@ucsd.edu. Correspondence to: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508. We thank Graham Elliott Dimitris Politis and Hal White for helpful comments.

1 Introduction

This paper studies spatial heteroskedasticity and autocorrelation consistent (HAC) estimation of covariance matrices of parameter estimators. As heteroskedasticity is a well known feature of cross sectional data (e.g. White (1980)), spatial dependence is also a common property due to interactions among economic agents. Therefore, robust inference in presence of heteroskedasticity and spatial dependence is an important problem in spatial data analysis.

The first discussion of spatial HAC estimation is Conley (1996, 1999). He proposes a spatial HAC estimator based on the assumption that each observation is a realization of a random process, which is stationary and mixing, at a point in a two-dimensional Euclidean space. Conley and Molinari (2007) examine the performance of this estimator using Monte Carlo simulation. Their results show that inference is robust to the measurement error in locations. Robinson (2005) considers nonparametric kernel spectral density estimation for weakly stationary processes on a d-dimensional lattice.

Kelejian and Prucha (2007, hereafter KP) also develop a spatial HAC estimator. As in many empirical studies, they model spatial dependence in terms of a spatial weighting matrix. The difference is that the weighting matrix is not assumed to be known and is not parametrized. Typical examples of this type of processes include the spatial autoregressive processes and spatial moving average processes. Local nonstationarity and heteroskedasticity are built-in features of these type of processes. This is in sharp contrast with Conley (1996, 1999) and Robinson (2005) in which the process is assumed to be stationary. KP employ an *economic distance* to characterize the decaying pattern of the spatial dependence. The covariance of random variables at locations i and j is a function of $d_{ij,n}$, the economic distance between them. As the economic distance increases, the covariance decreases in absolute value and vice versa. The existence of such an economic distance enables us to use the kernel method for the standard error estimation. The estimator is a weighted sum of sample covariances with weights depending on the relative distances, that is, $d_{ij,n}/d_n$ for some bandwidth parameter d_n.

We generalize the spatial HAC estimator proposed by KP to be applicable to general linear and nonlinear spatial models and establish its asymptotic properties. We provide the conditions for consistency and the rate of convergence. Let $E\ell_n$ denote the mean of the average number of pseudo-neighbors. By definition, two units are pseudo-neighbors if their distance is less than d_n. We show that the spatial HAC estimator is consistent if $E\ell_n = o(n)$ and $d_n \to \infty$ as $n \to \infty$. This result implies that the rate of convergence of the estimator is $E\ell_n/n$. Comparing our results with Andrews (1991), we find that the properties of the spatial HAC estimator we consider are interestingly parallel to those of the time series HAC estimator, even though they assume different DGPs and have different dependence structures.

We decompose the difference of the spatial HAC estimator from the true covariance matrix into three parts. The first part is due to the estimation error of model parameters and the second and third parts are bias and variance terms even if the model parameters are known. We derive the asymptotic bias and variance and show that the estimation error vanishes faster than the other two terms under some regularity conditions. As a result, the truncated Mean Squared Error (MSE) of the spatial HAC estimator is dominated by the bias and variance terms. This key result provides us the opportunity to select the bandwidth parameter to balance the asymptotic squared bias with variance. We find that the optimal bandwidth choice depends on the weighting matrix S_n used in the MSE criterion. Depending on which model parameter is the focus of interest, we suggest different choices of the weighting matrix. This scheme coincides with that suggested by Politis (2007).

We provide a data-driven implementation of the optimal bandwidth parameter and examine the finite sample properties of our spatial HAC estimator and the associated test via Monte Carlo simulation. We compare the performance of competing estimators using different choices of d_n and S_n. In addition, the effects of location errors and the performance of the plug-in procedure with mis-specified parametric model are examined. We also consider the case when the observations are located irregularly and compare the performance of the standard normal approximation with two naive bootstrap approximations for hypothesis testing.

In addition to KP, the paper that is most closely related to ours is Andrews (1991) who employ the asymptotic truncated MSE criterion to select the bandwidth parameter for time series HAC estimation. His paper in turn can be traced back to the literature on spectral density estimation. We extend Andrews (1991) to the spatial setting. The extension is nontrivial as spatial processes are more difficult to deal with, especially when they are not weakly stationary.

The remainder of the paper is as follows. Section 2 describes the estimation problem and the underlying spatial process we consider and introduces our spatial HAC estimator. Section 3 establishes the consistency, the rate of convergence, and the asymptotic truncated MSE of the spatial HAC estimator. Section 4 derives asymptotically optimal sequences of fixed bandwidth parameters and proposes a data-dependent implementation. Section 5 studies the consistency, the rate of convergence, and the asymptotic truncated MSE of the spatial HAC estimator with the estimated optimal bandwidth parameter. Section 6 presents Monte Carlo simulation results. Section 7 concludes.

2 Spatial Processes and HAC Estimators

In a general spatial model with moment restrictions, the asymptotic distribution of a parameter estimator often satisfies

$$(B_n J_n B_n')^{-\frac{1}{2}} \sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I_r), \text{ as } n \to \infty,$$

where n is the sample size, B_n is a nonstochastic $r \times p$ matrix and

$$J_n = \text{var}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_{i,n}(\theta_0) \right) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[V_{i,n}(\theta_0) V_{j,n}(\theta_0)' \right], \tag{1}$$

$V_{i,n}(\theta)$ is a random p-vector for each $\theta \in \Theta \subset R^r$. For IV estimation of a linear regression model, $V_{i,n}(\theta) = Z_{i,n}(Y_{i,n} - X_{i,n}'\theta)$ where $Z_{i,n}$ is the vector of instruments. For pseudo-ML estimation, $V_{i,n}(\theta)$ is the score function of the i^{th} observation. For GMM estimation, $V_{i,n}(\theta)$ is the moment vector. A prime example of this setting is the spatial linear regression:

$$Y_{i,n} = X_{i,n}'\theta_0 + u_{i,n},$$

where $E(u_{i,n}|X_{i,n}) = 0$. The OLS estimator of θ_0 is

$$\hat{\theta} = \left(\frac{1}{n} \sum_{i=1}^{n} X_{i,n} X_{i,n}' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} X_{i,n} Y_{i,n} \right).$$

Under some regularity conditions, $(B_n J_n B_n')^{-\frac{1}{2}}\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0, I_r)$ where

$$J_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left(X_{i,n}u_{i,n}\right)\left(X_{j,n}u_{j,n}\right)' \text{ and } B_n = \left(\frac{1}{n}\sum_{i=1}^{n} X_{i,n}X_{i,n}'\right)^{-1}.$$

We are interested in estimating the asymptotic variance of $\sqrt{n}(\hat{\theta}-\theta_0)$. As B_n is often easy to estimate by replacing θ_0 with $\hat{\theta}$, our focus is on consistent estimation of J_n. By extending the spatial HAC estimator proposed in KP, we can construct a spatial HAC estimator of J_n as follows

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{V}_{i,n}\hat{V}_{j,n}' K\left(\frac{d_{ij,n}}{d_n}\right), \tag{2}$$

where $\hat{V}_{i,n} = V_{i,n}(\hat{\theta})$ and $K(\cdot)$ is a real-valued kernel function. $d_{ij,n}$ is the economic distance between units i and j and d_n is a bandwidth or truncation parameter. We assume that the degree of spatial dependence is a function of $d_{ij,n}$. More specifically, if $d_{ij,n}$ is small, $V_{i,n}$ and $V_{j,n}$ are highly dependent. Whereas, if it is large, the two units are rather close to being independent.

We assume that $V_{i,n}(= V_{i,n}(\theta_0))$ for $i = 1, \ldots, n$ are generated from np common innovations:

$$V_{i,n} = \tilde{R}_{in}\tilde{\varepsilon}_n \tag{3}$$

where

$$\tilde{R}_{in} = \begin{bmatrix} \begin{pmatrix} \tilde{r}_{i1,n}^{(1)} & \ldots & \tilde{r}_{in,n}^{(1)} \end{pmatrix} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \begin{pmatrix} \tilde{r}_{i1,n}^{(p)} & \ldots & \tilde{r}_{in,n}^{(p)} \end{pmatrix} \end{bmatrix}$$

is a $p \times np$ block diagonal matrix with unknown elements, $\tilde{\varepsilon}_n^{(c)} = \left(\tilde{\varepsilon}_{1n}^{(c)}, ..., \tilde{\varepsilon}_{\ell n}^{(c)}, ..., \tilde{\varepsilon}_{n,n}^{(c)}\right)'$ and $\tilde{\varepsilon}_n = ((\tilde{\varepsilon}_n^{(1)})', ..., (\tilde{\varepsilon}_n^{(p)})')'$ is a $np \times 1$ vector of innovations. We assume that

$$\text{var}\left(\tilde{\varepsilon}_n^{(c)}\right) = \sigma_{cc} I_n,\ \text{cov}\left(\tilde{\varepsilon}_n^{(c)}, \tilde{\varepsilon}_n^{(d)}\right) = \sigma_{cd} I_n$$

so that the variance matrix of $\tilde{\varepsilon}_n$ is of the form

$$\text{var}(\tilde{\varepsilon}_n) = \Sigma \otimes I_n \text{ with } \Sigma = (\sigma_{ij}),$$

where $\otimes$ denotes the Kronecker product. The process exhibited in (3) allows nonstationarity and unconditional heteroskedasticity of $V_{i,n}$.

Let $R_{in} := \tilde{R}_{in}\left(\Sigma^{1/2} \otimes I_n\right)$ and $\varepsilon_n := (\varepsilon_{1,n}, ..., \varepsilon_{\ell,n}, ..., \varepsilon_{np,n}) = \left(\Sigma^{-1/2} \otimes I_n\right)\tilde{\varepsilon}_n$, then

$$V_{i,n} = R_{in}\varepsilon_n \text{ and } \text{var}(\varepsilon_n) = I_{np}.$$

The matrix R_{in} can be written more explicitly as

$$
\begin{aligned}
R_{in} &:= \begin{bmatrix} \begin{pmatrix} r^{(1)}_{i1,n} & \dots & r^{(1)}_{i,np,n} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} r^{(p)}_{i1,n} & \dots & r^{(p)}_{i,np,n} \end{pmatrix} \end{bmatrix} \\
&= \begin{bmatrix} \sigma^{11} \begin{pmatrix} \tilde{r}^{(1)}_{i1,n} & \dots & \tilde{r}^{(1)}_{in,n} \end{pmatrix} & \cdots & \sigma^{1p} \begin{pmatrix} \tilde{r}^{(p)}_{i1,n} & \dots & \tilde{r}^{(p)}_{in,n} \end{pmatrix} \\ \vdots & \ddots & \vdots \\ \sigma^{p1} \begin{pmatrix} \tilde{r}^{(p)}_{i1,n} & \dots & \tilde{r}^{(p)}_{in,n} \end{pmatrix} & \dots & \sigma^{pp} \begin{pmatrix} \tilde{r}^{(p)}_{i1,n} & \dots & \tilde{r}^{(p)}_{in,n} \end{pmatrix} \end{bmatrix}
\end{aligned}
$$

where σ^{ij} is the (i,j)-th element of $\Sigma^{1/2}$.

We make the following assumption on ε_n.

Assumption 1 *For each $n \geq 1$, $\{\varepsilon_{\ell,n}\}$ are i.i.d.$(0,1)$ with $E\varepsilon^4_{\ell,n} \leq c_E$ for a constant $c_E < \infty$.*

For simplicity, we assume that $\varepsilon_{i,n}$ is independent of $\varepsilon_{j,n}$ for $i \neq j$. Our results can be generalized but with more tedious calculations. Under Assumption 1, the covariance matrix between $V_{i,n}$ and $V_{j,n}$ is given by

$$
\Gamma_{ij,n} := \left(\gamma^{(cd)}_{ij,n}\right) = E[V_{i,n}V'_{j,n}] = R_{in}R'_{jn} \tag{4}
$$

where the (c,d)-th element of $\Gamma_{ij,n}$ is denoted by $\gamma^{(cd)}_{ij,n}$. Accordingly, equation (1) can be restated as

$$
J_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma_{ij,n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}R_{in}R'_{jn}
$$

and the (c,d)-th element of J_n is

$$
\begin{aligned}
J_n(c,d) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}E\sum_{m=1}^{np}\sum_{\ell=1}^{np} r^{(c)}_{im,n} r^{(d)}_{j\ell,n}\varepsilon_{m,n}\varepsilon_{\ell,n} \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{np} r^{(c)}_{im,n} r^{(d)}_{jm,n}.
\end{aligned}
$$

Assumption 2 *For all $j = 1,2,...,np$, and $s = 1,2,...,p$, $\sum_{k=1}^{n}\left|r^{(s)}_{kj,n}\right| < c_R$ for some constant c_R, $0 < c_R < \infty$.*

Assumption 3 *There exists $q_d > 0$ such that $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\Gamma_{ij,n}\| d^{q_d}_{ij,n} < \infty$ for all n, where $\|A\|$ denotes the Euclidean norm of matrix A.*

Assumption 4 *For $k = 1,2,...,n$*

$$
\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma_{ij,n} = \lim_{n\to\infty}\sum_{j=1}^{n}\Gamma_{kj,n}
$$

Assumptions 2 and 3 impose conditions on the persistence of the spatial process. If $\left|\sigma^{ij}\right| \leq C$ for some constant $C > 0$, then Assumption 2 holds if $\sum_{k=1}^{n} \left|\tilde{r}_{kj,n}^{(s)}\right| < c_R/C$. Since $\left|\tilde{r}_{kj,n}^{(s)}\right|$ can be regarded as the (absolute) change of $V_{i,n}^{(s)}$ in response to one unit change in $\tilde{\varepsilon}_{jn}^{(s)}$, the summability condition requires that the aggregate response be finite. The condition holds trivially if the set $\{\tilde{r}_{kj,n}^{(s)}, k = 1, 2, ..., n\}$ has only a finite number of nonzero elements. In this case, the dependence induced by the innovation $\tilde{\varepsilon}_{jn}^{(s)}$ are limited to a finite number of units. Assumption 3 states that $\Gamma_{ij,n}$ decays to zero fast enough such that $n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\Gamma_{ij,n}\| \, d_{ij,n}^{q_d}$ is finite for all n. This excludes the case in which the sample size increases because of more intensive sampling within a given distance. This condition enables us to truncate the sum $\sum_{j=1}^{n} \|\Gamma_{ij,n}\|$ and downweigh the summand without incurring a large error. As in the time series literature, this assumption helps us control the asymptotic bias of the spatial HAC estimator.

Assumption 4 states that in large samples the row sum of the covariance matrix is the same across different rows. This assumption is related to covariance stationarity but they are quite different. If the spatial process is covariance stationary and the units are located on a regular lattice, then the assumption holds trivially for every sample size n. On the other hand, the assumption holds for nonstationary processes such as the spatial AR processes as long as the units are not distributed very unevenly. In addition, the assumption is not for a finite sample size. It holds for any spatial process which is not stationary for a finite sample size but becomes approximately stationary in large samples. The assumption allows us to derive a neat expression for the asymptotic variance of the spatial HAC estimator. It can be relaxed if we are interested only in the consistency result.

The spatial HAC estimator we consider is based on (2) but it also allows for measurement errors in the economic distances as follows

$$\hat{J}_n = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{V}_{i,n} \hat{V}_{j,n}' K\left(\frac{d_{ij,n}^*}{d_n}\right), \tag{5}$$

where $d_{ij,n}^* = d_{ij,n} + \nu_{ij,n}$ and $\nu_{ij,n}$ denotes the measurement error. Data on economic distances available to econometricians usually contain measurement errors. For example, the economic distance between two countries may be measured by transportation cost in international trade and this inevitably involves some measurement error. Sometimes the economic distance may be estimated from another related model. The underlying estimation error is a special case of measurement errors.

Assumption 5 *(i)* $\{\nu_{ij,n}\}$ *are independent of* $\{\varepsilon_{\ell,n}\}$. *(ii)* $\nu_{ij,n} = o(d_n)$ *as* $d_n \to \infty$. *(iii)* $n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\Gamma_{ij,n}\| \, E\left|\nu_{ij,n}\right|^{q_d} < \infty$ *for all* n.

Assumption 5 is weaker than the restriction on measurement errors in KP and Conley (1999). They require the measurement errors to be bounded by a finite constant. However, there may be a case when the measurement error grows as the distance of two units becomes farther. We allow this. $\nu_{ij,n}$ can increase as $d_{ij,n}$ increases as long as Assumptions 5(ii) and (iii) hold. Under this assumption, it is straightforward that $n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\Gamma_{ij,n}\| \, E(d_{ij,n}^*)^{q_d} < \infty$ for all n. Essentially, measurement errors in distance can not be so large as to change the summability of $n^{-1} \sum_{i=1}^{n} \|\Gamma_{ij,n}\| \, E(d_{ij,n}^*)^{q}$.

Let $\ell_{i,n} = \sum_{j=1}^{n} 1\{d_{ij,n}^* \leq d_n\}$ and $\ell_n = 1/n \sum_{i=1}^{n} \ell_{i,n}$. If we call unit j a pseudo-neighbor of unit i if $d_{ij,n}^* \leq d_n$, then $\ell_{i,n}$ is the number of pseudo-neighbors that unit i has and ℓ_n is the average number of pseudo-neighbors. Here we use the terminology "pseudo-neighbor" in order to differentiate it from the common usage of "neighbor" in spatial modeling. We maintain the following assumption on the number of pseudo-neighbors.

Assumption 6 *(i) For all $i = 1, 2, ..., n$, $\ell_{i,n} \leq CE\ell_n$ for some constant C. (ii) $E\left|\ell_{i,n} - E\ell_n\right| = o(E\ell_n)$ as $d_n \to \infty$ and $n \to \infty$.*

Assumption 6(i) is very weak as C can be a large constant. Assumption 6(ii) says that the number of pseudo-neighbors for each unit i is close to the average number of pseudo-neighbors. This assumption is also weak as it allows $\ell_{i,n}$ to be different from ℓ_n as long as the difference does not grow too fast as n increases. This assumption rules out the case that the units are distributed very unevenly in space.

3 Asymptotic Properties of Spatial HAC Estimators

This section presents the consistency conditions, the rate of convergence, and the asymptotic truncated MSE of the fixed bandwidth kernel spatial HAC estimator. We begin by introducing the assumption on the kernel used in the spatial HAC estimator.

Assumption 7 *(i) The kernel $K : R \to [0,1]$ satisfies $K(0) = 1, K(x) = K(-x), K(x) = 0$ for $|x| \geq 1$. (ii) For all $x_1, x_2 \in R$ there is a constant, $c_L < 0$, such that*

$$\left|K(x_1) - K(x_2)\right| \leq c_L \left|x_1 - x_2\right|.$$

(iii) $(E\ell_n)^{-1} E \sum_{j=1}^{n} K^2\left(\frac{d_{ij,n}^*}{d_n}\right) \to \bar{K}$ for all i.

Examples of kernels which satisfy Assumptions 7 (i) and (ii) are the Bartlett, Tukey-Hanning and Parzen kernels. The quadratic spectral (QS) kernel does not satisfy Assumption 7(i) because it does not truncate. We may generalized our results to include the QS kernel but this requires a considerable amount of work. Assumption 7(iii) is more of an assumption on the distribution of the units. In the case of a 2-dimensional lattice structure, we have

$$\bar{K} = \frac{1}{\pi} \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1+x^2}} K^2(\sqrt{x^2+y^2}) dy dx = \int_{-1}^{1} K^2(r) dr$$

This relationship also holds for other structures if the units are not distributed very unevenly. In finite samples, we may use

$$\bar{K}_n = (n\ell_n)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} K^2\left(\frac{d_{ij,n}^*}{d_n}\right)$$

for $\bar{K}$.

Since the spatial process of $V_{i,n}$ is locally nonstationary, $\hat{J}_n$ is not a weighted average of the periodogram at different frequencies. Therefore, the kernel functions which generate positive semi-definite (psd) HAC estimators under the assumption of covariance stationarity do not necessarily guarantee the positive semi-definiteness of $\hat{J}_n$. KP introduce a class of kernel functions which generate psd spatial HAC estimators in finite samples when the distance measure corresponds to a Euclidean norm in $\mathbb{R}^p$, $p \geq 1$. Let $\varphi(x)$ be a function defined by

$$\varphi(x) = \Gamma\left(\frac{p}{2}\right)\int_0^\infty \left(\frac{2}{rx}\right)^{\frac{p-2}{2}} J_{\frac{p-2}{2}}(rx)dF(r), \quad x \geq 0$$

where F is a probability distribution function on $[0,\infty)$ and $J_{\frac{p-2}{2}}$ is a Bessel function of order $(p-2)/2$. Then

$$\left[\varphi\left(\|z_i - z_j\|\right)\right]_{i,j=1}^n$$

is psd for any points $z_1,\ldots,z_n$ in $\mathbb{R}^p$. Note that $\varphi : [0,\infty) \to \mathbb{R}$ with $\varphi(0) = 1$. If the kernel function $K(x)$ can be rewritten as $\varphi(x)$ for some probability distribution function $F_K(\cdot)$ and $d^*_{ij,n}/d_n = \|z_i - z_j\|_p$, then $\hat{J}_n$ is psd.

The asymptotic variance of $\hat{J}_n$ depends on g, the limit value of J_n :

$$g := \lim_{n\to\infty} J_n = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \Gamma_{ij,n}.$$

The asymptotic bias of $\hat{J}_n$ is determined by the smoothness of the kernel at zero and the rate of decaying of the spatial dependence as a function of the distance. Define

$$K_{q_0} = \lim_{x\to 0}\frac{1-K(x)}{|x|^{q_0}}, \qquad \text{for } q_0 \in [0,\infty).$$

and let $q = \max\{q_0 : K_{q_0} < \infty\}$ be the *Parzen characteristic exponent* of $K(x)$. The magnitude of q reflects the smoothness of $K(x)$ at $x = 0$. We assume $q \leq q_d$ throughout the paper. Let

$$g_n^{(q)} = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \Gamma_{ij,n}E(d^*_{ij,n})^q,\ g^{(q)} = \lim_{n\to\infty} g_n^{(q)} := \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \Gamma_{ij,n}E(d^*_{ij,n})^q.$$

Next we introduce additional assumptions required to obtain the asymptotic properties of $\hat{J}_n$.

Assumption 8 *(i)* $\sqrt{n}\left(\hat{\theta} - \theta_0\right) = O_p(1)$. *(ii)* $\sup_i E\sup_{\theta\in\Theta_n}\|V_{i,n}(\theta)\|^2 < \infty$ *where* Θ_n *is a small neighborhood around* θ_0. *(iii)* $\sup_i E\sup_{\theta\in\Theta_n}\|\frac{\partial}{\partial\theta'}V_{i,n}(\theta)\|^2 < \infty$. *(iv) For* $r = 1,\ldots,p$,
$\sup_i E\sup_{\theta\in\Theta_n}\|\frac{\partial^2}{\partial\theta\partial\theta'}V^{(r)}_{i,n}(\theta)\|^2 < \infty$. *(v)* $\sup_{i,a,b}\sum_{j=1}^\infty\left\|EV^{(r)}_{j,n}V^{(r)}_{b,n}\frac{\partial}{\partial\theta'}V^{(s)}_{i,n}\frac{\partial}{\partial\theta}V^{(s)}_{a,n}\right\| < \infty$ *for* $r,s = 1,\ldots,p$.

Assumption 8(i) usually holds by the asymptotic normality of parameter estimators. Assumption 8(ii) is implied by Assumptions 1 and 2. Assumptions 8(iii), (iv) and (v) are trivial in a linear regression case.

We define the MSE criterion as

$$MSE\left(\frac{n}{E\ell_n}, \hat{J}_n, S\right) = \frac{n}{E\ell_n}E\left[\text{vec}(\hat{J}_n - J_n)'S\text{vec}(\hat{J}_n - J_n)\right],$$

where S is some $p^2 \times p^2$ weighting matrix and vec$(\cdot)$ is the column by column vectorization function. We also define $\tilde{J}_n$ as the pseudo-estimator that is identical to $\hat{J}_n$ but is based on the true parameter, θ_0, instead of $\hat{\theta}$. That is,

$$\tilde{J}_n = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} V_{i,n} V_{j,n}' K \left(\frac{d_{ij,n}^*}{d_n} \right).$$

Under the assumptions above, the effect of using $\hat{\theta}$ instead of θ_0 on the asymptotic property is $o_p(1)$ as Theorem 1(c) states below. Therefore, we use $\tilde{J}_n$ to analyze the asymptotic properties of $\hat{J}_n$. Notwithstanding, if $\hat{\theta}$ has an infinite second moment, the underlying estimation error can dominate the MSE criterion. To circumvent the undue influence of $\hat{\theta}$ on the criterion of performance, we follow Andrews (1991) and replace the MSE criterion with a truncated MSE criterion. We define

$$MSE_h \left(\frac{n}{E\ell_n}, \hat{J}_n, S_n \right) = E \left[\min \left\{ \left| \frac{n}{E\ell_n} \mathrm{vec}(\hat{J}_n - J_n)' S_n \mathrm{vec}(\hat{J}_n - J_n) \right|, h \right\} \right]$$

where S_n is a $p^2 \times p^2$ weighting matrix that may be random. The criterion which we base on for the optimality result is the asymptotic truncated MSE, which is defined as

$$\lim_{h \to \infty} \lim_{n \to \infty} MSE_h \left(\frac{n}{E\ell_n}, \hat{J}_n, S_n \right).$$

This criterion yields the same value as the asymptotic MSE when $\hat{\theta}$ has well defined moments, but does not diverge to infinity when $\hat{\theta}$ has infinite second moments.

Assumption 9 *(i)* $E\varepsilon_{l,n}^8 < \infty$. *(ii)* $S_n \xrightarrow{p} S$ *for a positive definite matrix* S.

Let tr denote the trace function and K_{pp} the $p^2 \times p^2$ commutation matrix. Under the assumptions above, we have the following theorem.

Theorem 1 *Suppose that Assumptions 1-7 hold,* $E\ell_n$ *and* $d_n \to \infty$ *and* $E\ell_n/n \to 0$.

(a) $\lim_{n\to\infty} \frac{n}{E\ell_n} var \left(vec\tilde{J}_n \right) = \bar{K}(I + K_{pp})(g \otimes g)$.

(b) $\lim_{n\to\infty} d_n^q (E\tilde{J}_n - J_n) = -K_q g^{(q)}$.

(c) *If Assumption 8 holds and* $\frac{d_n^{2q} E\ell_n}{n} \to \tau \in (0, \infty)$*, then* $\sqrt{\frac{n}{E\ell_n}} \left(\hat{J}_n - J_n \right) = O_p(1)$ *and* $\sqrt{\frac{n}{E\ell_n}} \left(\hat{J}_n - \tilde{J}_n \right) = o_p(1)$.

(d) *Under the conditions of part (c) and Assumption 9,*

$$\begin{aligned} &\lim_{h\to\infty} \lim_{n\to\infty} MSE_h \left(\frac{n}{E\ell_n}, \hat{J}_n, S_n \right) \\ &= \lim_{h\to\infty} \lim_{n\to\infty} MSE_h \left(\frac{n}{E\ell_n}, \tilde{J}_n, S_n \right) \\ &= \lim_{n\to\infty} MSE \left(\frac{n}{E\ell_n}, \tilde{J}_n, S \right) \\ &= \frac{1}{\tau} K_q^2 \left(vecg^{(q)} \right)' S \left(vecg^{(q)} \right) + \bar{K} tr \left(S(I + K_{pp})(g \otimes g) \right). \end{aligned}$$

Proofs are given in the appendix. For each element of $\tilde{J}_n$, $\lim_{n\to\infty} \frac{n}{E\ell_n} \text{cov}\left(\tilde{J}_{rs,n}, \tilde{J}_{cd,n}\right) = \bar{K}\left(g_{rc}g_{sd} + g_{rd}g_{sc}\right)$ and $\lim_{n\to\infty} d_n^q (E\tilde{J}_{rs,n} - J_{rs,n}) = -K_q g_{rs}^{(q)}$. Theorem 1(a) and (b) show that the asymptotic variance and bias of $\tilde{J}_n$ depend on the choice of the bandwidth. When we increase the bandwidth, the bias decreases and the variance increases because $E\ell_n$ increases with d_n.

The second part of Theorem 1(c) shows that, compared with the variance term in part (a), the effect of using $V_{i,n}(\hat{\theta})$ instead of $V_{i,n}(\theta_0)$ in the construction of the spatial HAC estimator is of a smaller order. Therefore, the rate of convergence is obtained by balancing the variance and the squared bias. Accordingly, $E\ell_n = o(n)$ is the condition for the consistency of $\hat{J}_n$ and its rate of convergence is $\sqrt{E\ell_n/n}\ (= O(d_n^{-q}))$. If we assume that $E\ell_n = O(d_n^{\eta})$ for some $\eta > 0$, then the rate of convergence can be rewritten as $n^{q/(\eta+2q)}$. The results here are different from those provided by KP. In their paper, the condition for consistency is $E\ell_n = o(n^{\tau})$ where $\tau \leq \frac{1}{2}$ and the rate of convergence is $n^{q/(\eta+4q)}$. They obtain this slower rate of convergence by balancing the terms from the estimation error in $\hat{\theta}$ and the asymptotic bias. Their rate is not the best obtainable because their bound for the estimation error term is too loose.

It is also interesting that the asymptotic properties of the spatial HAC estimator are very similar to those of the time series HAC estimator even though their DGPs and dependence structures are different from each other. Instead of using d_n as the bandwidth parameter, we can also use $E\ell_n$ as the bandwidth parameter. In the time series case, $d_n = E\ell_n$. Substituting this relationship into Theorem 1, we obtain the same results as given in Parzen (1957), Hannan (1970) and Andrews (1991).

4 Optimal Bandwidth Parameter and Data Dependent Bandwidth Selection

This section presents a sequence of optimal bandwidth parameters which minimize the asymptotic truncated MSE of $\hat{J}_n$ and gives a data-driven implementation. We also consider the choice of the weighting matrix S_n.

We obtain the optimal bandwidth parameter directly as a corollary to Theorem 1(d). Let $d_n^\star$ be the optimal bandwidth parameter. Then

$$d_n^\star = \arg\min_{d_n} \frac{1}{d_n^{2q}} K_q^2 \left(\text{vec } g^{(q)}\right)' S_n \left(\text{vec } g^{(q)}\right) + \frac{E\ell_n}{n} \bar{K} \text{tr}\left(S_n(I + K_{pp})(g \otimes g)\right) \tag{6}$$

If the relation between $E\ell_n$ and d_n is specified, (6) can be restated in an explicit form. For example, we may assume that $E\ell_n = \alpha_n d_n^{\eta}$ and $\alpha_n = O(1)$ for some $\eta > 0$. Then (6) is reduced to:

$$\begin{aligned} d_n^\star &= \arg\min_{d_n} \frac{1}{d_n^{2q}} K_q^2 \left(\text{vec } g^{(q)}\right)' S_n \left(\text{vec } g^{(q)}\right) + \frac{\alpha_n d_n^{\eta}}{n} \bar{K} \text{tr}\left(S_n(I + K_{pp})(g \otimes g)\right) \\ &= \left(\frac{nqK_q^2 \kappa(q)}{\alpha_n \eta}\right)^{\frac{1}{2q+\eta}} \end{aligned} \tag{7}$$

where

$$\kappa(q) = \frac{\left(\text{vec } g^{(q)}\right)' S_n \left(\text{vec } g^{(q)}\right)}{\bar{K} \text{tr}\left(S_n(I + K_{pp})(g \otimes g)\right)}.$$

Corollary 2 *Suppose Assumptions 1-9 hold. Assume that* $E\ell_n = \alpha_n d_n^\eta$ *for some* $\eta > 0$, $\alpha_n = \alpha + o(1)$. *Then, for any sequence of bandwidth parameters* $\{d_n\}$ *such that* $\frac{d_n^{2q} E\ell_n}{n} \to \tau \in (0, \infty)$, $\{d_n^\star\}$ *is preferred in the sense that*

$$\lim_{h\to\infty} \lim_{n\to\infty} \left(MSE_h \left(n^{2q/(2q+\eta)}, \hat{J}_n(d_n), S_n \right) - MSE_h \left(n^{2q/(2q+\eta)}, \hat{J}_n(d_n^\star), S_n \right) \right) \geq 0.$$

The inequality is strict unless $d_n = d_n^\star + o(n^{1/(2q+\eta)})$.

In general, η is equal to the dimension of the space. In the time series case, $\eta = 1$ while in the two dimensional regular lattice case, $\eta = 2$. As a result, the optimal bandwidth $d_n^\star$ depends on the dimension of space. Given the nonparametric nature of our estimator, this is not surprising. In contrast, KP suggest using $d_n = [n^{1/4}]$, which is rate optimal only if $q = 1$ and $\eta = 2$. In general, both the rate and constant are suboptimal.

$d_n^\star$ is a function of g and $g^{(q)}$ which are unknown in finite samples. Therefore, the optimal bandwidth $d_n^\star$ is not feasible in practice. For this reason, a data dependent estimation procedure is needed for implementation. Among several data dependent bandwidth selection methods, plug-in methods are appropriate in this case because we consider the estimation of J_n at given data. In the plug-in methods, unknown parameters are estimated using a parametric or nonparametric method (e.g. Andrews (1991). Newey and West (1987, 1994)). The former yields a less variable bandwidth parameter but may introduce an asymptotic bias due to the mis-specification of the parametric model. In contrast, the latter does not require the knowledge of the DGP, but it converges more slowly than the former, which causes bandwidth selection to be less reliable. Since the optimal bandwidth involves $g^{(q)}$, a quantity that is very hard to estimate, we focus on the parametric plug-in method in this paper. In fact, the rate of convergence for a nonparametric estimator of $g^{(q)}$ is generally slower than that for g itself.

Figure 1 presents the percentage increase in MSE relative to the minimum MSE as a function of the bandwidth. The graph is based on the spatial AR(1) process $V_n = \rho W_n V_n + \varepsilon_n$ on a square grid of integers, where W_n is a contiguity matrix whose threshold is $\sqrt{2}$ and $\varepsilon_{i,n} \overset{i.i.d}{\sim} N(0, 1)$. The sample size is $n = 400$. As a standard practice, W_n is row-standardized and its diagonal elements are zero. The curve is U-shaped for each ρ and therefore our goal is to choose the bandwidth which is reasonably close to $d_n^\star$. As argued by Andrews (1991), good performance of a HAC estimator only requires the automatic bandwidth parameter to be near the optimal bandwidth value and not precisely equal to it.

The simplest and most popular approximating parametric model is the spatial AR(1) model for $V_n^{(c)}$, $c = 1, \ldots, p$. Depending on the correlation structure, spatial MA(q) or spatial ARMA(p,q) models can also be used. As an example, consider the case that $V_n^{(c)}$ follows a spatial AR(1) process of the form:

$$V_n^{(c)} = \rho_c W_n^{(c)} V_n^{(c)} + \tilde{\varepsilon}_n^{(c)} = (I_n - \rho_c W_n^{(c)})^{-1} \tilde{\varepsilon}_n^{(c)},$$

where $\tilde{\varepsilon}_{i,n}^{(c)} \overset{i.i.d}{\sim} (0, \sigma_\varepsilon^2)$ and $W_n^{(c)}$ is a spatial weight matrix. $W_n^{(c)}$ is determined a priori and by convention it is row-standardized and its diagonal elements are zero. See Anselin (1988). We can estimate ρ_c by quasi-maximum likelihood (QML) or spatial two stage least squares (2SLS) estimators (e.g. Kelejian and I.R. Prucha (1998)). In fact, a simple OLS estimator can be used. If the spatial AR(1) model is the true data generating process, then the OLS

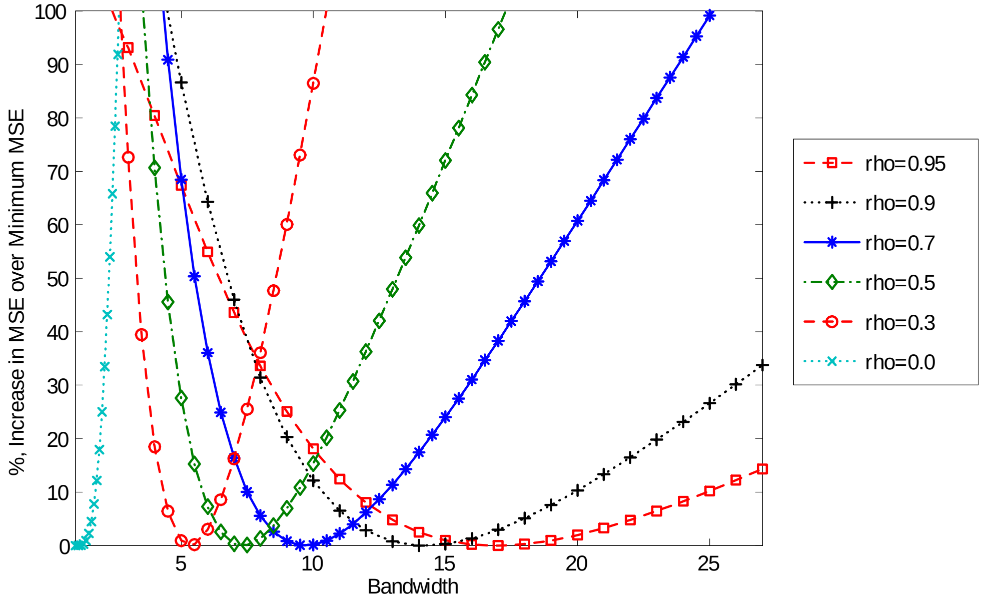

Figure 1: Spatial AR(1) Process : $V_n = \rho W_n V_n + \varepsilon_n$, $n = 400$

estimator is inconsistent while the QML and 2SLS estimators are consistent. Since the spatial AR(1) model is likely to be misspecified, the QML and 2SLS estimators are not necessarily preferred.

Let $\breve{\varepsilon}_n^{(c)} = \left(I_n - \hat{\rho}_c W_n^{(c)}\right) V_n^{(c)}$, $\breve{\varepsilon}_n = \left(\breve{\varepsilon}_n^{(1)}, ..., \breve{\varepsilon}_n^{(p)}\right)$ and $\hat{\Sigma} = n^{-1}\breve{\varepsilon}_n' \breve{\varepsilon}_n$. Define

$$\hat{A}_{cd} = \left[\frac{1}{n}\hat{V}_n^{(c)\prime}(I_n - \hat{\rho}_c W^{(c)})'(I_n - \hat{\rho}_d W_n^{(d)})\hat{V}_n^{(d)}\right]\left[(I_n - \hat{\rho}_c W_n^{(c)})^{-1}\right]\left[(I_n - \hat{\rho}_d W_n^{(d)})^{-1}\right]' \tag{8}$$

where its (i,j)-th element is denoted by $\hat{a}_{ij}^{(cd)}$ for $i, j = 1, \ldots, n$. Then, we estimate g_{cd} and $g_{cd}^{(q)}$ by

$$\hat{g}_{cd} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{a}_{ij}^{(cd)}, \hat{g}_{cd}^{(q)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{a}_{ij}^{(cd)}\left(d_{ij,n}^{*}\right)^{q}. \tag{9}$$

Consequently, the data dependent bandwidth parameter estimator, $\hat{d}_n$, based on the spatial AR(1) model is

$$\hat{d}_n = \arg\min_{d_n}\frac{1}{d_n^{2q}}K_q^2\left(\text{vec }\hat{g}^{(q)}\right)' S_n\left(\text{vec }\hat{g}^{(q)}\right) + \bar{K}\frac{\ell_n}{n}\text{tr}\left(S_n(I + K_{pp})(\hat{g}\otimes\hat{g})\right). \tag{10}$$

For spatial MA(1) and spatial ARMA(1,1) models, (8) is restated as

$$\hat{A}_{cd} = \left[\frac{1}{n}\hat{V}_n^{(c)\prime}\left(I_n + \hat{\lambda}_c M_n^{(c)\prime}\right)^{-1}(I_n + \hat{\lambda}_d M_n^{(d)})^{-1}\hat{V}_n^{(d)}\right]\left(I_n + \hat{\lambda}_c M_n^{(c)}\right)(I_n + \hat{\lambda}_d M_n^{(d)\prime})$$

$$\hat{A}_{cd} = \left[\frac{1}{n}\hat{V}_n^{(c)\prime}(I_n - \hat{\rho}_c W_n^{(c)\prime})\left(I_n + \hat{\lambda}_d M_n^{(c)\prime}\right)^{-1}(I_n + \hat{\lambda}_d M_n^{(d)})^{-1}(I - \hat{\rho}_d W_n^{(d)})\hat{V}_n^{(d)}\right] \times$$
$$(I_n - \hat{\rho}_c W_n^{(c)})^{-1}(I_n + \hat{\lambda}_c M_n^{(c)})(I_n + \hat{\lambda}_d M_n^{(d)\prime})(I_n - \hat{\rho}_d W_n^{(d)\prime})^{-1}$$

respectively. λ_c and $M_n^{(c)}$ are the coefficient and the $(n \times n)$ weighting matrix for the spatial MA component. Extension to spatial AR(p), spatial MA(q), spatial ARMA(p, q) models for $p, q \geq 2$ is straightforward.

The choice of the weighting matrix S_n is another important problem. A traditional choice suggested by Andrews (1991) is

$$\hat{S}_n = (\hat{B}_n \otimes \hat{B}_n)'\tilde{S}(\hat{B}_n \otimes \hat{B}_n)',$$

where $\tilde{S}$ is a $r^2 \times r^2$ diagonal weighting matrix. For this choice of $\hat{S}_n$, the asymptotic truncated MSE criterion reduces to the asymptotic truncated MSE of $\hat{B}_n\hat{J}_n\hat{B}_n'$ with weighting matrix $\tilde{S}$ provided that $\hat{B}_n - B_n = o_p(E\ell_n/n)$. When $\tilde{S}$ is an identity matrix, we obtain the MSE of the sum of the elements in $\hat{B}_n\hat{J}_n\hat{B}_n'$.

While $\hat{S}_n$ is consistent for the objective we are interested in, as Politis (2007) points out, it yields a single optimal bandwidth for estimating all elements of a covariance matrix but each element has its own individual optimal bandwidth. In particular, the cost of using a single optimal bandwidth increases when the process $V_n^{(s)}$ is significantly different for different s. This is typical in a spatial context. Considering this, we propose using different weighting matrices for different elements of the covariance matrix when V_n has a heterogenous dependence structure. Let $S_{rs,n}$ denote the weighting matrix for estimating $\hat{J}_{rs,n}$. Then, a natural choice of $S_{rs,n}$ is the diagonal matrix in which the element corresponding to $\hat{J}_{rs,n}$ is 1 and others are zero. We can also choose the weighting matrix such that the asymptotic truncated MSE criterion reduces to the asymptotic truncated MSE of a subvector of the parameter estimator $\hat{\theta}$.

One concern of this method is that it does not guarantee $\hat{J}_n$ to be psd, which is often regarded as a desirable property of $\hat{J}_n$. However, we can attain positive semi-definiteness with a simple modification suggested by Politis (2007). As $\hat{J}_n$ is symmetry, $\hat{J}_n(\hat{d}_n) = \hat{U}\hat{\Lambda}\hat{U}'$, where $\hat{U}$ is an orthogonal matrix and $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\hat{J}_n$. Let $\hat{\Lambda}^+ = \text{diag}(\hat{\lambda}_1^+, \ldots, \hat{\lambda}_p^+)$ where $\hat{\lambda}_s^+ = \max(\hat{\lambda}_s, 0)$. Then, we define our modified estimator as

$$\hat{J}_n(\hat{d}_n)^+ = \hat{U}\hat{\Lambda}^+\hat{U}'.$$

As each eigenvalue of $\hat{J}_n(\hat{d}_n)^+$ is nonnegative, it is psd. Theorem 4.1 in Politis (2007) shows that $\hat{J}_n(\hat{d}_n)^+$ converges J_n at the same rate as $\hat{J}_n(\hat{d}_n)$. In fact, it is not hard to show that the truncated AMSE of $\hat{J}_n(\hat{d}_n)^+$ is smaller than that of $\hat{J}_n$.

5 Properties of Data Dependent Bandwidth Parameter Estimators

In this section, we consider the consistency condition, rate of convergence, and asymptotic truncated MSE of spatial HAC estimators with the data dependent bandwidth parameter estimator. Let

$$\ddot{g}_{cd} = P\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{a}_{ij}^{(cd)}, \ddot{g}_{cd}^{(q)} = P\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{a}_{ij}^{(cd)}E\left(d_{ij,n}^{*}\right)^{q}.$$

be the probability limits of $\hat{g}_{cd}$ and $\hat{g}_{cd}^{(q)}$ respectively. Define

$$\ddot{d}_n = \arg\min_{d_n}\frac{1}{d_n^{2q}}K_q^2\left(\text{vec }\ddot{g}^{(q)}\right)' S_n\left(\text{vec }\ddot{g}^{(q)}\right) + \bar{K}\frac{E\ell_n}{n}\text{tr}\left(S_n(I+K_{pp})\left(\ddot{g}\otimes\ddot{g}\right)\right).$$

We study the properties of $\hat{J}_n(\hat{d}_n)$ by investigating $\hat{J}_n(\ddot{d}_n)$ because the asymptotic properties of $\hat{J}_n(\hat{d}_n)$ are equivalent to those of $\hat{J}_n(\ddot{d}_n)$ as stated in Theorem 2 below. For Theorem 2, we introduce the following assumption.

Assumption 10 $\sqrt{n}\left(\frac{\ddot{d}_n}{\hat{d}_n}-1\right) = O_p(1)$.

Since $\ddot{d}_n$ is the probability limit of the parametric plug-in estimator $\hat{d}_n$, the assumption holds if $\hat{g}_{cd}$ and $\hat{g}_{cd}^{(q)}$ converge to $\ddot{g}_{cd}$ and $\ddot{g}_{cd}^{(q)}$ respectively at the parametric rate. This is a rather weak assumption.

Let $\hat{\ell}_{i,n} = \sum_{j=1}^{n}1\left(d_{ij,n}^{*}\leq\hat{d}_n\right)$, $\ddot{\ell}_{i,n} = \sum_{j=1}^{n}1\{d_{ij,n}\leq\ddot{d}_n\}$, $\hat{\ell}_n = n^{-1}\sum_{i=1}^{n}\hat{\ell}_{i,n}$ and $\ddot{\ell}_n = n^{-1}\sum_{i=1}^{n}\ddot{\ell}_{i,n}$. The next theorem summarizes the properties of the spatial HAC estimator with $\hat{d}_n$.

Theorem 3 *Suppose Assumptions 1-10 hold.*

(a) $\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\hat{J}_n(\hat{d}_n)-J_n\right) = O_p(1)$ *and* $\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\hat{J}_n(\hat{d}_n)-\hat{J}_n(\ddot{d}_n)\right) = o_p(1)$.

(b) *Let* $\ddot{\tau} = \lim_{n\to\infty}\frac{\ddot{d}_n^{2q}E\ddot{\ell}_n}{n}$. *Then,*

$$\begin{aligned}
&\lim_{h\to\infty}\lim_{n\to\infty}MSE_h\left(\frac{n}{E\ddot{\ell}_n},\hat{J}_n(\hat{d}_n),S_n\right)\\
&=\lim_{h\to\infty}\lim_{n\to\infty}MSE_h\left(\frac{n}{E\ddot{\ell}_n},\hat{J}_n(\ddot{d}_n),S_n\right)\\
&=\frac{1}{\ddot{\tau}}K_q^2\left(vecg^{(q)}\right)' S\left(vecg^{(q)}\right)+\bar{K}tr\left(S(I+K_{pp})\left(g\otimes g\right)\right).
\end{aligned}$$

Proofs are given in the appendix. Theorem 2(a) implies that $\hat{J}_n(\hat{d}_n)\xrightarrow{p}J_n$ as long as $E\ddot{\ell}_n = o(n)$ and $\hat{J}_n(\hat{d}_n)$ and $\hat{J}_n(\ddot{d}_n)$ have the same asymptotic properties. If the approximating parametric model is correct, that is, $\hat{g}\xrightarrow{p}g$ and $\hat{g}^{(q)}\xrightarrow{p}g^{(q)}$, $\{\hat{d}_n\}$ has some optimality properties as a result of Theorem 1(d) and Corollary 1.

Corollary 4 *Suppose Assumptions 1-10 hold. Assume that* $E\ell_n = \alpha_n d_n^{\eta}$ *for some* $\eta > 0$ *and* $\alpha_n = \alpha + o(1)$. *Then for any sequence of data dependent bandwidth estimators* $\{\dot{d}_n\}$ *such that for some fixed sequence,* $\{d_n\}$, *which satisfies* $\lim_{n\to\infty}\frac{d_n^{2q}E\ell_n}{n}\to\tau\in(0,\infty)$ *we have*

$$\lim_{h\to\infty}\lim_{n\to\infty}\left(MSE_h\left(n^{2q/(2q+\eta)},\hat{J}_n(\dot{d}_n),S_n\right)-MSE_h\left(n^{2q/(2q+\eta)},\hat{J}_n(d_n),S_n\right)\right)=0,$$

$\hat{d}_n$ *is preferred in the sense that*

$$\lim_{h\to\infty}\lim_{n\to\infty}\left(MSE_h\left(n^{2q/(2q+\eta)},\hat{J}_n(\dot{d}_n),S_n\right)-MSE_h\left(n^{2q/(2q+\eta)},\hat{J}_n(\hat{d}_n),S_n\right)\right)\geq 0.$$

The inequality is strict unless $d_n = d_n^\star + o(n^{1/(2q+\eta)})$.

6 Monte Carlo Simulation

In this section, we study the properties of the spatial HAC estimator with Monte Carlo simulation. First, we compare the performance of the spatial HAC estimator based on $\hat{d}_n$ with other bandwidth selection procedures and the heteroskedasticity robust covariance estimator of White (1980). We evaluate them using the MSE criterion and the coverage accuracy of the associated CIs. Second, we examine the robustness of our bandwidth choice procedure to the mis-specification in the spatial weighting matrices and the approximating parametric model. We also examine its robustness to the presence of measurement errors in distance. Third, for studentized tests, we compare the normal approximation with some naive bootstrap approximations. Fourth, we evaluate the performance of the spatial HAC estimator with bandwidth parameter $\hat{d}_n$ when the units are distributed irregularly on the lattice. Finally, we use different weighting matrices in the MSE criterion and evaluate the effect of the resulting bandwidth choice on the MSE of a standard error estimator.

The data generating process we consider here is

$$y_n = X_n\theta_0 + u_n \tag{11}$$

$$u_n = \rho_0 W_{0n} u_n + \varepsilon_n,\ |\rho_0| < 1, \tag{12}$$

with $\varepsilon_{i,n} \overset{i.i.d.}{\sim} N(0,1)$. We assume a lattice structure, in which each unit is located on a square grid of integers. W_{0n} is a contiguity matrix and units i and j are neighbors if $d_{ij,n} \leq \sqrt{2}$. Following convention, it is row-standardized and its diagonal elements are zero.

We consider three different sizes of lattices, 20×20 ($n = 300, 400$), 25×25 ($n = 400$) and 32×32 ($n = 1024$). The ranges of d_n we consider are from 1 to 27 for the 20×20 lattice, from 1 to 34 for the 25×25 lattice and from 1 to 44 for the 32×32 lattice. We use a location model in the first part and a univariate regression model in the second part. The estimand of interest is the covariance matrix of $\sqrt{n}(\hat{\theta}-\theta_0)$. We use the Parzen kernel, which is defined as follows:

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

6.1 Location Model

For the location model, model (11) reduces to

$$y_{i,n} = \theta_0 + u_{i,n}.$$

Without loss of generality, we set $\theta_0 = 1$. A natural estimator of θ_0 is $\hat{\theta} = n^{-1}\sum_{i=1}^n y_{i,n}$ and $\hat{u}_n = y_n - \hat{\theta}$.

We use the spatial AR(1) as the approximating parametric model. The concentrated log-likelihood function for the spatial AR(1) process is

$$\log L(\hat{u}_n|\rho) = -\frac{n}{2}\log\left(\hat{u}_n - \rho W_n \hat{u}_n\right)'\left(\hat{u}_n - \rho W_n \hat{u}_n\right) + \log|I_n - \rho W_n| + const.$$

See Lee (2004). For a given spatial weighting matrix W_n, we estimate ρ by the QML method, that is

$$\hat{\rho} = \hat{\rho}\left(W_n\right) = \arg\max_{\rho} \log L(\hat{u}_n|\rho).$$

Depending on the choice of W_n, we obtain a different $\hat{\rho}$ and hence a different bandwidth parameter $\hat{d}_n(W_n)$ from equation (10). To find $\hat{d}_n$, we search the minimizer numerically instead of using the plug-in version of (7). In our simulation experiment, we take W_n to be the contiguity matrix in which units i and j are neighbors if $d_{ij,n} \leq \mathcal{D}$, a threshold parameter. We consider three values for the threshold: $\mathcal{D} = 1, \sqrt{2}, 2$, leading to three bandwidth choices $\hat{d}_n^{(\ell)}, \hat{d}_n$ and $\hat{d}_n^{(h)}$. Note that when $\mathcal{D} = \sqrt{2}$, the spatial weighting matrix is equal to the true spatial weighting matrix W_{0n}. We also consider the case with measure errors in distance. When $d_{ij,n} > 1$, we take $P(\nu_{ij,n} = -1) = P(\nu_{ij,n} = 0) = P(\nu_{ij,n} = 1) = 1/3$. We use the contiguity matrix as the weighting matrix and take the threshold parameter to be $\sqrt{2}$. This gives us the data driven bandwidth estimator $\hat{d}^{(e)}$.

KP suggest taking $d_n^{KP} = [n^{\frac{1}{4}}]$, where $[z]$ denotes the largest integer that is less than or equal to z. We compare the performances of $\hat{d}_n^{(\ell)}, \hat{d}_n$, $\hat{d}_n^{(h)}$ and $\hat{d}^{(e)}$ with that of d_n^{KP}. We also include the heteroskedasticity consistent estimator of White (1980, 1984) in our comparison. The estimator is defined to be

$$\text{INID} = \frac{1}{n-1}\sum_{i=1}^{n} \hat{u}_{i,n}^2,$$

which can be regarded as the SHAC estimator with bandwidth set to be 0.

Table 1 presents the ratio of the MSE of the spatial HAC estimator with different bandwidth choices to the spatial HAC estimator with the infeasible finite sample optimal bandwidth $\tilde{d}_n$. It also reports the average bandwidth choice in each scenario. As in the time series case, $\hat{d}_n$ performs much better with positive spatial dependence than with negative spatial dependence. When ρ is positive, the ratio is usually less than 1.20 even with incorrect W_n and measurement errors. When ρ is negative, the ratio is higher than 1.20 for all cases. Table 1 also illustrates how mis-specification in the spatial weighting matrix affects our choice of bandwidth. If W_n includes fewer units as neighbors, the bandwidth estimator tends to be smaller than the one with correct W_n. In contrast, if W_n includes more units as neighbors, the bandwidth estimator tends to be larger. This coincides with our intuition. If we think we have a larger neighborhood, we need to choose a larger bandwidth to reflect the dependence structure.

Table 1 also presents the performance of the spatial HAC estimator with measurement errors. The effects of measurement errors are related to the mis-specification of W_n. For a given bandwidth parameter, positive measurement errors lead to a smaller number of neighbors and vice versa. Whereas, in contrast to the mis-specification of W_n, measurement errors are different across different individuals. Table 1 shows that the estimator contaminated by measurement errors performs very poorly compared to other estimators when ρ is negative, while it performs reasonably well when ρ is positive.

Table 1 also compares $\hat{d}_n$ with d_n^{KP}. As d_n^{KP} depends only on the sample size, it is invariant to the spatial dependence. Thus, it performs relatively well when it is close to $\tilde{d}_n$ (e.g. $\rho = -0.3$) but it is inferior to $\hat{d}_n$ in most senarios.

Table 2 provides the bias, variance and MSE of the spatial HAC estimators with different bandwidth selection and those of INID. We use SHAC_0, SHAC_l, SHAC_h, SHAC_e and SHAC_{KP} to denote the spatial HAC estimators with $\hat{d}_n$, $\hat{d}_n^{(l)}$, $\hat{d}_n^{(h)}$, $\hat{d}_n^{(e)}$ and $\hat{d}_n^{KP}$ respectively. We can see that SHAC_0 is reasonably accurate in general but that it suffers from severe underestimation when ρ is extremely high. Spatial HAC estimators do not capture high dependence well even if we choose a large bandwidth since spatial HAC estimators are constructed with the estimated residuals not the true disturbances. Our asymptotic theory does not capture the effect of demeaning on the SHAC estimator. This is analogous to the time series case, see for example, Sun, Phillips and Jin (2008) and Sun and Phillips (2008).

When there is no spatial dependence ($\rho = 0$), SHAC_0 is quite reliable in that the RMSE is only 12% of the true value even though INID is slightly more accurate. When there exists some spatial dependence, SHAC_0 is much more accurate than INID. Furthermore, INID is rarely improved with an increasing sample size, which is in sharp contrast to SHAC_0. For example, when $\rho = 0.3$ and $n = 400$, the MSE of SHAC_0 is less than a third of that of INID. When $n = 1024$, the difference increases with the former less than a fifth of the latter. Therefore, when there is no spatial dependence, the loss of efficiency from using a spatial HAC estimator with data dependent bandwidth is small. Whereas, there is a remarkable reduction in RMSE by using a spatial HAC estimator when there exists spatial dependence.

Table 2 also shows how mis-specification in W_n and measurement errors affect the performance of the spatial HAC estimator using the bandwidth choice we suggest. Comparing SHAC_e with SHAC_0, we find that measurement errors lead to higher MSE. However, the difference in MSE is not very large, reflecting the robustness of the SHAC to the presence of measurement errors. Similarly, mis-specification in W_n is not critical in our simulation design. Among the three bandwidth choices $\hat{d}_n^{(l)}$, $\hat{d}_n^{(h)}$, $\hat{d}_n^{(e)}$, none of them performs consistently better than others and the difference gets smaller when $n = 1024$. Compared to SHAC_{KP}, all of them tend to yield smaller MSEs especially when $n = 400$ and ρ is high.

Table 3 reports the empirical coverage probabilities of CIs associated with different spatial HAC estimators. The results in this table are similar to the ones in Table 2. All of the estimators yield very accurate CIs when there is no spatial dependence. In contrast, when there is spatial dependence, INID is clearly inferior to spatial HAC estimators. As the sample size increases, the coverage accuracy improves for all of the estimators except INID. Compared to SHAC_{KP}, spatial HAC estimators using our data dependent bandwidth choice are more reliable as the dependence increases even in the presence of measurement errors or mis-specification in the spatial weighting matrix.

Table 3 shows that, when $\rho = 0.9$ or 0.95, the error in coverage probability (ECP) is substantial. For example, when $\rho = 0.95$, the ECP for the 95% CI with SHAC_0 is 16.2% even when $n = 1024$. As seen in Table 2, the downward bias of spatial HAC estimators becomes very large when spatial dependence is very high. For this reason, the CIs tend to be very tight. The ECP comes from two sources. First, the spatial HAC estimator is biased downward. Second, the CIs are based on the asymptotic normal approximation. In order to alleviate this problem, we investigate the performance of some bootstrap procedures in Table 5.

Table 4 shows the performance of $\hat{d}_n$ with misspecified parametric models. As the parametric plug-in method is likely to biased, robustness of the spatial HAC estimator to the mis-specification of the approximating parametric model is a highly desirable property. Consider the case that u_n follows a SAR(p) process:

$$u_n = \rho W_{n1} u_n + \rho^2 W_{n2} u_n + \cdots + \rho^p W_{np} u_n + \varepsilon_n.$$

The thresholds for W_{1n}, W_{2n}, W_{3n} and W_{4n} are $d_{ij,n} \le \sqrt{2}$, $\sqrt{2} < d_{ij,n} \le 2$, $2 < d_{ij,n} < \sqrt{5}$ and $\sqrt{5} < d_{ij,n} \le 2\sqrt{2}$ respectively. Regardless of the number of lags the true process has, we use spatial AR(1) as the approximating parametric model. Table 4 illustrates that as the number of lags increases, the accuracy of the spatial HAC estimator using the spatial AR(1) model becomes lower. However, comparison with d_n^{KP} clearly shows that the plug-in method using spatial AR(1) model performs reasonably well. For example, when $\rho = 0.4$ and the DGP is SAR(4) the empirical coverage probability of the 99% CI with SHAC_0 is 91.9% and that with SHAC_{KP} is 86.5%.

Table 5 examines bootstrap approximation as an alternative to the normal approximation. Both i.i.d. *naive* bootstrap and wild bootstrap are considered. The procedure for the i.i.d. *naive* bootstrap we use here is as follows:

(S.1) At each location i, draw $y_{i,n}^*$ randomly from $\{y_{i,n}, i = 1, \ldots, n\}$ with replacement.

(S.2) Estimate the model parameter θ by $\hat{\theta}^* = n^{-1} \sum y_{i,n}^*$.

(S.3) Construct the spatial HAC estimator based on the bootstrap sample but use the bandwidth parameter $\hat{d}_n$.

(S.4) Compute the t-stat in the bootstrap world.

(S.5) Repeat S.1-S.4 to obtain the empirical distribution of the bootstrapped t-stat.

(S.6) Use critical values from the empirical distribution in (S.5) to construct CIs.

We also implement the wild bootstrap, which is proposed by Liu (1988) to account for unknown form of heteroskedasticity. The procedure is the same as that for the iid bootstrap except that (S.1) is replaced by (W.1)

(W.1) At each location, compute the residual $\hat{u}_{i,n} = y_{i,n} - \hat{\theta}$ and generate the bootstrap observation $y_{i,n}^*$:

$$y_{i,n}^* = \begin{cases} \hat{\theta} + \hat{u}_{i,n} & \text{with probability 0.5,} \\ \hat{\theta} - \hat{u}_{i,n} & \text{with probability 0.5.} \end{cases}$$

See Davidson and Flachaire (2001) for more details.

(S.1) and (W.1) eliminate spatial dependence of the bootstrap sample. Gonçalves and Vogelsang (2008) show that the i.i.d. *naive* bootstrap provides a valid approximation to the "fixed-b" asymptotic distribution in time series regressions. Under the "fixed-b" specification, the bandwidth is set proportional to the sample size and the associated test statistic converges to a non-standard limiting distribution (e.g. Kiefer and Vogelsang (2002, 2005)). Gonçalves and Vogelsang (2008) introduce a naive bootstrap procedure to obtain the critical values from the non-standard distribution. Bester, Conley, Hansen and Vogelsang (2008) have extended the "fixed-b" asymptotics and the naive bootstrap procedure to spatial HAC estimation. Their results are not applicable to our setting for two reasons. First, we adopt the traditional asymptotics framework in which the bandwidth or the number of pseudo-neighbors grows at a slower rate than the sample size. Second, the spatial processes we consider allow for nonstationarity and heteroskedasticity which are ruled out in Bester, Conley, Hansen and Vogelsang (2008). However, the idea of using bootstrap to capture the randomness of the HAC estimator is still applicable. When the bandwidth is large, the bias of the HAC estimator is small and the main task is to capture the finite sample variation of the HAC estimator. By ignoring the spatial dependence hence the bias of the HAC estimator, the iid bootstrap and wild bootstrap do exactly this.

The bootstrap method can be justified in the traditional framework. Under some regularity assumptions and $E\ell_n = o(n)$, the t-statistic or Wald statistic converges in distribution to the standard normal distribution or a chi-square distribution. In the bootstrap world, the corresponding test statistic obviously converges to the same distribution. Therefore, the iid bootstrap and wild bootstrap can be viewed as a valid method to obtain critical values from the standard normal or Chi-square distribution. Whether the critical values are second order correct, however, is beyond the scope of this paper.

Table 5 shows that the bootstrap methods implemented here improve the accuracy of the CIs compared to the standard normal approximation, especially when the dependence is extremely high. As we have seen in previous tables, the standard normal approximation yields a large size distortion when spatial dependence is very high. However, we don't find this problem from the bootstrap procedures. Between the i.i.d. *naive* bootstrap and the wild bootstrap, there is no significant difference. For example, when $\rho = 0.95$, the empirical coverage probabilities of the 95% CI by the i.i.d. *naive* bootstrap and the wild bootstrap are 87.0% and 84.9% respectively, while that of CLT is 65.9%.

Table 6 illustrates the performance of the spatial HAC estimator with $\hat{d}_n$ when the units are located irregularly on the lattice. We generate u_n using the spatial AR(1) process on 20×20 and 25×25 lattices and randomly sample 300 and 400 locations from the lattices respectively without replacement. We estimate the location model with the observations on those 300 and 400 locations. We condition on the same set of locations we sample in each simulation. Table 6 shows that irregularity in location does not adversely affect the performance of the spatial HAC estimators with $\hat{d}_n$. The result is confirmed by comparing Table 6 with Tables 2 and 3 in which the observations are regularly spaced. This corroborates our asymptotic results as they do not require a regular lattice structure.

6.2 Univariate Model

In the second part, the regression model we consider is

$$y_{i,n} = \alpha + \beta x_{i,n} + u_{i,n}$$

where $\alpha = 1$, $\beta = 5$, $x_n = (x_{i,n})$ is the standardized version of $\tilde{x}_n$, which follows a spatial process of the form:

$$\tilde{x}_n = \psi W_{0n} \tilde{x}_n + \zeta_n,$$

with $\zeta_{in} \overset{i.i.d}{\sim} U[0,1]$. Here we assume the spatial process of $\tilde{x}_n$ and u_n have the same weighting matrix W_{0n}. Let X_n be the design matrix with i-th row $X_{i,n} = [1, x_{i,n}]$. In view of the standardization, $n^{-1} X_n' X_n$ is the 2×2 identity matrix.

We consider two different weighting matrices: $S_n = \check{S}_n$ or $\hat{S}_n$ where

$$\check{S}_n = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \hat{S}_n = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The first choice $\check{S}_n$ is suggested by Andrews (1991) in time series HAC estimation. For this choice, the MSE criterion reduces to the MSE of $\hat{J}_{11,n} + \hat{J}_{22,n} + 2\hat{J}_{12,n}$. The second choice is designed to select the variance of $\hat{\beta}$ and the corresponding MSE is the MSE of $\hat{J}_{22,n}$.

Table 7 reports the bias and MSE of the SHAC estimator $\hat{J}_{22,n}$ for the above two weighting matrices. The coverage probability of the associated 95% CI is also reported. When $\psi = 0.3$, $\hat{S}_n$ always yields more accurate $\hat{J}_{22,n}$ if $\rho > 0$. In the case that ρ is very high, the reduction in MSE and improvement in coverage accuracy by using $\hat{S}_n$ over $\check{S}_n$ are remarkable. For example, when $\rho = 0.95$ and $n = 400$, the MSE of $\hat{J}_{22,n}$ with weighting matrix $\hat{S}_n$ is 27.23 while that with $\check{S}_n$ is 50.10. The empirical coverage probability of the CI with $\hat{S}_n$ is 91.2% and that with $\check{S}_n$ is 76.6%. When $n = 1024$, the difference is still very large but become less dramatic. When $\psi = 0.9$, $\hat{S}_n$ performs better than $\check{S}_n$ in most cases although the margin of improvement is small.

7 Conclusion

In this paper, we study the asymptotic properties of the spatial HAC estimator. We establish the consistency conditions, the rate of convergence and the asymptotic truncated MSE of the estimator. We also determine the optimal bandwidth parameter which minimizes the asymptotic truncated MSE. As this optimal bandwidth parameter is not feasible in practice, we suggest a data dependent bandwidth parameter estimator using a parametric plug-in method. Monte Carlo simulation results show that the data dependent bandwidth choice we suggest performs reasonably well compared to other bandwidth selection procedures in terms of both the MSE criterion and the coverage accuracy of CIs. They also confirm the robustness of our bandwidth choice procedure to the mis-specification in the spatial weighting matrix and the approximating parametric model, irregularity and sparsity in spatial locations, and the presence of measurement errors.

Instead of using the asymptotic truncated MSE criterion, we can study the optimal bandwidth selection based on a criterion that is most suited for hypothesis testing or CI construction. It is interesting to extend the methods by Sun, Phillips and Jin (2008) and Sun and Phillips (2008) on time series HAC estimation to the spatial setting.

Table 1: Ratio of the MSE of Spatial HAC Estimators with Different Bandwidths to the MSE of the Spatial HAC Estimator with Finite Sample Optimal Bandwidth, $\tilde{d}_n$

	ρ							
	-0.5	-0.3	0	0.3	0.5	0.7	0.9	0.95
$\hat{d}_n$	1.43	1.49	3.36	1.09	1.06	1.04	1.21	1.14
	(4.8)	(4.2)	(2.7)	(5.1)	(6.9)	(9.3)	(21.9)	(26.8)
$\hat{d}_n^{(l)}$	2.31	1.92	2.60	1.14	1.11	1.06	1.05	1.12
	(4.2)	(3.6)	(2.4)	(4.4)	(6.0)	(8.1)	(14.7)	(25.3)
$\hat{d}_n^{(h)}$	1.24	1.28	5.09	1.14	1.12	1.15	1.32	1.14
	(5.5)	(4.8)	(3.4)	(5.9)	(8.3)	(12.1)	(26.9)	(27.0)
$\hat{d}_n^{(e)}$	7.22	2.27	5.27	1.22	1.19	1.21	1.33	1.14
	(3.2)	(3.4)	(3.1)	(6.0)	(8.6)	(12.7)	(27.0)	(27.0)
d_n^{KP}	2.40	1.40	4.44	1.20	1.68	2.07	1.99	1.73
$d_n^{\star}$	1.31	1.33	1.00	1.02	1.00	1.01	1.33	1.14
	(4.9)	(4.1)	(1.0)	(4.9)	(7.0)	(9.8)	(27.0)	(27.0)
$\tilde{d}_n$	(6.3)	(5.3)	(1.3)	(5.4)	(7.1)	(9.2)	(13.7)	(16.6)

Notes: (1) sample size $n = 400$. (2) $d_n^{KP} = 4$ when $n = 400$. (3) number in parenthesis represents the average value of bandwidth choice.

Table 2: Bias, Variance, and MSE of Spatial HAC Estimators with Different Bandwidth Choices in a Location Model with Spatial AR(1) Error

	Bias		Variance		MSE (RMSE/True Value)			
Estimator	n=400	n=1024	n=400	n=1024	n=400		n=1200	
				$\rho = 0$				
$SHAC_0$	-0.008	-0.004	0.015	0.007	0.015	(0.12)	0.007	(0.08)
$SHAC_l$	-0.006	-0.003	0.012	0.006	0.012	(0.11)	0.006	(0.07)
$SHAC_h$	-0.015	-0.006	0.023	0.011	0.023	(0.15)	0.011	(0.10)
$SHAC_e$	-0.014	-0.006	0.024	0.012	0.024	(0.15)	0.012	(0.11)
$SHAC_{KP}$	-0.018	-0.020	0.020	0.019	0.020	(0.14)	0.020	(0.14)
INID	-0.001	0.001	0.005	0.002	0.005	(0.07)	0.002	(0.04)
				$\rho = 0.3$				
$SHAC_0$	-0.399	-0.323	0.140	0.080	0.299	(0.27)	0.184	(0.21)
$SHAC_l$	-0.448	-0.375	0.113	0.064	0.314	(0.27)	0.205	(0.22)
$SHAC_h$	-0.346	-0.271	0.193	0.109	0.313	(0.27)	0.182	(0.21)
$SHAC_e$	-0.561	-0.283	0.207	0.118	0.337	(0.28)	0.198	(0.22)
$SHAC_{KP}$	-0.519	-0.322	0.060	0.068	0.329	(0.28)	0.172	(0.20)
INID	-1.001	-1.000	0.005	0.002	1.006	(0.49)	1.000	(0.49)
				$\rho = 0.5$				
$SHAC_0$	-1.066	-0.846	0.805	0.465	1.941	(0.35)	1.180	(0.27)
$SHAC_l$	-1.186	-0.965	0.631	0.363	2.039	(0.36)	1.294	(0.28)
$SHAC_h$	-0.960	-0.732	1.125	0.641	2.046	(0.36)	1.178	(0.27)
$SHAC_e$	-0.986	-0.753	1.196	0.688	2.169	(0.37)	1.255	(0.28)
$SHAC_{KP}$	-1.703	-1.128	0.174	0.217	3.074	(0.44)	1.490	(0.30)
INID	-2.859	-2.855	0.008	0.004	8.182	(0.71)	8.153	(0.71)
				$\rho = 0.7$				
$SHAC_0$	-3.900	-3.063	9.004	5.409	24.21	(0.44)	14.79	(0.34)
$SHAC_l$	-4.195	-3.421	7.149	4.043	24.75	(0.44)	15.75	(0.36)
$SHAC_h$	-3.753	-2.714	12.747	7.811	26.83	(0.46)	15.18	(0.35)
$SHAC_e$	-3.896	-2.767	12.992	8.114	28.17	(0.47)	15.77	(0.36)
$SHAC_{KP}$	-6.876	-5.054	0.871	1.219	48.15	(0.62)	26.76	(0.46)
INID	-9.731	-9.705	0.024	0.010	94.71	(0.87)	94.20	(0.87)
				$\rho = 0.9$				
$SHAC_0$	-60.21	-40.81	871.2	708.4	4496.6	(0.66)	2374.1	(0.48)
$SHAC_l$	-55.92	-43.28	781.5	566.4	3908.4	(0.61)	2439.2	(0.49)
$SHAC_h$	-62.90	-50.72	958.3	886.6	4914.3	(0.69)	3459.0	(0.58)
$SHAC_e$	-63.17	-50.48	945.6	869.2	4936.6	(0.69)	3417.0	(0.58)
$SHAC_{KP}$	-85.73	-73.70	26.7	42.7	7376.6	(0.84)	5474.0	(0.73)
INID	-98.49	-97.92	0.5	0.2	9700.0	(0.97)	9587.7	(0.97)
				$\rho = 0.95$				
$SHAC_0$	-285.0	-221.5	10590	11099	91790	(0.74)	60140	(0.61)
$SHAC_l$	-283.6	-209.1	10372	10443	90810	(0.74)	54150	(0.57)
$SHAC_h$	-285.0	-247.2	10606	13318	91850	(0.74)	74400	(0.67)
$SHAC_e$	-285.9	-247.6	10469	13245	92210	(0.74)	74560	(0.67)
$SHAC_{KP}$	-373.4	-339.8	215	365	139610	(0.92)	115810	(0.84)
INID	-402.3	-399.4	4	1	161880	(0.99)	159490	(0.99)

Table 3: Empirical Coverage Probabilities of Nominal 99%, 95% and 90% CIs Constructed Using Spatial HAC Estimators in a Location Model with Spatial AR(1) Error

	99%		95%		90%	
Estimator	n=400	n=1024	n=400	n=1024	n=400	n=1024
			$\rho = 0$			
SHAC_0	98.8	99.1	95.2	95.1	90.5	90.7
SHAC_l	98.8	99.1	95.3	95.1	90.9	90.6
SHAC_h	98.8	99.1	95.2	95.2	90.4	90.4
SHAC_e	99.0	99.1	95.2	94.9	90.6	90.0
SHAC_{KP}	98.9	98.9	95.4	94.9	90.5	90.0
INID	98.9	99.0	95.8	95.7	91.3	91.3
			$\rho = 0.3$			
SHAC_0	97.6	98.2	91.9	93.1	86.7	86.7
SHAC_l	97.2	98.1	91.9	92.7	86.6	86.1
SHAC_h	97.8	98.4	92.1	93.5	87.3	87.2
SHAC_e	97.8	98.0	91.9	93.3	86.4	86.9
SHAC_{KP}	97.2	98.2	91.4	93.2	85.7	86.8
INID	94.0	94.0	84.6	84.0	76.4	77.6
			$\rho = 0.5$			
SHAC_0	96.8	97.7	90.6	91.8	83.5	85.4
SHAC_l	96.5	97.3	90.3	91.4	82.6	84.8
SHAC_h	96.7	97.6	90.7	92.0	83.4	86.2
SHAC_e	96.7	97.6	90.2	91.5	83.1	85.9
SHAC_{KP}	95.0	96.9	87.4	90.7	79.1	83.6
INID	84.3	83.8	70.8	72.3	63.9	64.6
			$\rho = 0.7$			
SHAC_0	95.3	96.9	87.7	89.4	79.7	83.4
SHAC_l	95.0	96.4	86.9	89.0	79.8	82.5
SHAC_h	95.5	96.9	86.3	89.8	80.0	83.6
SHAC_e	95.1	96.8	85.8	89.4	79.5	83.3
SHAC_{KP}	89.6	94.2	77.8	85.3	69.2	78.1
INID	66.5	66.7	53.3	54.2	46.2	47.7
			$\rho = 0.9$			
SHAC_0	84.5	93.3	73.3	84.2	66.7	78.8
SHAC_l	87.2	93.3	77.3	84.4	69.6	78.3
SHAC_h	81.5	89.3	71.0	79.4	64.1	74.0
SHAC_e	81.4	89.5	70.8	79.8	64.0	74.6
SHAC_{KP}	68.8	82.1	56.8	70.4	48.9	62.5
INID	35.2	37.4	27.2	28.8	23.2	24.8
			$\rho = 0.95$			
SHAC_0	77.0	87.9	65.9	78.8	59.7	72.2
SHAC_l	77.5	89.5	66.7	80.8	59.9	73.6
SHAC_h	76.9	83.4	65.9	73.2	59.7	66.6
SHAC_e	76.9	83.4	65.9	73.1	59.7	66.6
SHAC_{KP}	55.1	70.3	43.6	59.5	36.3	52.2
INID	24.5	25.8	18.6	17.8	15.0	14.8

Table 4: Performance of the Spatial HAC Estimator with Bandwidth $\hat{d}_n$ under Misspecified Approximating Parametric Model

	SAR(2)		SAR(3)		SAR(4)	
	SHAC_0	SHAC_{KP}	SHAC_0	SHAC_{KP}	SHAC_0	SHAC_{KP}
			$\rho = 0.2$			
Bias	-0.357	-0.400	-0.385	-0.431	-0.392	-0.437
MSE	0.208	0.204	0.249	0.231	0.238	0.237
	(0.26)	(0.26)	(0.27)	(0.27)	(0.27)	(0.27)
99%	97.8	97.5	97.4	97.3	97.4	97.3
95%	92.0	91.6	91.8	91.5	91.8	91.5
90%	87.1	86.3	86.5	85.9	86.5	85.8
			$\rho = 0.3$			
Bias	-0.795	-1.020	-0.962	-1.231	-1.115	-1.307
MSE	0.880	1.123	1.222	1.607	1.498	1.802
	(0.35)	(0.39)	(0.37)	(0.43)	(0.39)	(0.44)
99%	96.5	95.9	96.2	95.1	95.9	94.7
95%	90.1	88.0	89.9	87.5	89.7	87.1
90%	83.3	80.6	82.2	79.3	82.2	78.6
			$\rho = 0.4$			
Bias	-1.999	-2.850	-3.204	-4.495	-3.971	-5.480
MSE	5.141	8.335	12.40	20.52	18.59	30.41
	(0.44)	(0.56)	(0.49)	(0.64)	(0.52)	(0.67)
99%	94.1	91.9	93.3	88.6	91.9	86.5
95%	87.0	81.4	84.2	77.0	82.6	73.6
90%	80.2	73.4	77.3	68.6	75.4	66.9
			$\rho = 0.5$			
Bias	-8.067	-11.95	-41.95	-57.33	-202.8	-249.9
MSE	78.65	144.0	1988.0	3294.3	43653	62499
	(0.55)	(0.74)	(0.68)	(0.88)	(0.79)	(0.95)
99%	91.2	81.1	83.1	63.6	72.3	43.8
95%	81.4	68.3	72.6	50.2	62.6	34.5
90%	74.4	60.7	65.8	42.8	53.4	28.1

Note: (1) Number in parenthesis represents the ratio of the RMSE to the true value. (2) $d_n^{KP} = 4$.

Table 5: Empirical Coverage Probability of Nominal 99%, 95%, 90% Confidence Intervals Constructed Using the Bootstrap and Standard Normal Approximations

	Normal	i.i.d. Bootstrap	Wild Bootstrap
		$\rho = 0.0$	
99%	98.8	99.0	98.9
95%	95.2	95.4	95.5
90%	90.5	90.8	91.4
		$\rho = 0.3$	
99%	97.6	97.9	98.1
95%	91.9	92.5	93.1
90%	86.7	87.8	88.1
		$\rho = 0.5$	
99%	96.8	97.7	97.8
95%	90.6	91.8	92.0
90%	83.5	86.2	86.2
		$\rho = 0.7$	
99%	95.3	96.8	97.2
95%	87.7	90.6	90.6
90%	79.7	83.6	83.9
		$\rho = 0.9$	
99%	84.5	96.0	95.4
95%	73.3	88.6.	θ87.3
90%	66.7	82.5	81.2
		$\rho = 0.95$	
99%	77.0	95.0	94.2
95%	65.9	87.0	84.9
90%	59.7	80.8	77.4

Table 6: Performance of the Spatial HAC Estimator with Bandwidth $\hat{d}_n$ in the Presence of Irregulairty and Sparsity in Spatial Locations

	$n = 300, (20 \times 20)$		$n = 400, (25 \times 25)$	
	SHAC$_0$	SHAC$_{KP}$	SHAC$_0$	SHAC$_{KP}$
		$\rho = 0$		
Bias	-0.011	-0.021	-0.007	-0.027
MSE	0.015	0.022	0.010	0.023
	(0.12)	(0.15)	(0.10)	(0.15)
99%	99.1	98.8	99.1	99.1
95%	95.2	95.0	96.1	96.0
90%	90.1	89.5	92.8	92.0
		$\rho = 0.3$		
Bias	-0.323	-0.364	-0.270	-0.229
MSE	0.195	0.186	0.126	0.108
	(0.25)	(0.24)	(0.22)	(0.20)
99%	97.6	97.5	97.6	97.9
95%	92.3	91.7	92.3	86.7
90%	86.6	86.2	86.7	86.8
		$\rho = 0.5$		
Bias	-0.773	-1.107	-0.613	-0.667
MSE	1.013	1.364	0.591	0.588
	(0.32)	(0.37)	(0.28)	(0.28)
99%	96.6	95.7	98.0	97.9
95%	90.0	88.5	92.3	92.3
90%	84.3	80.8	85.2	85.0
		$\rho = 0.7$		
Bias	-2.340	-4.204	-1.737	-2.564
MSE	9.67	18.29	4.845	7.222
	(0.40)	(0.55)	(0.35)	(0.42)
99%	94.0	89.9	96.3	94.9
95%	87.2	80.4	88.9	84.6
90%	80.3	70.8	80.5	77.1
		$\rho = 0.9$		
Bias	-27.75	-49.62	-18.48	-32.66
MSE	1189.7	2480.2	507.7	1084.5
	(0.55)	(0.80)	(0.48)	(0.70)
99%	87.7	69.6	90.8	78.3
95%	77.6	57.1	80.4	63.8
90%	70.3	50.4	71.1	55.4
		$\rho = 0.95$		
Bias	-139.8	-213.2	-87.4	-145.4
MSE	25120	45588	10210	21268
	(0.66)	(0.89)	(0.57)	(0.82)
99%	80.6	55.2	85.2	62.3
95%	69.2	44.6	72.6	49.6
90%	62.4	38.5	63.9	42.4

Note: Number in parenthesis represents the ratio of RMSE to the true value

Table 7: Bias and MSE of $\hat{J}_{22,n}$ and Empirical Coverage of the Associated CIs with Bandwidth Selected Using Different Weighting Matrices in the Truncated AMSE Criterion

				ρ					
n	ψ			0	0.3	0.5	0.7	0.9	0.95
400	0.3	$\hat{S}_n$	Bias	-0.016	-0.056	-0.110	-0.235	-0.922	-2.036
			MSE	0.021	0.032	0.062	0.208	3.909	27.23
			95%	93.8	93.8	93.2	92.6	91.5	91.2
		$\check{S}_n$	Bias	-0.012	-0.063	-0.121	-0.271	-2.060	-5.290
			MSE	0.017	0.044	0.109	0.406	8.535	50.10
			95%	94.0	93.0	92.1	91.0	82.5	76.6
	0.9	$\hat{S}_n$	Bias	-0.035	-0.311	-0.721	-2.021	-13.89	-41.41
			MSE	0.035	0.226	0.998	7.140	321.7	2859.4
			95%	93.1	90.7	89.7	86.8	82.4	78.8
		$\check{S}_n$	Bias	-0.026	-0.318	-0.724	-2.009	-14.92	-48.84
			MSE	0.028	0.219	0.967	7.013	355.4	3540.7
			95%	93.2	91.1	89.9	87.1	79.2	70.5
1024	0.3	$\hat{S}_n$	Bias	-0.006	-0.044	-0.090	-0.190	-0.734	-1.661
			MSE	0.009	0.014	0.031	0.105	1.783	12.06
			95%	95.2	95.2	94.8	94.6	94.0	93.0
		$\check{S}_n$	Bias	-0.004	-0.041	-0.073	-0.165	-0.876	-3.1326
			MSE	0.007	0.022	0.058	0.210	3.503	24.15
			95%	95.2	95.1	94.3	94.0	92.4	88.4
	0.9	$\hat{S}_n$	Bias	-0.018	-0.245	-0.548	-1.507	-10.24	-30.42
			MSE	0.014	0.127	0.567	4.092	191.0	1780.9
			95%	95.9	94.1	92.7	91.4	88.7	86.9
		$\check{S}_n$	Bias	-0.013	-0.247	-0.540	-1.456	- 9.731	-29.58
			MSE	0.011	0.122	0.544	3.986	195.9	1976.7
			95%	95.8	94.1	93.0	91.6	89.2	85.7

APPENDIX

Proof of Theorem 1

For notational simplicity, we re-order the individuals and make new indices. For $i_{(j)} = 1, ..., \ell_{j,n}$, $d^*_{i_{(j)}j,n} \leq d_n$, and for $i_{(j)} = \ell_{j+1,n}, \ldots, n$, $d^*_{i_{(j)}j,n} > d_n$.

(a) Asymptotic Variance: $\lim_{n\to\infty} \frac{n}{E\ell_n} var\left(\text{vec}\left(\tilde{J}_n\right)\right) = \bar{K}(I + K_{pp})(g \otimes g)$

Let $\varphi_{lkrs,n} = \sum_{i=1}^n \sum_{j=1}^n r^{(r)}_{il,n} r^{(s)}_{jk,n} K\left(\frac{d^*_{ij,n}}{d_n}\right)$ and $N_n = \{\nu_{ij,n} | i, j = 1, \ldots, n\}$ be the set of measurement errors in distance. By Assumption 5(i), we have

$$
\begin{aligned}
&\frac{n}{E\ell_n} cov\left(\tilde{J}_{rs,n}, \tilde{J}_{cd,n}\right) \\
&= \frac{1}{nE\ell_n} E\left[E\left(\left(\sum_{l=1}^{np}\sum_{k=1}^{np} \varphi_{lkrs,n}(\varepsilon_{l,n}\varepsilon_{k,n} - E\varepsilon_{l,n}\varepsilon_{k,n})\right)\left(\sum_{e=1}^{np}\sum_{f=1}^{np} \varphi_{efcd,n}(\varepsilon_{e,n}\varepsilon_{f,n} - E\varepsilon_{e,n}\varepsilon_{f,n})\right)\Big| N_n\right)\right] \\
&= \frac{1}{nE\ell_n} E\Bigg[\Bigg(\sum_{l=1}^{np}\sum_{k=1}^{np}\sum_{e=1}^{np}\sum_{f=1}^{np} \varphi_{lkrs,n}\varphi_{efcd,n}(\varepsilon_{l,n}\varepsilon_{k,n}\varepsilon_{e,n}\varepsilon_{f,n} - \varepsilon_{l,n}\varepsilon_{k,n}E\varepsilon_{e,n}\varepsilon_{f,n} - \varepsilon_{e,n}\varepsilon_{f,n}E\varepsilon_{l,n}\varepsilon_{k,n} \\
&\quad + E\varepsilon_{l,n}\varepsilon_{k,n}E\varepsilon_{e,n}\varepsilon_{f,n})\Big| N_n\Bigg)\Bigg] \\
&= \frac{n}{E\ell_n} E\left[\frac{1}{n^2}\left(\sum_{l=1}^{np} \varphi_{llrs,n}\varphi_{llcd,n}\left(E\varepsilon^4_{l,n} - 3\right) + \sum_{l=1}^{np}\sum_{k=1}^{np} \varphi_{lkrs,n}\varphi_{lkcd,n} + \sum_{l=1}^{np}\sum_{k=1}^{np} \varphi_{lkrs,n}\varphi_{klcd,n}\right)\right] \\
&:= C_{1,n} + C_{2,n} + C_{3,n},
\end{aligned}
$$

where

$$
\begin{aligned}
C_{1,n} &= \frac{1}{nE\ell_n}\sum_{l=1}^{np} E\left[\left(\sum_{i=1}^n\sum_{j=1}^n r^{(r)}_{il,n} r^{(s)}_{jl,n} K\left(\frac{d^*_{ij,n}}{d_n}\right)\right)\left(\sum_{a=1}^n\sum_{b=1}^n r^{(c)}_{al,n} r^{(d)}_{bl,n} K\left(\frac{d^*_{ab,n}}{d_n}\right)\right)\right]\left(E\varepsilon^4_{l,n} - 3\right), \\
C_{2,n} &= \frac{1}{nE\ell_n}\sum_{l=1}^{np}\sum_{k=1}^{np} E\left[\left(\sum_{i=1}^n\sum_{j=1}^n r^{(r)}_{il,n} r^{(s)}_{jk,n} K\left(\frac{d^*_{ij,n}}{d_n}\right)\right)\left(\sum_{a=1}^n\sum_{b=1}^n r^{(c)}_{al,n} r^{(d)}_{bk,n} K\left(\frac{d^*_{ab,n}}{d_n}\right)\right)\right], \\
C_{3,n} &= \frac{1}{nE\ell_n}\sum_{l=1}^{np}\sum_{k=1}^{np} E\left[\left(\sum_{i=1}^n\sum_{j=1}^n r^{(r)}_{il,n} r^{(s)}_{jk,n} K\left(\frac{d^*_{ij,n}}{d_n}\right)\right)\left(\sum_{a=1}^n\sum_{b=1}^n r^{(c)}_{ak,n} r^{(d)}_{bl,n} K\left(\frac{d^*_{ab,n}}{d_n}\right)\right)\right].
\end{aligned}
$$

$C_{1,n}$ can be restated as

$$
\frac{1}{E\ell_n}\sum_{i=1}^n\sum_{j=1}^n\sum_{a=1}^n\sum_{b=1}^n E\left[K\left(\frac{d^*_{ij,n}}{d_n}\right)K\left(\frac{d^*_{ab,n}}{d_n}\right)\right]\left(\frac{1}{n}\sum_{l=1}^{np} r^{(r)}_{il,n} r^{(s)}_{jl,n} r^{(c)}_{al,n} r^{(d)}_{bl,n}\left(E\varepsilon^4_{l,n} - 3\right)\right).
$$

Therefore,

$$
\begin{aligned}
|C_{1,n}| &\le \frac{1}{nE\ell_n}\sum_{l=1}^{np}\left|E\varepsilon_{l,n}^4-3\right|\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}E\left[K\left(\frac{d^*_{ij,n}}{d_n}\right)K\left(\frac{d^*_{ab,n}}{d_n}\right)\right]\left|r^{(r)}_{il,n}r^{(s)}_{jl,n}r^{(c)}_{al,n}r^{(d)}_{bl,n}\right| \\
&\le \frac{1}{nE\ell_n}\sum_{l=1}^{np}\left|E\varepsilon_{l,n}^4-3\right|\left(\sum_{i=1}^{n}\left|r^{(r)}_{il,n}\right|\right)\left(\sum_{j=1}^{n}\left|r^{(s)}_{jl,n}\right|\right)\left(\sum_{a=1}^{n}\left|r^{(c)}_{al,n}\right|\right)\left(\sum_{b=1}^{n}\left|r^{(d)}_{bl,n}\right|\right) \\
&\le \frac{c_R^4}{E\ell_n}\frac{1}{n}\sum_{l=1}^{np}\left|E\varepsilon_{l,n}^4-3\right| \le \frac{c_R^4 c_E p}{E\ell_n}=o(1)
\end{aligned}
$$

using Assumptions 1 and 2.

$C_{2,n}$ can be restated as

$$
\begin{aligned}
&\frac{1}{nE\ell_n}E\left[\sum_{l=1}^{np}\sum_{k=1}^{np}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}r^{(r)}_{il,n}r^{(s)}_{jk,n}r^{(c)}_{al,n}r^{(d)}_{bk,n}K\left(\frac{d^*_{ij,n}}{d_n}\right)K\left(\frac{d^*_{ab,n}}{d_n}\right)\right] \\
&=\frac{1}{nE\ell_n}E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}K\left(\frac{d^*_{ij,n}}{d_n}\right)K\left(\frac{d^*_{ab,n}}{d_n}\right)\left(\sum_{l=1}^{np}r^{(r)}_{il,n}r^{(c)}_{al,n}\right)\left(\sum_{k=1}^{np}r^{(s)}_{jk,n}r^{(d)}_{bk,n}\right)\right] \\
&=\frac{1}{nE\ell_n}E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}}K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right]. \qquad \text{(A.1)}
\end{aligned}
$$

In order to prove that (A.1) converges to $\bar{K}g_{rc}g_{sd}$, it suffices to first show that

$$
\begin{aligned}
&\lim_{n\to\infty}\frac{1}{nE\ell_n}E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}}K^2\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right] \\
&=\lim_{n\to\infty}\frac{1}{nE\ell_n}E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}}K^2\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right] \\
&=\bar{K}g_{rc}g_{sd}, \qquad \text{(A.2)}
\end{aligned}
$$

and then show that

$$
\begin{aligned}
&\lim_{n\to\infty}\frac{1}{nE\ell_n}E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}}K^2\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right] \\
&=\lim_{n\to\infty}\frac{1}{nE\ell_n}E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}}K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right]. \qquad \text{(A.3)}
\end{aligned}
$$

Let $\gamma^{(sd)}_{\bar{\imath},b_{(a)},n} = \frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}} \gamma^{(sd)}_{j_{(i)}b_{(a)},n}$. Then,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{nE\ell_n} E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}} K^2\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right] \\
&= \lim_{n\to\infty} \frac{1}{nE\ell_n} E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}} K^2\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{\bar{\imath},b_{(a)},n}\right] \\
&+ \lim_{n\to\infty} \frac{1}{nE\ell_n} E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}} K^2\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\left(\gamma^{(sd)}_{j_{(i)}b,n} - \gamma^{(sd)}_{\bar{\imath},b_{(a)},n}\right)\right]. \qquad \text{(A.4)}
\end{aligned}$$

For the first term in (A.4), under the Assumption 7(i),

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{nE\ell_n} E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}} K^2\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{\bar{\imath},b_{(a)},n}\right] \\
&= \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{n}\gamma^{(rc)}_{ia,n}\sum_{b_{(a)}=1}^{\ell_{a,n}}\gamma^{(sd)}_{\bar{\imath},b_{(a)},n}\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}} K^2\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)\right] \\
&= \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{n}\gamma^{(rc)}_{ia,n}\right) E\left(\frac{1}{E\ell_n}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{j_{(i)}=1}^{\ell_{i,n}}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right)\left[\frac{1}{E\ell_n}\sum_{j=1}^{n} K^2\left(\frac{d^*_{ij,n}}{d_n}\right)\right]. \qquad \text{(A.5)}
\end{aligned}$$

Note that

$$\begin{aligned}
&\left|\frac{1}{E\ell_n}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{j_{(i)}=1}^{\ell_{i,n}}\gamma^{(sd)}_{j_{(i)}b_{(a)},n} - \frac{1}{E\ell_n}\sum_{b_{(a)}=1}^{E\ell_n}\sum_{j_{(i)}=1}^{E\ell_n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right| \le \left|\frac{1}{E\ell_n}\sum_{b_{(a)}=\ell_{a,n}+1}^{E\ell_n}\sum_{j_{(i)}=1}^{E\ell_n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right| \qquad \text{(A.6)} \\
&+ \left|\frac{1}{E\ell_n}\sum_{b_{(a)}=1}^{E\ell_n}\sum_{j_{(i)}=\ell_{i,n}+1}^{E\ell_n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right| + \left|\frac{1}{E\ell_n}\sum_{b_{(a)}=\ell_{a,n}+1}^{E\ell_n}\sum_{j_{(i)}=\ell_{i,n}+1}^{E\ell_n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right|.
\end{aligned}$$

We proceed to show that the expected value of each term is $o(1)$. We consider the first term only as the proofs for the other two terms are similar. By the Markov inequality,

$$P\left(\frac{|\ell_{a,n} - E\ell_n|}{E\ell_n} \ge \varepsilon\right) \le \frac{1}{\varepsilon}E\left|\frac{\ell_{a,n}}{E\ell_n} - 1\right| \to 0$$

using Assumption 6 (ii). That is, for any $\varepsilon > 0$, there exists a $N_0 > 0$ such that for $n \ge N_0$

$$P\left(\ell_{a,n} \notin B(E\ell_n, \varepsilon)\right) \le \varepsilon,$$

where $B(E\ell_n,\varepsilon) = (\lfloor (1-\varepsilon)\, E\ell_n \rfloor\,,\ \ \lceil (1+\varepsilon)\, E\ell_n \rceil)$. Now

$$\begin{aligned}
&E\left|\frac{1}{E\ell_n}\sum_{b_{(a)}=\ell_{a,n}+1}^{E\ell_n}\sum_{j_{(i)}=1}^{E\ell_n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right| \le E\frac{1}{E\ell_n}\sum_{b_{(a)}=\ell_{a,n}+1}^{E\ell_n}\left(\sum_{j_{(i)}=1}^{E\ell_n}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right|\right) \\
&= E\left(\frac{1}{E\ell_n}\sum_{b_{(a)}=\ell_{a,n}+1}^{E\ell_n}\left(\sum_{j_{(i)}=1}^{E\ell_n}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right|\right)\middle|\, \ell_{a,n}\in B(E\ell_n,\varepsilon)\right)P\left(\ell_{a,n}\in B(E\ell_n,\varepsilon)\right) \\
&+ E\left(\frac{1}{E\ell_n}\sum_{b_{(a)}=\ell_{a,n}+1}^{E\ell_n}\left(\sum_{j_{(i)}=1}^{E\ell_n}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right|\right)\middle|\, \ell_{a,n}\notin B(E\ell_n,\varepsilon)\right)P\left(\ell_{a,n}\notin B(E\ell_n,\varepsilon)\right) \\
&\le 2\varepsilon\left[\frac{1}{2\varepsilon E\ell_n}\sum_{b_{(a)}=\lfloor(1-\varepsilon)E\ell_n\rfloor}^{\lceil(1+\varepsilon)E\ell_n\rceil}\left(\sum_{j_{(i)}=1}^{E\ell_n}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right|\right)\right] + O(1)P\left(\ell_{a,n}\notin B(E\ell_n,\varepsilon)\right),
\end{aligned}$$

which can be made arbitrarily small when $n\to\infty$. So the first term in (A.6) is indeed $o_p(1)$. Hence

$$E\left|\frac{1}{E\ell_n}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{j_{(i)}=1}^{\ell_{i,n}}\gamma^{(sd)}_{j_{(i)}b_{(a)},n} - \frac{1}{E\ell_n}\sum_{b_{(a)}=1}^{E\ell_n}\sum_{j_{(i)}=1}^{E\ell_n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right| = o(1)\,. \tag{A.7}$$

Since $(E\ell_n)^{-1}\sum_{j=1}^{n}K^2\left(d^*_{ij,n}/d_n\right) = \left(\ell_{i,n}/E\ell_n\right)\ell_{i,n}^{-1}\sum_{j=1}^{n}K^2\left(d^*_{ij,n}/d_n\right)$ is bounded, we also have

$$E\left[\left|\frac{1}{E\ell_n}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{j_{(i)}=1}^{\ell_{i,n}}\gamma^{(sd)}_{j_{(i)}b_{(a)},n} - \frac{1}{E\ell_n}\sum_{b_{(a)}=1}^{E\ell_n}\sum_{j_{(i)}=1}^{E\ell_n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right|\frac{1}{E\ell_n}\sum_{j=1}^{n}K^2\left(\frac{d^*_{ij,n}}{d_n}\right)\right] = o(1).$$

As a result

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{nE\ell_n}E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}}K^2\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{\bar{\imath},b_{(a)},n}\right] \\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{n}\gamma^{(rc)}_{ia,n}E\left[\left(\frac{1}{E\ell_n}\sum_{b_{(a)}=1}^{E\ell_n}\sum_{j_{(i)}=1}^{E\ell_n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right)\left(\frac{1}{E\ell_n}\sum_{j=1}^{n}K^2\left(\frac{d^*_{ij,n}}{d_n}\right)\right)\right] \\
&= \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{n}\gamma^{(rc)}_{ia,n}\right)\lim_{n\to\infty}\left(\frac{1}{E\ell_n}\sum_{b_{(a)}=1}^{E\ell_n}\sum_{j_{(i)}=1}^{E\ell_n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right)\lim_{n\to\infty}E\left[\frac{1}{E\ell_n}\sum_{j=1}^{n}K^2\left(\frac{d^*_{ij,n}}{d_n}\right)\right] \\
&= \bar{K}g_{rc}g_{sd},
\end{aligned}$$

using Assumption 7(ii).

For the second term in (A.4), we have

$$
\begin{aligned}
&\left|\frac{1}{nE\ell_n}E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}}K^2\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\left(\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\overline{\imath},b_{(a)},n}\right)\right]\right| \\
&=\left|\frac{1}{nE\ell_n}E\left[\sum_{i=1}^{n}\sum_{a=1}^{n}\gamma^{(rc)}_{ia,n}\sum_{j_{(i)}=1}^{\ell_{i,n}}K^2\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)\sum_{b_{(a)}=1}^{\ell_{a,n}}\left(\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\overline{\imath},b_{(a)},n}\right)\right]\right| \\
&=\left|\frac{1}{n}E\left[\sum_{i=1}^{n}\sum_{a=1}^{n}\gamma^{(rc)}_{ia,n}\left(\frac{1}{\ell_{i,n}}\sum_{j_{(i)}=1}^{\ell_{i,n}}K^2\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)\right)\frac{\ell_{i,n}}{E\ell_n}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left(\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\overline{\imath},b_{(a)},n}\right)\right]\right| \\
&\leq C\frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{n}\left|\gamma^{(rc)}_{ia,n}\right|E\left|\sum_{b_{(a)}=1}^{\ell_{a,n}}\left(\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\overline{\imath},b_{(a)},n}\right)\right| \qquad \text{(A.8)}
\end{aligned}
$$

by Assumption 6. Note that for some generic constant C

$$
\begin{aligned}
&\left|\sum_{b_{(a)}=1}^{\ell_{a,n}}\left(\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\overline{\imath},b_{(a)},n}\right)\right| \\
&\leq\left|\sum_{b=1}^{n}\gamma^{(sd)}_{j_{(i)}b,n}1\left\{d^*_{ab,n}<d_n\right\}\right|+\left|\frac{1}{E\ell_n}\sum_{b=1}^{n}\sum_{j=1}^{n}\gamma^{(sd)}_{jb,n}1\left\{d^*_{ab,n}<d_n,d^*_{ij,n}<d_n\right\}\right| \\
&\leq\sum_{b=1}^{n}\left|\gamma^{(sd)}_{j_{(i)}b,n}\right|+\sup_b\sum_{j=1}^{n}\left|\gamma^{(sd)}_{jb,n}\right|\frac{1}{E\ell_n}\sum_{b=1}^{n}1\left\{d^*_{ab,n}<d_n\right\} \\
&\leq\sum_{b=1}^{n}\left|\gamma^{(sd)}_{j_{(i)}b,n}\right|+\left(\sup_b\sum_{j=1}^{n}\left|\gamma^{(sd)}_{jb,n}\right|\right)\frac{\ell_{a,n}}{E\ell_n}\leq C+C\frac{\ell_{a,n}}{E\ell_n}.
\end{aligned}
$$

By Assumption 6(i), $E\lim_{n\to\infty}\frac{\ell_{a,n}}{E\ell_n}\leq\lim_{n\to\infty}\frac{E\ell_{a,n}}{E\ell_n}\leq C$. Invoking the dominated convergence theorem and Assumption 4 yields

$$
\lim_{n\to\infty}E\left|\sum_{b_{(a)}=1}^{\ell_{a,n}}\left(\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\overline{\imath},b_{(a)},n}\right)\right|=EP\lim_{n\to\infty}\left|\sum_{b_{(a)}=1}^{\ell_{a,n}}\left(\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\overline{\imath},b_{(a)},n}\right)\right|.
$$

Using the same argument for proving (A.7) and combining the result with Assumption 4, we deduce that $P\lim_{n\to\infty}\left|\sum_{b_{(a)}=1}^{\ell_{a,n}}\left(\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\overline{\imath},b_{(a)},n}\right)\right|=0$. Hence the second term in (A.4) is $o_p(1)$.

By a symmetric argument, we obtain the result that

$$
\lim_{n\to\infty}\frac{1}{nE\ell_n}E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}}K^2\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{jb,n}\right]=\bar{K}g_{rc}g_{sd}.
$$

The next step is to prove (A.3). In view of previous derivations, it suffices to show that

$$
\lim_{n\to\infty}E\left[\frac{1}{nE\ell_n}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left[K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)-K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\right]^2\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{jb,n}\right]=0. \qquad \text{(A.9)}
$$

But

$$
\begin{aligned}
&E\left[\frac{1}{nE\ell_n}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left[K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)-K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\right]^2\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{jb,n}\right]\\
&=E\left[\frac{1}{nE\ell_n}\sum_{(i,j(i),a,b(a))\in\mathcal{I}_1}\left[K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)-K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\right]^2\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{jb,n}\right]\\
&+\frac{1}{nE\ell_n}\sum_{(i,j(i),a,b(a))\in\mathcal{I}_2}\left[K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)-K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\right]^2\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{jb,n}\\
&:=F_{1n}+F_{2n},
\end{aligned}
$$

where

$$
\mathcal{I}_1=\left\{(i,j(i),a,b(a)):\left|d^*_{ij_{(i)},n}-d^*_{ab_{(a)},n}\right|\le 2c_n\right\},
$$

and

$$
\mathcal{I}_2=\left\{(i,j(i),a,b(a)):\left|d^*_{ij_{(i)},n}-d^*_{ab_{(a)},n}\right|> 2c_n\right\}.
$$

For F_{1n}, we have

$$
\begin{aligned}
F_{1n}&\le E\left|\frac{1}{nE\ell_n}\sum_{(i,j(i),a,b(a))\in\mathcal{I}_1}\left[K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)-K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\right]^2\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{jb,n}\right|\\
&\le\frac{c_L^2}{nE\ell_n}\sum_{(i,j(i),a,b(a))\in\mathcal{I}_1}\left|\frac{d^*_{ij_{(i)},n}}{d_n}-\frac{d^*_{ab_{(a)},n}}{d_n}\right|^2\left|\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{jb,n}\right|\\
&\le\frac{4c_L^2c_n^2}{d_n^2}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{n}\left|\gamma^{(rc)}_{ia,n}\right|\right)\left(\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left|\gamma^{(rc)}_{jb,n}\right|\right)=O\left(\frac{c_n^2}{d_n^2}\right)
\end{aligned}
$$

using equation (A.7). For F_{2n} we note that if $\left|d^*_{ij_{(i)},n}-d^*_{ab_{(a)},n}\right|>2c_n$, then either $d^*_{ia,n}>c_n$ or $d^*_{j_{(i)}b_{(a)},n}>c_n$. Otherwise, if both $d^*_{ia,n}\le c_n$ and $d^*_{j_{(i)}b_{(a)},n}\le c_n$, then

$$
d^*_{ij_{(i)},n}-d^*_{ab_{(a)},n}\le d^*_{ia,n}+d^*_{ab_{(a)},n}+d^*_{b_{(a)}j_{(i)},n}-d^*_{ab_{(a)},n}\le 2c_n,
$$

and

$$
d^*_{ij_{(i)},n}-d^*_{ab_{(a)},n}\ge d^*_{ij_{(i)},n}-d^*_{ia,n}-d^*_{ij_{(i)},n}-d^*_{j_{(i)}b_{(a)},n}\ge -2c_n.
$$

These two inequalities imply that $\left|d^*_{ij_{(i)},n}-d^*_{ab_{(a)},n}\right|\le 2c_n$, a contradiction. Without the loss of generality, we assume that $d^*_{ia,n} > c_n$ for $\left(i, j_{(i)}, a, b_{(a)}\right) \in \mathcal{I}_2$. In this case

$$\begin{aligned}
F_{2n} &\leq E\left|\frac{1}{nE\ell_n}\sum_{(i,j(i),a,b(a))\in\mathcal{I}_2}\left[K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right) - K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\right]^2 \gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right| \\
&\leq E\frac{1}{nE\ell_n}\sum_{(i,j(i),a,b(a))\in\mathcal{I}_2}\left|\left(d^*_{ia,n}\right)^q \gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\left(d^*_{ia,n}\right)^{-q}\right| \\
&= E\frac{1}{nE\ell_n}\sum_{i=1}^{n}\sum_{a:d^*_{ia,n}\in[c_n,d_n]}\left(d^*_{ia,n}\right)^q\left|\gamma^{(rc)}_{ia,n}\right|\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right|\left(d^*_{ia,n}\right)^{-q} \\
&\leq E\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{n}\left(d^*_{ia,n}\right)^q\left|\gamma^{(rc)}_{ia,n}\right|\right)\left(\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right|\right)c_n^{-q} \\
&= o(c_n^{-q}).
\end{aligned}$$

By choosing c_n such that $c_n \to \infty$ but $c_n/d_n \to 0$, we have

$$F_{1n} = o\,(1) \text{ and } F_{2n} = o(1)$$

and (A.3) is proved.

With the same procedure, it is straightforward that $\lim_{n\to\infty} C_{3,n} = \bar{K}g_{rd}g_{sc}$. Therefore,

$$\lim_{n\to\infty}\frac{n}{E\ell_n}cov\left(\tilde{J}_{rs,n}, \tilde{J}_{cd,n}\right) = \bar{K}(g_{rc}g_{sd} + g_{rd}g_{sc}).$$

In terms of matrix form,

$$\lim_{n\to\infty}\frac{n}{E\ell_n}\text{var}\left(\text{vec}\left(\tilde{J}_n\right)\right) = \bar{K}(I + K_{pp})\left(g\otimes g\right),$$

where $g = [g_{rs}]$, $r, s = 1, \ldots, n$.

(b) Asymptotic Bias: $\lim_{n\to\infty} d_n^p(E\tilde{J}_n - J_n)$ when $d_n \to \infty$.

By Assumption 5(ii) and the dominated convergence theorem, we have

$$\begin{aligned}
d_n^q\left(E\tilde{J}_n - J_n\right) &= -E\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma_{ij,n}\left(d^*_{ij,n}\right)^q\left[\frac{1-K\left(\frac{d^*_{ij,n}}{d_n}\right)}{\left(\frac{d^*_{ij,n}}{d_n}\right)^q}\right]\right) \\
&= -K_q\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma_{ij,n}E\left(d^*_{ij,n}\right)^q + o(1).
\end{aligned}$$

Therefore,

$$\lim_{n\to\infty} d_n^q(E\tilde{J}_n - J_n) = -K_q\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma_{ij,n}E\left(d^*_{ij,n}\right)^q = -K_q g^{(q)},$$

where $g^{(q)}_{rs}$ is (r, s)-th element of $g^{(q)}$.

(c) $\sqrt{\frac{n}{E\ell_n}}\left(\hat{J}_n - J_n\right) = O_p(1)$ **and** $\sqrt{\frac{n}{E\ell_n}}\left(\hat{J}_n - \tilde{J}_n\right) = o_p(1)$

By (a) and (b) the first part of (c) is implied by the second part. Therefore, it suffices to show that $\sqrt{\frac{n}{E\ell_n}}\left(\hat{J}_n - \tilde{J}_n\right) = o_p(1)$. This holds if and only if $\sqrt{\frac{n}{E\ell_n}}\left(b'\hat{J}_n b - b'\tilde{J}_n b\right) = o_p(1)$ for any $b \in R^p$. In consequence, we can consider the case that J_n is a scalar random variable without loss of generality. Using a Taylor expansion, we have

$$
\begin{aligned}
&\sqrt{\frac{n}{E\ell_n}}\left(\hat{J}_n - \tilde{J}_n\right) = \sqrt{\frac{n}{E\ell_n}}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{d^*_{ij,n}}{d_n}\right)\left[\hat{V}_{i,n}\hat{V}'_{j,n} - V_{i,n}V'_{j,n}\right] \\
&:= 2L_{1,n}\sqrt{n}\left(\hat{\theta} - \theta_0\right) + \sqrt{n}\left(\hat{\theta} - \theta_0\right)' L_{2,n}\sqrt{n}\left(\hat{\theta} - \theta_0\right) \\
&+ \sqrt{n}\left(\hat{\theta} - \theta_0\right)' L_{3,n}\sqrt{n}\left(\hat{\theta} - \theta_0\right) \qquad \text{(A.10)}
\end{aligned}
$$

where

$$
\begin{aligned}
L_{1,n} &= \sqrt{\frac{E\ell_n}{n}}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} V_{j,n}\left(\frac{1}{E\ell_n}\sum_{i=1}^{n} K\left(\frac{d^*_{ij,n}}{d_n}\right)\frac{\partial}{\partial\theta'}V_{i,n}\right), \\
L_{2,n} &= \frac{1}{\sqrt{nE\ell_n}}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{d^*_{ij,n}}{d_n}\right)\left(\frac{\partial^2}{\partial\theta\partial\theta'}V_{i,n}(\overline{\theta})\right)V_{j,n}(\overline{\theta}), \\
L_{3,n} &= \frac{1}{\sqrt{nE\ell_n}}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{d^*_{ij,n}}{d_n}\right)\left(\frac{\partial}{\partial\theta}V_{i,n}(\overline{\theta})\right)\left(\frac{\partial}{\partial\theta}V_{j,n}(\overline{\theta})\right)'.
\end{aligned}
$$

Therefore, under Assumption 8(i) it suffices to show that $L_{1,n} = o_p(1)$, $L_{2,n} = o_p(1)$ and $L_{3,n} = o_p(1)$.

For $L_{2,n}$, we have, using the Cauchy inequality,

$$
\begin{aligned}
\|L_{2,n}\|^2 &= \left\|\frac{1}{\sqrt{nE\ell_n}}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{d^*_{ij,n}}{d_n}\right)\left(\frac{\partial^2}{\partial\theta\partial\theta'}V_{i,n}(\overline{\theta})\right)V_{j,n}(\overline{\theta})\right\|^2 \\
&\le \left[\frac{1}{\sqrt{nE\ell_n}}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} 1(d^*_{ij,n} \le d_n)\left\|\frac{\partial^2}{\partial\theta\partial\theta'}V_{i,n}(\overline{\theta})\right\|\left\|V_{j,n}(\overline{\theta})\right\|\right]^2 \\
&\le \left[\sqrt{\frac{E\ell_n}{n}}\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}V_{i,n}(\overline{\theta})\right\|\left(\frac{1}{E\ell_n}\sum_{j=1}^{n} 1(d^*_{ij,n} \le d_n)\left\|V_{j,n}(\overline{\theta})\right\|\right)\right]^2 \\
&\le \frac{E\ell_n}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{\theta}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}V_{i,n}(\theta)\right\|^2\right)\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{E\ell_n}\sum_{j=1}^{n} 1(d^*_{ij,n} \le d_n)\left\|V_{j,n}(\overline{\theta})\right\|\right)^2 \\
&= \frac{E\ell_n}{n}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{E\ell_n}\sum_{j=1}^{n} 1(d^*_{ij,n} \le d_n)\left\|V_{j,n}(\overline{\theta})\right\|\right)^2 O_p(1) \qquad \text{(A.11)}
\end{aligned}
$$

where the last equality follows from Assumption 8(iv). By Assumption 8(ii), we have

$$
\begin{aligned}
&P\left(\frac{1}{E\ell_n}\sum_{j=1}^{n}1\left(d_{ij,n}^{*}\leq d_n\right)\left\|V_{j,n}(\overline{\theta})\right\|>\Delta\right)\\
&\leq\frac{1}{\Delta E\ell_n}\sum_{j=1}^{n}E1\left(d_{ij,n}^{*}\leq d_n\right)E\left(\left\|V_{j,n}(\overline{\theta})\right\|\right)\\
&\leq\frac{1}{\Delta}E\left[\frac{\ell_{i,n}}{E\ell_n}\frac{1}{\ell_{i,n}}\sum_{j_{(i)}=1}^{\ell_{i,n}}\left(E\left\|V_{j_{(i)},n}(\overline{\theta})\right\|^{2}\right)^{\frac{1}{2}}\right]\\
&\leq\frac{1}{\Delta}\left(\frac{E\ell_{i,n}}{E\ell_n}\right)\left(\sup_{j}E\sup_{\theta}\left\|V_{j,n}(\theta)\right\|^{2}\right)^{\frac{1}{2}}\to 0,
\end{aligned}
$$

as $\Delta\to\infty$. This implies that

$$
\frac{1}{E\ell_n}\sum_{j=1}^{n}1(d_{ij,n}^{*}\leq d_n)\left\|V_{j,n}(\overline{\theta})\right\|=O_p\left(1\right)
$$

uniformly over i. Hence $L_{2,n}=o_p\left(1\right)$. Using the same procedure, we can show that $L_{3,n}=o_p\left(1\right)$ under Assumption 8(iii).

The next step is to show $L_{1,n}=o_p(1)$. By Markov inequality and Assumption 8(v):

$$
\begin{aligned}
&P\left(\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}V_{j,n}\left(\frac{1}{E\ell_n}\sum_{i=1}^{n}K\left(\frac{d_{ij,n}^{*}}{d_n}\right)\frac{\partial}{\partial\theta'}V_{i,n}\right)\right\|>\delta\right)\\
&\leq\frac{1}{\delta^{2}}E\left[\frac{1}{nE\ell_n^{2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}K\left(\frac{d_{ij,n}^{*}}{d_n}\right)K\left(\frac{d_{ab,n}^{*}}{d_n}\right)V_{j,n}V_{b,n}\frac{\partial}{\partial\theta'}V_{i,n}\frac{\partial}{\partial\theta}V_{a,n}\right]\\
&\leq\frac{1}{\delta^{2}}\frac{1}{nE\ell_n^{2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}\left|E\left(K\left(\frac{d_{ij,n}^{*}}{d_n}\right)K\left(\frac{d_{ab,n}^{*}}{d_n}\right)\right)\right|\left\|E\left(V_{j,n}V_{b,n}\frac{\partial}{\partial\theta'}V_{i,n}\frac{\partial}{\partial\theta}V_{a,n}\right)\right\|\\
&\leq\frac{1}{\delta^{2}}\frac{1}{nE\ell_n^{2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}E\left[1(d_{ij,n}^{*}\leq d_n)1(d_{ab,n}^{*}\leq d_n)\right]\left\|E\left(V_{j,n}V_{b,n}\frac{\partial}{\partial\theta'}V_{i,n}\frac{\partial}{\partial\theta}V_{a,n}\right)\right\|\\
&\leq\frac{1}{\delta^{2}}E\left[\frac{\ell_{j,n}\ell_{b,n}}{E\ell_n^{2}}\frac{1}{n\ell_{j,n}\ell_{b,n}}\sum_{j=1}^{n}\sum_{i_{(j)}=1}^{\ell_{j,n}}\sum_{b=1}^{n}\sum_{a_{(b)}=1}^{\ell_{b,n}}\left\|E\left(V_{j,n}V_{b,n}\frac{\partial}{\partial\theta'}V_{i_{(j)},n}\frac{\partial}{\partial\theta}V_{a_{(b)},n}\right)\right\|\right]\\
&\leq\frac{1}{\delta^{2}}\left(\frac{E\ell_{j,n}\ell_{b,n}}{E\ell_n^{2}}\right)\sup_{i,a,b}\sum_{j=1}^{n}\left\|E\left(V_{j,n}V_{b,n}\frac{\partial}{\partial\theta'}V_{i,n}\frac{\partial}{\partial\theta}V_{a,n}\right)\right\| \qquad\qquad \text{(A.12)}\\
&\leq\frac{C}{\delta^{2}}\frac{\left(E\ell_{j,n}^{2}\right)^{1/2}\left(E\ell_{b,n}^{2}\right)^{1/2}}{E\ell_n^{2}}\leq\frac{C'}{\delta^{2}}
\end{aligned}
$$

where the last inequality follows from Assumption 6. Combining this with the definition of L_{1n}, we obtain $L_{1,n}=O_p\left(\sqrt{\frac{E\ell_n}{n}}\right)=o_p\left(1\right)$.

(d) Asymptotic Truncated MSE

To establish the first and second equalities of Theorem 1(d), we introduce two lemmas from Andrews (1991). For proofs, see Lemmas A1 and A2 in Andrews (1991).

Lemma 5 *If $\{\xi_n\}$ is bounded sequence of random variables such that $\xi_n \xrightarrow{p} 0$, then $E\xi_n \to 0$.*

Lemma 6 *Let $\{X_n\}$ be a sequence of nonnegative rv's for which $\sup_{n\geq 1} EX_n^{1+\delta} < \infty$ for some $\delta > 0$. Then, $\lim_{h\to\infty} \lim_{n\to\infty}(E\min\{X_n, h\} - EX_n) = 0$.*

In our setting,

$$\begin{aligned}
\xi_n &= \min\left\{\frac{n}{E\ell_n}\left|\mathrm{vec}(\hat{J}_n - J_n)' S_n \mathrm{vec}(\hat{J}_n - J_n)\right|, h\right\} \\
&- \min\left\{\frac{n}{E\ell_n}\left|\mathrm{vec}(\tilde{J}_n - J_n)' S_n \mathrm{vec}(\tilde{J}_n - J_n)\right|, h\right\} \\
&= \min\left\{\frac{n}{E\ell_n}\left|\mathrm{vec}(\hat{J}_n - \tilde{J}_n + \tilde{J}_n - J_n)' S_n \mathrm{vec}(\hat{J}_n - \tilde{J}_n + \tilde{J}_n - J_n)\right|, h\right\} \\
&- \min\left\{\frac{n}{E\ell_n}\left|\mathrm{vec}(\tilde{J}_n - J_n)' S_n \mathrm{vec}(\tilde{J}_n - J_n)\right|, h\right\} \\
&= \min\left\{\frac{n}{E\ell_n}\left|\mathrm{vec}(\tilde{J}_n - J_n)' S_n \mathrm{vec}(\tilde{J}_n - J_n)\right| + o_p(1), h\right\} \\
&- \min\left\{\frac{n}{E\ell_n}\left|\mathrm{vec}(\tilde{J}_n - J_n)' S_n \mathrm{vec}(\tilde{J}_n - J_n)\right|, h\right\} \xrightarrow{p} 0,
\end{aligned}$$

as $n \to \infty$. Here the $o_p(1)$ term follows from Theorem 1(c). Also $|\xi_n| \leq h$. By Lemma 5 , $E\xi_n \to 0$. Since this holds for all h, the first equality of Theorem 1(d) holds.

The second equality of Theorem 1(d) is obtained by showing that

$$\lim_{h\to\infty}\lim_{n\to\infty}\left(MSE_h\left(\frac{n}{E\ell_n}, \tilde{J}_n, S_n\right) - MSE_h\left(\frac{n}{E\ell_n}, \tilde{J}_n, S\right)\right) = 0 \tag{A.13}$$

$$\lim_{h\to\infty}\lim_{n\to\infty} MSE_h\left(\frac{n}{E\ell_n}, \tilde{J}_n, S\right) = \lim_{n\to\infty} MSE\left(\frac{n}{E\ell_n}, \tilde{J}_n, S\right). \tag{A.14}$$

Under Assumption 8(ii), (A.13) holds by applying Lemma 5. Equation (A.14) holds by applying Lemma 6 with

$$X_n = \left|\frac{n}{E\ell_n}\mathrm{vec}(\tilde{J}_n - J_n)' S(\tilde{J}_n - J_n)\right|.$$

It is easy to see that $\sup_{n\geq 1} EX_n^2 < \infty$, as required by Lemma 6, if $E\left[\sqrt{\frac{n}{E\ell_n}}(\tilde{J}_{rs,n} - J_{rs,n})\right]^4 = O(1)\ \forall r, s \leq p$. Note that

$$E\left[\sqrt{\frac{n}{E\ell_n}}(\tilde{J}_{rs,n} - J_{rs,n})\right]^4 = E\left[(I_1 + I_2)^4\right].$$

where

$$I_1 = \sqrt{\frac{n}{E\ell_n}} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{d_{ij,n}^*}{d_n}\right) (V_{i,n}^{(r)} V_{j,n}^{(s)} - \gamma_{ij,n}^{(rs)})$$

$$I_2 = \sqrt{\frac{n}{E\ell_n}} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(K\left(\frac{d_{ij,n}^*}{d_n}\right) - 1\right) \gamma_{ij,n}^{(rs)}$$

Therefore, it suffices to show that the following terms are all $O(1)$:

$$D_{1n} = E\left[I_1^4\right],\ D_{2n} = E\left[I_1^3 I_2\right],\ D_{3n} = E\left[I_1^2 I_2^2\right],\ D_{4n} = E\left[I_1 I_2^3\right], D_{5n} = E\left[I_2^4\right].$$

By Theorem 1(a) and (b), it is straightforward to show that D_{3n}, D_{4n} and D_{5n} are $O(1)$ under Assumption 9(ii). The proofs for D_{1n} and D_{2n} are similar and we focus only on D_{1n} here. As denoted before, let $\varphi_{lkrs,n} = \sum_{i=1}^{n} \sum_{j=1}^{n} K(\frac{d_{ij,n}^*}{d_n}) r_{il,n}^{(r)} r_{jk,n}^{(s)}$. Then,

$$\begin{aligned}
D_{1n} &= E\left[\sqrt{\frac{n}{E\ell_n}} \left(\frac{1}{n} \sum_{l=1}^{np} \sum_{k=1}^{np} \varphi_{lkrs,n}(\varepsilon_{l,n}\varepsilon_{k,n} - E\varepsilon_{l,n}\varepsilon_{k,n})\right)\right]^4 \\
&\leq 8E\left[\sqrt{\frac{n}{E\ell_n}} \left(\frac{1}{n} \sum_{l=1}^{np} \varphi_{llrs,n}(\varepsilon_{l,n}^2 - E\varepsilon_{l,n}^2)\right)\right]^4 \\
&+ 8E\left[\sqrt{\frac{n}{E\ell_n}} \left(\frac{1}{n} \sum_{l=1}^{np} \sum_{k\neq l}^{np} \varphi_{lkrs,n}(\varepsilon_{l,n}\varepsilon_{k,n} - E\varepsilon_{l,n}\varepsilon_{k,n})\right)\right]^4 \\
&:= 8G_{1n} + 8G_{2n}
\end{aligned}$$

Let $\tilde{\varepsilon}_{l_1,n}^2 = \varepsilon_{l,n}^2 - E\varepsilon_{l,n}^2$. For G_{1n}, we have:

$$\begin{aligned}
G_{1n} &= \frac{1}{n^2 \left(E\ell_n\right)^2} E \sum_{l_1,l_2,l_3,l_4} \varphi_{l_1 l_1 rs,n}\varphi_{l_2 l_2 rs,n}\varphi_{l_3 l_3 rs,n}\varphi_{l_4 l_4 rs,n} E\tilde{\varepsilon}_{l_1,n}^2 \tilde{\varepsilon}_{l_2,n}^2 \tilde{\varepsilon}_{l_3,n}^2 \tilde{\varepsilon}_{l_4,n}^2 \\
&= \frac{C}{n^2 \left(E\ell_n\right)^2} \left[E\left(\sum_l \varphi_{llrs,n}^4\right) + 3E \sum_{l_1 \neq l_2} \varphi_{l_1 l_1 rs,n}^2 \varphi_{l_2 l_2 rs,n}^2\right] \\
&\leq \frac{C}{n^2 \left(E\ell_n\right)^2} E\left(\sum_l \varphi_{llrs,n}^2\right)^2 = O\left[\frac{1}{\left(E\ell_n\right)^2}\right] = o(1)
\end{aligned}$$

using the fact that

$$\begin{aligned}
\sum_l \varphi_{llrs,n}^2 &= \sum_l \left(\sum_{i=1}^{n} \sum_{j=1}^{n} K(\frac{d_{ij,n}^*}{d_n}) r_{il,n}^{(r)} r_{jl,n}^{(s)}\right)^2 \\
&\leq \sum_l \left(\sum_{i=1}^{n} \left|r_{il,n}^{(r)}\right|\right)^2 \left(\sum_{j=1}^{n} \left|r_{jl,n}^{(r)}\right|\right)^2 = O(n)
\end{aligned}$$

by Assumption 2.

For G_{2n}, we have:

$$G_{2n} = \frac{1}{n^2 \left(E\ell_n\right)^2} E\Bigg[\sum_{l_1 \neq l_2} \sum_{l_3 \neq l_4} \sum_{l_5 \neq l_6} \sum_{l_7 \neq l_8} \varphi_{l_1 l_2 rs,n} \varphi_{l_3 l_4 rs,n} \varphi_{l_5 l_6 rs,n} \varphi_{l_7 l_8 rs,n}$$
$$\times (\varepsilon_{l_1,n}\varepsilon_{l_2,n} - E\varepsilon_{l_1,n}\varepsilon_{l_2,n})(\varepsilon_{l_3,n}\varepsilon_{l_4,n} - E\varepsilon_{l_3,n}\varepsilon_{l_4,n})$$
$$(\varepsilon_{l_5,n}\varepsilon_{l_6,n} - E\varepsilon_{l_5,n}\varepsilon_{l_6,n})(\varepsilon_{l_7,n}\varepsilon_{l_8,n} - E\varepsilon_{l_7,n}\varepsilon_{l_8,n})\Bigg]$$
$$= \frac{1}{n^2 \left(E\ell_n\right)^2} \sum_{l_1 \neq l_2} \sum_{l_3 \neq l_4} \sum_{l_5 \neq l_6} \sum_{l_7 \neq l_8} E\left[\varphi_{l_1 l_2 rs,n} \varphi_{l_3 l_4 rs,n} \varphi_{l_5 l_6 rs,n} \varphi_{l_7 l_8 rs,n}\right]$$
$$\times E[(\varepsilon_{l_1,n}\varepsilon_{l_2,n} - E\varepsilon_{l_1,n}\varepsilon_{l_2,n})(\varepsilon_{l_3,n}\varepsilon_{l_4,n} - E\varepsilon_{l_3,n}\varepsilon_{l_4,n})$$
$$(\varepsilon_{l_5,n}\varepsilon_{l_6,n} - E\varepsilon_{l_5,n}\varepsilon_{l_6,n})(\varepsilon_{l_7,n}\varepsilon_{l_8,n} - E\varepsilon_{l_7,n}\varepsilon_{l_8,n})]$$

where the last equality holds by the independence of $\{\varepsilon_{\ell,n}\}$ from $\{\nu_{ij,n}\}$. Since $\{\varepsilon_{l,n}\}$ is independent and $E\varepsilon_{l,n}^8 < \infty$, it suffices to show that

$$\frac{1}{n^2 \left(E\ell_n\right)^2} E \sum_{l_1=1}^{np} \sum_{l_2 \neq l_1}^{np} \varphi_{l_1 l_2 rs,n}^4 < \infty, \tag{A.15}$$

$$\frac{1}{n^2 \left(E\ell_n\right)^2} E \sum_{l_1=1}^{np} \sum_{l_2=1}^{np} \sum_{l_3=1}^{np} \sum_{l_4=1}^{np} \varphi_{l_1 l_2 rs,n} \varphi_{l_1 l_2 rs,n} \varphi_{l_3 l_4 rs,n} \varphi_{l_3 l_4 rs,n} < \infty, \tag{A.16}$$

$$\frac{1}{n^2 \left(E\ell_n\right)^2} E \sum_{l_1=1}^{np} \sum_{l_2=1}^{np} \sum_{l_3=1}^{np} \sum_{l_4=1}^{np} \varphi_{l_1 l_2 rs,n} \varphi_{l_2 l_4 rs,n} \varphi_{l_4 l_3 rs,n} \varphi_{l_3 l_1 rs,n} < \infty. \tag{A.17}$$

Equation (A.15) is true because

$$\frac{1}{n^2 \left(E\ell_n\right)^2} E \sum_{l_1=1}^{np} \sum_{l_2 \neq l_1}^{np} \varphi_{l_1 l_2 rs,n}^4 = \frac{1}{n^2 \left(E\ell_n\right)^2} \sum_{l_1=1}^{np} \sum_{l_2 \neq l_1}^{np} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} K(\frac{d_{ij,n}^*}{d_n}) r_{il_1,n}^{(r)} r_{jl_2,n}^{(s)} \right)^4$$
$$\leq \frac{1}{n^2 \left(E\ell_n\right)^2} \sum_{l_1=1}^{np} \sum_{l_2 \neq l_1}^{np} \left(\sum_{i=1}^{n} \left| r_{il_1,n}^{(r)} \right| \right)^4 \left(\sum_{j=1}^{n} \left| r_{jl_2,n}^{(s)} \right| \right)^4$$
$$= O\left[\frac{1}{\left(E\ell_n\right)^2}\right]$$

where the last equality follows from Assumption 2.

For equation (A.16), we have

$$\frac{1}{n^2 E\ell_n^2} E \sum_{l_1=1}^{np} \sum_{l_2=1}^{np} \sum_{l_3=1}^{np} \sum_{l_4=1}^{np} \varphi_{l_1 l_2 rs,n} \varphi_{l_1 l_2 rs,n} \varphi_{l_3 l_4 rs,n} \varphi_{l_3 l_4 rs,n}$$
$$= \left(\frac{1}{nE\ell_n} E \sum_{l_1=1}^{np} \sum_{l_2=1}^{np} \varphi_{l_1 l_2 rs,n} \varphi_{l_1 l_2 rs,n} \right)^2$$

and

$$
\begin{aligned}
&\frac{1}{nE\ell_n}E\sum_{l_1=1}^{np}\sum_{l_2=1}^{np}\varphi_{l_1l_2rs,n}\varphi_{l_1l_2rs,n} \\
&=\frac{1}{nE\ell_n}E\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}K\left(\frac{d^*_{ij,n}}{d_n}\right)K\left(\frac{d^*_{ab,n}}{d_n}\right)\left(\sum_{l_1=1}^{np}r^{(r)}_{il_1,n}r^{(r)}_{al_1,n}\right)\left(\sum_{l_2=1}^{np}r^{(s)}_{jl_2,n}r^{(s)}_{bl_2,n}\right) \\
&=\frac{1}{nE\ell_n}E\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}K\left(\frac{d^*_{ij,n}}{d_n}\right)K\left(\frac{d^*_{ab,n}}{d_n}\right)\gamma^{(rr)}_{ia,n}\gamma^{(ss)}_{jb,n}=O(1)
\end{aligned}
$$

using equations (A.2) and (A.3). Hence (A.16) holds.

Finally, for equation (A.17), we note that

$$
\begin{aligned}
&\frac{1}{n^2E\ell_n^2}E\sum_{l_1=1}^{np}\sum_{l_2=1}^{np}\sum_{l_3=1}^{np}\sum_{l_4=1}^{np}\varphi_{l_1l_2rs,n}\varphi_{l_2l_4rs,n}\varphi_{l_4l_3rs,n}\varphi_{l_3l_1rs,n} \\
&=\frac{1}{n^2E\ell_n^2}E\sum_{i,j,a,b}\sum_{o,p,q,m}K\left(\frac{d^*_{ij,n}}{d_n}\right)K\left(\frac{d^*_{ab,n}}{d_n}\right)K\left(\frac{d^*_{op,n}}{d_n}\right)K\left(\frac{d^*_{qm,n}}{d_n}\right)\gamma^{(sr)}_{ja,n}\gamma^{(sr)}_{bo,n}\gamma^{(sr)}_{pq,n}\gamma^{(sr)}_{mi,n} \\
&=\frac{1}{n^2E\ell_n^2}E\sum_{i,a}\sum_{o,q}\left(\sum_{j_{(i)}=1}^{\ell_{i,n}}K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)\gamma^{(sr)}_{j_{(i)}a,n}\right)\left(\sum_{b_{(a)}=1}^{\ell_{a,n}}K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\gamma^{(sr)}_{b_{(a)}o,n}\right) \\
&\times\left(\sum_{p_{(o)}=1}^{\ell_{o,n}}K\left(\frac{d^*_{op_{(o)},n}}{d_n}\right)\gamma^{(sr)}_{p_{(o)}q,n}\right)\left(\sum_{m_{(q)}=1}^{\ell_{q,n}}K\left(\frac{d^*_{qm_{(q)},n}}{d_n}\right)\gamma^{(sr)}_{m_{(q)}i,n}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{j_{(i)}=1}^{\ell_{i,n}}K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)\gamma^{(sr)}_{j_{(i)}a,n}&=\sum_{\left\{j:d^*_{ij,n}<d_n\right\}}K\left(\frac{d^*_{ij,n}}{d_n}\right)\gamma^{(sr)}_{ja,n} \\
&=\sum_{\left\{j:d^*_{ij,n}<d_n,d^*_{ja,n}<d_n\right\}}K\left(\frac{d^*_{ij,n}}{d_n}\right)\gamma^{(sr)}_{ja,n}+\sum_{\left\{j:d^*_{ij,n}<d_n,d^*_{ja,n}\geq d_n\right\}}K\left(\frac{d^*_{ij,n}}{d_n}\right)\gamma^{(sr)}_{ja,n} \\
&=\sum_{j_{(a)}=1}^{\ell_{a,n}}K\left(\frac{d^*_{ij_{(a)},n}}{d_n}\right)\gamma^{(sr)}_{j_{(a)}a,n}+O_p\left(d_n^{-q}\right) \\
&=\sum_{j_{(a)}=1}^{\ell_{a,n}}\gamma^{(sr)}_{j_{(a)}a,n}+O_p\left(d_n^{-q}\right)+\sum_{j_{(a)}=1}^{\ell_{a,n}}\frac{K(d^*_{ij_{(a)},n}/d_n)-1}{\left(d^*_{aj_{(a)},n}/d_n\right)^q}\gamma^{(sr)}_{j_{(a)}a,n}\left(d^*_{aj_{(a)},n}\right)^q\left(d_n^{-q}\right) \\
&=\sum_{j_{(a)}=1}^{\ell_{a,n}}\gamma^{(sr)}_{j_{(a)}a,n}+O_p\left(d_n^{-q}\right)
\end{aligned}
$$

where the $O(\cdot)$ term also satisfies $EO_p\left(d_n^{-q}\right)=O\left(d_n^{-q}\right)$. So

$$
\begin{aligned}
&\frac{1}{n^2E\ell_n^2}E\sum_{l_1=1}^{np}\sum_{l_2=1}^{np}\sum_{l_3=1}^{np}\sum_{l_4=1}^{np}\varphi_{l_1l_2rs,n}\varphi_{l_2l_4rs,n}\varphi_{l_4l_3rs,n}\varphi_{l_3l_1rs,n} \\
&=\frac{1}{n^2E\ell_n^2}\sum_{i,a}\sum_{o,q}\left(\sum_{j_{(a)}=1}^{\ell_{a,n}}\gamma_{j_{(a)}a,n}^{(sr)}\right)\left(\sum_{b_{(o)}=1}^{\ell_{o,n}}\gamma_{b_{(o)}o,n}^{(sr)}\right)\left(\sum_{p_{(q)}=1}^{\ell_{q,n}}\gamma_{p_{(q)}q,n}^{(sr)}\right)\left(\sum_{m_{(i)}=1}^{\ell_{i,n}}\gamma_{m_{(i)}i}^{(sr)}\right)(1+o(1)) \\
&=\left(\frac{1}{nE\ell_n}\sum_{i,a}\sum_{j_{(a)}=1}^{\ell_{a,n}}\sum_{m_{(i)}=1}^{\ell_{i,n}}\gamma_{j_{(a)}a,n}^{(sr)}\gamma_{m_{(i)}i,n}^{(sr)}\right)^2(1+o(1))=O(1)
\end{aligned}
$$

using equations (A.2) and (A.3).

Combining the above proof, we obtain $G_{2n}=O(1)$. Hence $D_{1n}=O(1)$.

For the last equality of Theorem 1(d), since

$$
\frac{n}{E\ell_n}=\frac{d_n^{2q}}{d_n^{2q}E\ell_n/n}=\frac{d_n^{2q}}{\tau+o(1)},
$$

we have

$$
\begin{aligned}
&\lim_{n\to\infty}MSE\left(\frac{n}{E\ell_n},\tilde{J}_n,S_n\right) \\
&=\lim_{n\to\infty}\frac{n}{E\ell_n}\text{vec }\left(E\tilde{J}_n-J_n\right)'S_n\text{vec }\left(E\tilde{J}_n-J_n\right)+\lim_{n\to\infty}\frac{n}{E\ell_n}\bar{K}\text{tr}\left(S_n\text{var}(\text{vec }\tilde{J}_n)\right) \\
&=\frac{1}{\tau}K_q^2\left(\text{vec }g^{(q)}\right)'S\left(\text{vec }g^{(q)}\right)+\bar{K}\text{tr}\left(S(I+K_{pp})(g\otimes g)\right),
\end{aligned}
$$

where the last equality holds by Theorem 1(a) and (b).

Proof of Corollary 1

The proof is very close to the proof of Corollary 1 in Andrews (1991). As

$$
n^{\frac{2q}{2q+\eta}}=\alpha_n^{\frac{2q}{2q+\eta}}\left(\frac{d_n^{2q}E\ell_n}{n}\right)^{\frac{\eta}{2q+\eta}}\frac{n}{E\ell_n}=\alpha_n^{\frac{2q}{2q+\eta}}\left(\tau^{\frac{\eta}{2q+\eta}}+o(1)\right)\frac{n}{E\ell_n},
$$

by Theorem 1(d), we obtain

$$
\begin{aligned}
&\lim_{h\to\infty}\lim_{n\to\infty}MSE_h\left(n^{2q/(2q+\eta)},\hat{J}_n(d_n),S_n\right) \\
&=\alpha^{\frac{2q}{2q+\eta}}\tau^{\frac{\eta}{2q+\eta}}\left(\frac{1}{\tau}K_q^2\left(\text{vec }g^{(q)}\right)'S\left(\text{vec }g^{(q)}\right)+\bar{K}\text{tr}\left(S(I+K_{pp})(g\otimes g)\right)\right),
\end{aligned}
$$

It is straightforward to show that this is uniquely minimized over $\tau\in(0,\infty)$ by $\tau^\star=qK_q^2\kappa(q)/\eta$ (provided $0<\kappa<\infty$ and S is psd) and that a sequence $\{d_n\}$ satisfies $\frac{d_n^{2q}E\ell_n}{n}\to\tau^\star$ if and only if $d_n=d_n^\star+o(n^{1/(2q+\eta)})$.

Proof of Theorem 2

(a) $\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\hat{J}_n(\hat{d}_n)-J_n\right)=O_p(1)$ **and** $\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\hat{J}_n(\hat{d}_n)-\hat{J}_n(\ddot{d}_n)\right)=o_p(1)$

By Theorem 1(c), $\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\hat{J}_n(\ddot{d}_n)-J_n\right)=O_p(1)$. Therefore, it suffices to show the second part of Theorem 2(a). Without loss of generality, we assume J_n is a scalar random variable. Note that

$$
\begin{aligned}
&\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\hat{J}_{rs,n}(\hat{d}_n)-\hat{J}_{rs,n}(\ddot{d}_n)\right)\\
&=\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(K\left(\frac{d^*_{ij,n}}{\hat{d}_n}\right)-K\left(\frac{d^*_{ij,n}}{\ddot{d}_n}\right)\right)\hat{V}_{i,n}\hat{V}_{j,n}\right)\\
&:=M_{1n}+M_{2n}+M_{3n}
\end{aligned}
$$

where

$$
M_{1n}=\sqrt{\frac{n}{E\ddot{\ell}_n}}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left(K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)-K\left(\frac{d^*_{ij_{(i)},n}}{\ddot{d}_n}\right)\right)\left(\hat{V}_{i,n}\hat{V}'_{j_{(i)},n}-V_{i,n}V'_{j_{(i)},n}\right)\right],
$$

$$
M_{2n}=\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left(K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)-K\left(\frac{d^*_{ij_{(i)},n}}{\ddot{d}_n}\right)\right)V_{i,n}V'_{j_{(i)},n}\right),
$$

$$
M_{3n}=\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j_{(i)}=\ddot{\ell}_{i,n}+1}^{n}K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)\hat{V}_{i,n}\hat{V}'_{j_{(i)}}\right).
$$

The third term M_{3n} is zero when $\hat{d}_n\le\ddot{d}_n$. We assume that $\hat{d}_n>\ddot{d}_n$ below. Therefore, it suffices to show $M_{1n}=o_p(1)$, $M_{2n}=o_p(1)$ and $M_{3n}=o_p(1)$. We consider the case that $V_{i,n}$ is a scalar here as the proof for the vector case is similar.

Note that

$$
\begin{aligned}
\|M_{1n}\|&=\sqrt{\frac{n}{E\ddot{\ell}_n}}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left(K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)-K\left(\frac{d^*_{ij_{(i)},n}}{\ddot{d}_n}\right)\right)\left(\hat{V}_{i,n}\hat{V}'_{j_{(i)},n}-V_{i,n}V'_{j_{(i)},n}\right)\right]\\
&=\sqrt{\frac{1}{nE\ddot{\ell}_n}}\left\|\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left(K\left(\frac{d^*_{ij,n}}{\hat{d}_n}\right)-K\left(\frac{d^*_{ij,n}}{\ddot{d}_n}\right)\right)\left(\frac{\partial V_{i,n}(\bar{\theta})}{\partial\theta'}V_{j_{(i)},n}(\bar{\theta})+V_{i,n}(\bar{\theta})\frac{\partial V_{j_{(i)},n}(\bar{\theta})}{\partial\theta'}\right)\left(\hat{\theta}-\theta\right)\right\|\\
&\le\sqrt{\frac{E\ddot{\ell}_n}{n}}O_p(1)\frac{1}{E\ddot{\ell}_n\sqrt{n}}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left|\frac{d^*_{ij,n}}{\hat{d}_n}-\frac{d^*_{ij,n}}{\ddot{d}_n}\right|\left(\left\|\frac{\partial}{\partial\theta}V_{i,n}(\bar{\theta})V_{j_{(i)},n}(\bar{\theta})\right\|+\left\|V_{i,n}(\bar{\theta})\frac{\partial}{\partial\theta}V_{j_{(i)},n}(\bar{\theta})\right\|\right)\\
&\le o_p(1)\sqrt{n}\left|\frac{\ddot{d}_n}{\hat{d}_n}-1\right|\frac{1}{E\ddot{\ell}_n n}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left(\left\|\frac{\partial}{\partial\theta}V_{i,n}(\bar{\theta})V_{j_{(i)},n}(\bar{\theta})\right\|+\left\|V_{i,n}(\bar{\theta})\frac{\partial}{\partial\theta}V_{j_{(i)},n}(\bar{\theta})\right\|\right)\\
&=o_p(1)\frac{\ddot{\ell}_{i,n}}{E\ddot{\ell}_n}\frac{1}{\ddot{\ell}_{i,n}n}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left(\left\|\frac{\partial}{\partial\theta}V_{i,n}(\bar{\theta})V_{j_{(i)},n}(\bar{\theta})\right\|+\left\|V_{i,n}(\bar{\theta})\frac{\partial}{\partial\theta}V_{j_{(i)},n}(\bar{\theta})\right\|\right),
\end{aligned}
\tag{A.18}
$$

where the first inequality uses Assumption 7(ii) and the $O_p(1)$ and $o_p(1)$ terms hold as $\sqrt{n}\left(\hat{\theta}-\theta\right)=O_p(1)$, $\sqrt{n}\left(\ddot{d}_n/\hat{d}_n-1\right)=O_p(1)$. Since $\ddot{\ell}_{i,n}/E\ddot{\ell}_n=O_p(1)$, it now suffices to show that

$$\frac{1}{\ddot{\ell}_{i,n}n}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left(\left\|\frac{\partial}{\partial\theta}V_{i,n}(\bar{\theta})V_{j_{(i)},n}(\bar{\theta})\right\|+\left\|V_{i,n}(\bar{\theta})\frac{\partial}{\partial\theta}V_{j_{(i)},n}(\bar{\theta})\right\|\right)=O_p(1).$$

Using Assumption 8(ii) and (iii), we have

$$\begin{aligned}
&P\left(\frac{1}{\ddot{\ell}_{i,n}n}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left(\left\|\frac{\partial}{\partial\theta}V_{i,n}(\bar{\theta})V_{j_{(i)},n}(\bar{\theta})\right\|+\left\|V_{i,n}(\bar{\theta})\frac{\partial}{\partial\theta}V_{j_{(i)},n}(\bar{\theta})\right\|\right)>\Delta\right)\\
&\leq\frac{1}{\Delta}\frac{1}{\ddot{\ell}_{i,n}n}E\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left(E\left\|\frac{\partial}{\partial\theta}V_{i,n}(\bar{\theta})V_{j_{(i)},n}(\bar{\theta})\right\|+E\left\|V_{i,n}(\bar{\theta})\frac{\partial}{\partial\theta}V_{j_{(i)},n}(\bar{\theta})\right\|\right)\\
&\leq\frac{1}{\Delta}E\frac{1}{\ddot{\ell}_{i,n}n}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left(\left[E\left(\frac{\partial}{\partial\theta}V_{i,n}(\bar{\theta})\right)^2\right]^{\frac{1}{2}}\left[E\left(V_{j_{(i)},n}(\bar{\theta})\right)^2\right]^{\frac{1}{2}}\right)\\
&+\frac{1}{\Delta}E\frac{1}{\ddot{\ell}_{i,n}n}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left(\left[E\left(V_{i,n}(\bar{\theta})\right)^2\right]^{\frac{1}{2}}\left[E\left(\frac{\partial}{\partial\theta}V_{j_{(i)},n}(\bar{\theta})\right)^2\right]^{\frac{1}{2}}\right)\\
&\leq\frac{2}{\Delta}\sup_i\left(E\left[\sup_\theta\left(\frac{\partial}{\partial\theta}V_{i,n}(\theta)\right)^2\right]\right)^{\frac{1}{2}}\sup_j\left(E\sup_\theta\left(V_{j,n}(\theta)\right)^2\right)^{\frac{1}{2}}\to 0
\end{aligned}$$

as n and Δ grows. Thus, $M_{1n}=o_p(1)$.

We now consider M_{2n}. Since $\sqrt{n}\left(\ddot{d}_n/\hat{d}_n-1\right)=O_p(1)$, we have $P\left(\sqrt{n}\left|\ddot{d}_n/\hat{d}_n-1\right|>C\right)\to 0$ as $C\to\infty$. That is, for any $\varepsilon>0$, there exists a constant $C>0$ such that $P\left(\sqrt{n}\left|\ddot{d}_n/\hat{d}_n-1\right|>C\right)<\varepsilon$ for sufficiently large n. Hence we can focus on the event that $\mathcal{E}=\left\{\sqrt{n}\left|\ddot{d}_n/\hat{d}_n-1\right|<C\right\}$ and we do so in the following derivation:

$$\begin{aligned}
&P\left(\left\|\frac{1}{\sqrt{nE\ddot{\ell}_n}}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{in}}\left(K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)-K\left(\frac{d^*_{ij_{(i)},n}}{\ddot{d}_n}\right)\right)V_{i,n}V_{j_{(i)},n}\right\|>\delta\right)\\
&\leq\frac{1}{\delta^2}\frac{1}{nE\ddot{\ell}_n}E\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ddot{\ell}_{a,n}}\left\|E\left(V_{i,n}V_{j_{(i)},n}V_{a,n}V_{b_{(a)},n}\right)\right\|\\
&\times\left|\left(K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)-K\left(\frac{d^*_{ij_{(i)},n}}{\ddot{d}_n}\right)\right)\left(K\left(\frac{d^*_{ab_{(a)},n}}{\hat{d}_n}\right)-K\left(\frac{d^*_{ab_{(a)},n}}{\ddot{d}_n}\right)\right)\right|\\
&\leq\frac{C}{\delta^2}\frac{1}{nE\ddot{\ell}_n}E\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ddot{\ell}_{a,n}}\left\|E\left(V_{i,n}V_{j_{(i)},n}V_{a,n}V_{b_{(a)},n}\right)\right\|\frac{d^*_{ij_{(i)},n}}{\ddot{d}_n}\frac{d^*_{ab_{(a)},n}}{\ddot{d}_n}\left(\frac{\ddot{d}_n}{\hat{d}_n}-1\right)^2\\
&\leq\frac{C}{\delta^2}\frac{1}{E\ddot{\ell}_n}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}\left\|E\left(V_{i,n}V_{j,n}V_{a,n}V_{b,n}\right)\right\|. \qquad (A.19)
\end{aligned}$$

To compute the order of the above upper bound, we note that

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}\left\|E\left(V_{i,n}V_{j,n}V_{a,n}V_{b,n}\right)\right\| \\
&=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}\left|E\left(\sum_{\ell_1=1}^{np} r_{i\ell_1,n}\varepsilon_{\ell_1}\right)\left(\sum_{\ell_2=1}^{np} r_{j\ell_2,n}\varepsilon_{\ell 2}\right)\left(\sum_{\ell_3=1}^{np} r_{a\ell_3,n}\varepsilon_{\ell 3}\right)\left(\sum_{\ell_4=1}^{np} r_{b\ell_2,n}\varepsilon_{\ell 4}\right)\right| \\
&=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}\sum_{\ell=1}^{np}\left|r_{i\ell,n}r_{j\ell,n}r_{a\ell,n}r_{b\ell,n}\right|E\varepsilon_\ell^4+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}\left|\gamma_{i,a}\right|\left|\gamma_{j,b}\right| \\
&+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}\left|\gamma_{i,j}\right|\left|\gamma_{a,b}\right|+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a=1}^{n}\sum_{b=1}^{n}\left|\gamma_{i,b}\right|\left|\gamma_{j,a}\right| \\
&\leq C\left(\sum_{i=1}^{n}\left|r_{i\ell,n}\right|\right)^4\frac{p}{n}+3\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\gamma_{i,j}\right|\right)^2=O\left(1\right), \qquad \text{(A.20)}
\end{aligned}
$$

so the upper bound in (A.19) is $o\left(1\right)$, which implies that $M_{2n}=o_p\left(1\right).$

The next step is to show $M_{3n}=o_p(1)$. As before, we can focus on the event $\mathcal{E}=\left\{\sqrt{n}\left|\ddot{d}_n/\hat{d}_n-1\right|<C\right\}$. For any given $\delta>0$,

$$
\begin{aligned}
P\left(\left\|M_{3n}\right\|\geq\delta\right)&=P\left(\left\|\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j_{(i)}=\ddot{\ell}_{i,n}+1}^{n}K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)\hat{V}_{i,n}\hat{V}_{j_{(i)},n}\right)\right\|\geq\delta\right) \\
&=P\left(\left\|\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j_{(i)}=\ddot{\ell}_{i,n}+1}^{n}K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)V_{i,n}V_{j_{(i)},n}\left(1+o_p\left(1\right)\right)\right)\right\|\geq\delta\right) \\
&\leq P\left(\left\|\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j_{(i)}=\ddot{\ell}_{i,n}+1}^{n}K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)V_{i,n}V_{j_{(i)},n}\right)\right\|\geq\delta,\mathcal{E}\right)+P\left(\mathcal{E}^c\right) \\
&\leq\delta^{-2}E\left[\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j_{(i)}=\ddot{\ell}_{i,n}+1}^{\hat{\ell}_{i,n}}K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)V_{i,n}V_{j_{(i)},n}\right)\middle|\mathcal{E}\right]^2+P\left(\mathcal{E}^c\right)
\end{aligned}
$$

But

$$E\left[\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j_{(i)}=\ddot{\ell}_{i,n}+1}^{\hat{\ell}_{i,n}} K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right) V_{i,n}V_{j_{(i)},n}\right)\mathcal{E}\right]^2$$

$$\leq \frac{1}{nE\ddot{\ell}_n}E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=\ddot{\ell}_{i,n}+1}^{\hat{\ell}_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=\ddot{\ell}_{a,n}+1}^{\hat{\ell}_{a,n}}\left|K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)-K(1)\right|\left|K\left(\frac{d^*_{ab_{(a)},n}}{\hat{d}_n}\right)-K(1)\right|\right.$$

$$\left.\left|E\left(V_{i,n}V_{j_{(i)},n}V_{a,n}V_{b_{(a)},n}\right)\right|\mathcal{E}\right]$$

$$\leq \frac{c_L^2}{nE\ddot{\ell}_n}E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=\ddot{\ell}_{i,n}+1}^{\hat{\ell}_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=\ddot{\ell}_{a,n}+1}^{\hat{\ell}_{a,n}}\left|\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}-1\right|\times\left|\frac{d^*_{ab_{(a)},n}}{\hat{d}_n}-1\right|\left|E\left(V_{i,n}V_{j_{(i)},n}V_{a,n}V_{b_{(a)},n}\right)\right|\mathcal{E}\right]$$

$$\leq \frac{c_L^2}{nE\ddot{\ell}_n}E\left[\sum_{i=1}^{n}\sum_{j_{(i)}=\ddot{\ell}_{i,n}+1}^{\hat{\ell}_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=\ddot{\ell}_{a,n}+1}^{\hat{\ell}_{a,n}}\left|\frac{\ddot{d}_n}{\hat{d}_n}-1\right|\times\left|\frac{\ddot{d}_n}{\hat{d}_n}-1\right|\left|E\left(V_{i,n}V_{j_{(i)},n}V_{a,n}V_{b_{(a)},n}\right)\right|\mathcal{E}\right]$$

$$\leq \frac{c_L^2}{E\ddot{\ell}_n}E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{a=1}^{n}\sum_{j_{(i)}=\ddot{\ell}_{i,n}+1}^{\hat{\ell}_{i,n}}\sum_{b_{(a)}=\ddot{\ell}_{a,n}+1}^{\hat{\ell}_{a,n}}\left|E\left(V_{i,n}V_{j_{(i)},n}V_{a,n}V_{b_{(a)},n}\right)\right|\right]$$

$$\leq \frac{c_L^2}{E\ddot{\ell}_n}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{a=1}^{n}\sum_{j=1}^{n}\sum_{b=1}^{n}\left|E\left(V_{i,n}V_{j,n}V_{a,n}V_{b,n}\right)\right| = o(1),$$

using equation (A.20). Hence $M_{3n} = o_p(1)$. Consequently, $\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\hat{J}_n(\hat{d}_n)-\hat{J}_n(\ddot{d}_n)\right) = o_p(1)$.

The first equality of Theorem 2(b) holds by applying Lemma 5 in the same way as in proof of the first equality of Theorem 1(d). Then the second equality of Theorem 2(b) holds by Theorem 1(d).

Proof of Corollary 2

By Corollary 4 and Theorem 2(b),

$$\begin{aligned}&\lim_{h\to\infty}\lim_{n\to\infty}\left(MSE_h\left(n^{2q/(2q+\eta)},\hat{J}_n(\dot{d}_n),S_n\right)-MSE_h\left(n^{2q/(2q+\eta)},\hat{J}_n(\hat{d}_n),S_n\right)\right)\\&=\lim_{h\to\infty}\lim_{n\to\infty}\left(MSE_h\left(n^{2q/(2q+\eta)},\hat{J}_n(d_n),S_n\right)-MSE_h\left(n^{2q/(2q+\eta)},\hat{J}_n(\ddot{d}_n),S_n\right)\right)\end{aligned} \tag{A.21}$$

Since $\ddot{g}=g$ and $\ddot{g}^{(q)}=g^{(q)}$, $\ddot{d}_n=d^\star_n$. Corollary 1 implies that the expression in (A.21) is ≥ 0 with the inequality being strict unless $d_n=d^\star_n+o(n^{1/(2q+\eta)})$.