Heteroskedasticity and Spatiotemporal Dependence Robust Inference for Linear Panel Models with Fixed Effects (Job Market Paper)

Min Seong Kim*
Department of Economics, UC San Diego

November 14, 2010

Abstract

This paper studies robust inference for linear panel models with fixed effects in the presence of heteroskedasticity and spatiotemporal dependence of unknown forms. We propose a bivariate kernel covariance estimator, which is flexible to nest existing estimators as special cases with certain choices of bandwidths. For distributional approximations, we consider two different types of asymptotics. When the level of smoothing is assumed to increase with the sample size, the proposed estimator is consistent and the associated Wald statistic converges to a χ^2 distribution. We show that our covariance estimator improves upon existing estimators in terms of robustness and efficiency. When we assume the level of smoothing to be held fixed, the covariance estimator has a random limit and we show by asymptotic expansion that the limiting distribution of the test statistic depends on the bandwidth parameters, the kernel function, and the number of restrictions being tested. As this distribution is nonstandard, we establish the validity of an F-approximation to this distribution, which greatly facilitates the test. For optimal bandwidth selection, we propose a procedure based on the upper bound of asymptotic mean square error criterion. The flexibility of our estimator and proposed bandwidth selection procedure make our estimator adaptive to the dependence structure in data. This *adaptiveness* automates the selection of covariance estimator. That is, our estimator reduces to the existing estimators which are designed to cope with the particular dependence structures. Simulation results show that the F-approximation and the adaptiveness work reasonably well.

*Email: msk003@ucsd.edu. I am sincerely grateful to my advisor Yixiao Sun for his invaluable guidance and support. I thank Brendan Beare, Otilia Boldea, Pierre-Andre Chiappori, Gordon Dahl, Feico Drost, Graham Elliott, Patrik Guggenberger, Jinyong Hahn, James Hamilton, Hiroaki Kaido, Ivana Komunjer, Jan Magnus, Bertrand Melenberg, Philip Neary, Andres Santos, Halbert White and the participants of Panel Data Conference at Amsterdam, Econometric Society World Congress at Shanghai and seminars at Tilburg university and UCSD. Any comment or suggestion is appreciated.

1 Introduction

This paper studies robust inference for linear panel models with fixed effects in the presence of heteroskedasticity and spatiotemporal dependence of unknown forms. As economic data is potentially heterogeneous and correlated in unknown ways across individuals and time, the robust inference in the panel setting is a very important issue. See, for example, Betrand, Duflo and Mullainathan (2004) and Petersen (2009). The main interest in this problem lies in (i) how to construct covariance estimators that take the correlation structure into account; (ii) how to approximate the sampling distribution of the associated test statistic; and (iii) how to select smoothing parameters in finite samples.

Regarding covariance estimation, we propose a bivariate kernel estimator. In order to utilize the kernel in the spatial dimension, we need some *a priori* knowledge about the dependence structure. It is often assumed that the covariance of two random variables at locations i and j is a decreasing function of some observable distance measure d_{ij} between them. An example of d_{ij} is the economic distance. The idea of using a distance measure to characterize spatial dependence is common in the spatial econometrics literature. See, for example, Conley (1999), Kelejian and Prucha (2007, KP hereafter), Bester, Conley, Hansen and Vogelsang (2009, BCHV hereafter) and Kim and Sun (2010).

There are several robust covariance estimators with correlated panel data. Liang and Zeger (1986) and Arellano (1987) propose the clustered covariance estimator (CCE) by extending White standard error (White, 1980) to account for serial correlation. Wooldridge (2003) provides a concise review on this estimator and Kèzdi (2003) explores its properties in panel models with fixed effects. Driscoll and Kraay (1998, DK hereafter) suggest a different approach that uses a time series HAC estimator (Domowitz and White, 1982; and Newey and West, 1987) with the cross-sectional averages of moment conditions. Gonçlaves (2008) examines the properties of this estimator in linear panel models with fixed effects. Another approach considered in this paper is the extension of the spatial HAC estimator with serial averages of moment conditions. This is symmetric to the DK estimator. The spatial HAC estimator is firstly proposed by Conley (1999), and KP argue that it can be extended to the panel setting with fixed T.

As our estimator is based on the bivariate kernel, it nests these existing estimators as special cases, reducing to each of them with certain bandwidth selection. We refer to this as *flexibility.* If the sequence of the bandwidth in the spatial dimension, d_n, is assumed to increase fast enough, then our estimator with the rectangular kernel converges to the DK estimator. Similarly, if we assume the sequence of the bandwidth in the time dimension, d_T, to increase fast enough, then our estimator with the rectangular kernel converges to the KP estimator. On the other hand, if d_n is assumed to approach to zero, our estimator reduces to the generalized CCE.

For distributional approximations, we consider both the increasing smoothing asymptotics and fixed smoothing asymptotics. Let $\ell_{i,n}$ denote the number of individuals whose distance from individual i is less than or equal to d_n and ℓ_n be the average of $\ell_{i,n}$ across i. If $d_n, d_T \to \infty$ as $n, T \to \infty$ but slowly so that $\ell_n \ell_T / nT \to 0$, then the level of smoothing increases with the sample size. As a result, our covariance estimator is consistent and the limiting distribution of the associated Wald statistic is the χ^2 distribution.

The alternative estimators are also consistent under some regularity conditions, but each approach has an important limitation in practice. The properties of the CCE heavily depend on spatial correlation. While this estimator is quite efficient with zero correlation in the spatial dimension, even moderate spatial correlation may lead to the substantial bias of the estimator and hence size distortion in statistical inference. Though zero correlation between individuals is sometimes assumed for convenience, they are generally not independent due to, for example, spill-over effects, competition and so on.[1] For the DK estimator, collapsing spatial dependence by the cross-sectional averaging, it is robust to arbitrary form of spatial dependence. However, when spatial dependence decreases with some distance measure, this estimator is not efficient because it does not downweigh or truncate the covariance between spatially remote units. The variance of the DK estimator does not vanish as n increases. For the KP estimator, in contrast, it does not employ downweighing or truncation in the time domain.

The proposed estimator improves upon the above estimators by employing a bivariate kernel. It does not require zero spatial correlation for consistency in contrast to the CCE and more efficient than the DK and KP estimators in general. More specifically, if individuals are located on a 2-dimensional lattice and the Bartlett kernel is used, our estimator is more efficient than the DK estimator if $T = o(n^{3/2})$ and than the KP estimator if $n = o(T^4)$. The conditions are more generous with the second order kernels, such as the Parzen kernel, i.e., $T = o(n^{5/2})$ and $n = o(T^6)$.

If $\ell_n \ell_T / nT$ is assumed to be held fixed, then the level of smoothing is fixed with the sample size. Under fixed smoothing asymptotics, the covariance estimator converges in distribution to a random matrix and the limiting distribution of Wald statistic is nonstandard but pivotal. The fixed smoothing asymptotic approximation is firstly suggested by Kiefer, Vogelsang and Bunzel (2000) and Kiefer and Vogelsang (2002a, 2002b, 2005) in the time series context. This is usually referred to as 'fixed-b' asymptotics where b denotes the ratio of the bandwidth parameter d_T to the sample size T. They show by simulation that the fixed-b asymptotic approximation is more accurate in size than the conventional asymptotic χ^2 approximation. Jansson (2004), Sun, Phillips and Jin (2008), and Sun and Phillips (2009) provide its theoretical analyses.

We adapt the fixed smoothing asymptotics in the panel setting with our covariance estimator. Using asymptotic expansion we show that the deviation of this limiting distribution from the χ^2 distribution depends on the smoothing parameters, kernel function and the number of restrictions being tested. We can also accomodate the estimation uncertainty of the parameter estimator under fixed smoothing asymptotics. As the limiting distribution is nonstandard, we extend Sun (2010) to establish the validity of an F-approximation to this distribution. Under fixed smoothing asymptotics, the covariance estimator converges in distribution to an infinite weighted sum of independent Wishart distributions. We approximate the infinite weighted sum of independent Wishart distributions with a single Wishart distribution with an 'equivalent degree of freedom.' With this result, the fixed smoothing limiting distribution of the scaled Wald statistic with some correction factor becomes approximately F distributed. This F-approximation greatly facilitates the testing procedure because we can obtain the critical values without simulation.

Several testing methods using the fixed smoothing asymptotics are recently proposed in the spatial or panel setting. BCHV extend the fixed-b asymptotics to the spatial context where dependence is indexed in more than one dimension, and propose an *i.i.d.* bootstrap method to obtain the critical values. Vogelsang (2008) develops a fixed-b asymptotic theory for statistics based on the generalized CCE and the DK estimator. Besides the kernel methods, Hansen (2007) and Bester, Conley and

[1]Cameron, Gelbach, and Miller (2006) and Thompson (2009) address this problem by clustering on the time and spatial dimensions simultaneously. While this allows for both the serial and spatial correlations, observations on different individuals in different time are assumed to be uncorrelated (Peterson, 2009).

Hansen (2009) apply the fixed smoothing asymptotics to the testing procedure with the CCE. They assume the number of clusters to be fixed and the number of observations per cluster to increase with the sample size. Ibragimov and Müller (2010) consider fixed smoothing asymptotics for the Fama and MacBeth (1973) type procedure by fixing the number of groups. Sun and Kim (2010) considers a testing procedure using the series covariance estimator in the spatial setting. They show that, when the number of basis functions is assumed to be held fixed, the series estimator converges in distribution to a Wishart distribution, and that the scaled Wald statistic converges to F distribution. The motivation of our F-approximation arises from this series method. All the kernel covariance estimators with positive semi-definite kernels can be written as series covariance estimators (Percival and Walden, 1993, p. 353). Actually, the other two 'non-kernel' methods also use the t or F distribution as the reference distribution while the kernel methods by BCHV and Vogelsang (2008) have to simulate the critical values. From this point of view, our paper fills the gap in the literature, providing F-approximation for the kernel method in the panel setting.

As the finite sample performance of our estimator may heavily depend on the choice of bandwidth parameters, we propose an optimal bandwidth selection procedure based on the upper bound of the asymptotic mean square error (AMSE*) criterion. Though it is standard practice to use the asymptotic mean square error (AMSE) criterion for bandwidth selection (e.g. Andrews, 1991 and Newey and West, 1994), this criterion is not tractable for the proposed estimator. Our minimax criterion is simple to implement and makes the bias and variance tradeoff transparent. It is interesting to note that the level of persistence in each dimension affects both $d_T^\star$ and $d_n^\star$, the optimal bandwidth parameters in the time and spatial dimensions respectively, but in the opposite direction. That is, for example, if a process becomes persistent in the spatial dimension, the criterion does not only increase $d_n^\star$ to capture the bias but also decreases $d_T^\star$ to reduce the variance. We suggest a parametric plug-in procedure for practical implementation. Four different types of spatiotemporal parametric models in Anselin (2001) are considered.

Our bandwidth selection procedure does not apply directly to the rectangular and, more broadly, flat-top kernel estimators. However, it is interesting to consider flat-top kernel estimators because they are higher order accurate (Politis, 2010). This is particularly more important in our setting because the flexibility is the main ingredient of the adaptiveness of our estimator which is explained below. In the literature, Andrews (1991, footnote on p. 834) and Lin and Sakata (2009) suggest a practical rule for the rectangular kernel estimator based on the AMSE criterion. Sun and Kaplan (2010) explore this problem rigorously and provide a bandwidth selection procedure which is testing optimal. Both the methods lead the rectangular kernel estimator to better asymptotic properties than any finite order kernel estimator we target. We modify our procedure to be applicable to the rectangular kernel based on these preceding procedures.

The flexibility of our estimator and proposed optimal bandwidth selection procedure make our estimator adaptive to the dependence structure in data. That is, our estimator reduces to the existing estimators that are designed to cope with the particular dependence structures. This *adaptiveness* is the salient feature of our method. As it practically automates the selection of covariance estimator, our estimation procedure can be safely used in the presence of very general forms of spatiotemporal dependences. This is confirmed by our Monte Carlo study.

The papers that are closely related to ours are Andrews (1991) and Kim and Sun (2010). We present a rigorous analysis of the properties of the panel HAC estimator as they do in the time series and spatial settings. Another paper which is closely related is Sun (2010) who develops the

F-approximation to the fixed-b asymptotic distribution in the time series context. Extension of his results to the panel setting is important. As the critical values from the nonstandard distributions depend not only on the kernels and chosen values of smoothing parameters but also on the distribution of observed locations in the panel setting, it is much more complicated to simulate the critical values than in the times series case.

The remainder of the paper is as follows. Section 2 introduces the panel model, covariance estimator and hypothesis testing we consider. In section 3, we examine the properties of our estimator and the associated test statistic under increasing smoothing asymptotics. Section 4 develops an optimal bandwidth selection procedure. Section 5 examines the properties of the existing estimators. The flexibility and adaptiveness of our estimator are illustrated in section 6. In section 7, we study the limit theory for our covariance estimator and the associated test statistic under fixed smoothing asymptotics. We also establish its F-approximation. Section 8 reports simulation evidence. The last section concludes.

2 Panel model, covariance estimator and hypothesis testing

In this paper, we consider a linear panel regression model with fixed effects[2]:

$$Y_{it} = X_{it}'\beta_0 + \alpha_i + f_t + u_{it}, \tag{1}$$

where X_{it} and β are p-vectors and α_i and f_t denote scalar individual and time effects respectively. When X_{it} is correlated with α_i and f_t, we may use a fixed effects estimation approach. Let $\bar{Z}_i = T^{-1}\sum_{t=1}^{T} Z_{it}$, $\bar{Z}_t = n^{-1}\sum_{i=1}^{n} Z_{it}$ and $\bar{Z} = (nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} Z_{it}$. We also define $\tilde{Z}_{it} = Z_{it} - \bar{Z}_i - \bar{Z}_t + \bar{Z}$. Then, the fixed effects estimator, $\hat{\beta}$, is defined as

$$\hat{\beta} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}'\right)^{-1} \sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{Y}_{it}. \tag{2}$$

Under some regularity conditions, the asymptotic distribution of $\hat{\beta}$ is

$$\left(Q_{nT} J_{nT} Q_{nT}'\right)^{-\frac{1}{2}} \sqrt{nT}\left(\hat{\beta} - \beta_0\right) \xrightarrow{d} N\left(0, I_p\right) \text{ as } n, T \to \infty,$$

where

$$Q_{nT} = \left((nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} E\left[\tilde{X}_{it}\tilde{X}_{it}'\right]\right)^{-1} \text{ and } J_{nT} = var\left((nT)^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}u_{it}\right).$$

To make inference on β_0, we have to estimate unknown quantities in the asymptotic variance of $\hat{\beta}$. Since Q_{nT} is consistently estimated with its sample analog, our central interest is on J_{nT}. Letting

[2] Our analysis can potentially be generalized to the GMM setting. We focus on a static linear panel model to be free from the incidental parameters problem (Neyman and Scott, 1948) which the fixed effects estimators of nonlinear and dynamic panel models usually suffer from. See Arellano and Hahn (2006) for detail.

$V_{(i,t)} = \tilde{X}_{it} u_{it}$, J_{nT} can be rewritten as

$$J_{nT} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[V_{(i,t)} V'_{(j,s)}\right]$$
$$:= \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \Gamma_{(it,js)}. \tag{3}$$

We propose a bivariate kernel covariance estimator which is given as

$$\hat{J}_{nT} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right) K\left(\frac{d_{ts}}{d_T}\right) \hat{V}_{(i,t)} \hat{V}'_{(j,s)}, \tag{4}$$

where $\hat{V}_{(i,t)} = \tilde{X}_{it}(\tilde{Y}_{it} - \tilde{X}'_{it}\hat{\beta})$ and $K(\cdot)$ is a real-valued kernel function.[3] d_{ij} and d_{ts} denote the distance measures in the spatial and time dimensions and d_n and d_T are the corresponding bandwidth parameters. Whereas it is natural to define $d_{ts} = |t - s|$, what is used to measure d_{ij} differs with applications. Geographic distance is one of the most common measures (e.g. BCHV and Barrios, Diamond, Imbens and Kolesar, 2010), but other measures can also be considered, e.g. transportation cost (Conley and Ligon, 2000) and similarity of input and output structure (Chen and Conley, 2001 and Conley and Dupor, 2003).

Consider null hypothesis $H_0 : R\beta = r_0$ and the alternative hypothesis $H_1 : R\beta \neq r_0$ where R is a $g \times p$ matrix and r_0 is a g-vector. For hypothesis testing, we use the Wald statistic

$$W_{nT} = \sqrt{nT}\left(R\hat{\beta} - r_0\right)' \left(R\hat{Q}_{nT}\hat{J}_{nT}\hat{Q}'_{nT}R'\right)^{-1} \sqrt{nT}\left(R\hat{\beta} - r_0\right)$$

where $\hat{Q}_{nT} = \left((nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}'_{it}\right)^{-1}$, and its F-test version

$$F_{nT} = W_{nT}/g.$$

3 Increasing smoothing asymptotics

3.1 Basic setting

We employ the linear transformation of nTp common innovations to represent the process of $V_{(i,t)}$ as follows:

$$V_{(i,t)} = \tilde{R}_{(i,t)}\tilde{\varepsilon}, \tag{5}$$

where

$$\tilde{R}_{(i,t)} = \begin{bmatrix} \left(\tilde{r}^{(1)}_{(it,1,1)}, \tilde{r}^{(1)}_{(it,2,1)}, \ldots, \tilde{r}^{(1)}_{(it,n,T)}\right) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \left(\tilde{r}^{(p)}_{(it,1,1)}, \tilde{r}^{(p)}_{(it,2,1)}, \ldots, \tilde{r}^{(p)}_{(it,n,T)}\right) \end{bmatrix}$$

[3]For simplicity of our analysis, the same kernel is assumed to be used in both the spatial and time dimensions.

is a $p \times nTp$ block diagonal matrix with unknown elements and $\tilde{\varepsilon} = \left(\left(\tilde{\varepsilon}^{(1)} \right)', \ldots, \left(\tilde{\varepsilon}^{(p)} \right)' \right)'$ in which $\tilde{\varepsilon}^{(c)} = \left(\tilde{\varepsilon}^{(c)}_{(1,1)}, \ldots, \tilde{\varepsilon}^{(c)}_{(n,1)}, \tilde{\varepsilon}^{(c)}_{(1,2)}, \ldots, \tilde{\varepsilon}^{(c)}_{(n,T)} \right)'$. As in Kim and Sun (2009), we assume that

$$\operatorname{var}\left(\tilde{\varepsilon}^{(c)} \right) = \sigma_{cc} I_{nT}, \ \operatorname{cov}\left(\tilde{\varepsilon}^{(c)}, \tilde{\varepsilon}^{(d)} \right) = \sigma_{cd} I_{nT}$$

and

$$\operatorname{var}\left(\tilde{\varepsilon} \right) = \Sigma \otimes I_{nT} \text{ with } \Sigma = \left(\sigma_{cd} \right),$$

where $c, d = 1, \ldots, p$ and $\otimes$ denotes the Kronecker product. This type of linear array processes allows for nonstationarity and unconditional heteroskedasticity of $V_{(i,t)}$ and includes many spatiotemporal parametric models such as spatial dynamic models (Anselin, 2001) as special cases. It also treats the temporal and spatial dependence in a symmetric way.

Let $R_{(i,t)} := \tilde{R}_{(i,t)} \left(\Sigma^{1/2} \otimes I_{nT} \right)$ and $\varepsilon := (\varepsilon_1, \ldots, \varepsilon_l, \ldots, \varepsilon_{nTp})' = \left(\Sigma^{-1/2} \otimes I_{nT} \right) \tilde{\varepsilon}$. Then,

$$V_{(i,t)} = R_{(i,t)} \varepsilon \text{ and } \operatorname{var}\left(\varepsilon \right) = I_{nTp}. \tag{6}$$

The matrix $R_{(i,t)}$ can be written more explicitly as

$$\begin{aligned} R_{(i,t)} &:= \begin{bmatrix} \begin{pmatrix} r^{(1)}_{(i,t),1} & \cdots & r^{(1)}_{(i,t),nTp} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} r^{(p)}_{(i,t),1} & \cdots & r^{(p)}_{(i,t),nTp} \end{pmatrix} \end{bmatrix} \\ &= \begin{bmatrix} \sigma^{11} \begin{pmatrix} \tilde{r}^{(1)}_{(it,1,1)} & \cdots & \tilde{r}^{(1)}_{(it,n,T)} \end{pmatrix} & \cdots & \sigma^{1p} \begin{pmatrix} \tilde{r}^{(1)}_{(it,1,1)} & \cdots & \tilde{r}^{(1)}_{(it,n,T)} \end{pmatrix} \\ \vdots & \ddots & \vdots \\ \sigma^{p1} \begin{pmatrix} \tilde{r}^{(p)}_{(it,1,1)} & \cdots & \tilde{r}^{(p)}_{(it,n,T)} \end{pmatrix} & \cdots & \sigma^{pp} \begin{pmatrix} \tilde{r}^{(p)}_{(it,1,1)} & \cdots & \tilde{r}^{(p)}_{(it,n,T)} \end{pmatrix} \end{bmatrix} \end{aligned}$$

where σ^{cd} denotes the (c,d)-th element of $\Sigma^{1/2}$. We make the following assumption on ε_l.

Assumption I1 *For all $l = 1, \ldots, nTp$, $\varepsilon_l \overset{i.i.d.}{\sim} (0,1)$ with $E\left[\varepsilon_l^4 \right] \leq c_E$ for some constant $c_E < \infty$.*

For simplicity, we assume that ε_l is independent of ε_k for $l \neq k$. We can relax independence assumption to zero correlation but with more tedious calculations. Under Assumption I1, the covariance matrix of $V_{(i,t)}$ and $V_{(j,s)}$ is given by

$$\Gamma_{(it,js)} := \left(\gamma^{(cd)}_{(it,js)} \right) = E\left[V_{(i,t)} V'_{(j,s)} \right] = R_{(i,t)} R'_{(j,s)}, \tag{7}$$

where the (c,d)-th element of $\Gamma_{(it,js)}$ is denoted by $\gamma^{(cd)}_{(it,js)}$. Accordingly, the covariance matrix can be restated as

$$J_{nT} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} R_{(i,t)} R'_{(j,s)}, \tag{8}$$

and the (c,d)-th element of J_{nT} is

$$J_{nT}(c,d) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\sum_{l=1}^{nTp} r_{(i,t),l}^{(c)} r_{(j,s),l}^{(d)}\right). \tag{9}$$

Assumption I2 *For all $l = 1, ..., nTp$, $c = 1, ..., p$, n and T, $\sum_{i=1}^{n}\sum_{t=1}^{T}\left|r_{(i,t),l}^{(c)}\right| < c_R$ for some constant c_R, $0 < c_R < \infty$.*

Assumption I3 *There exist $q_1, q_2 > 0$ such that*

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\|\Gamma_{(it,js)}\right\| d_{ij}^{q_1} < \infty \text{ and } \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\|\Gamma_{(it,js)}\right\| d_{ts}^{q_2} < \infty$$

for all n and T, where $\|A\|$ denotes the Euclidean norm of matrix A.

Assumptions I2 and I3 impose the conditions on the persistence of the process. If $|\sigma^{cd}| \leq C$ for a constant $C > 0$, then Assumption I2 holds if $\sum_{i=1}^{n}\sum_{t=1}^{T}|\tilde{r}_{(it,j,s)}^{(d)}| < c_R/C$. Since $|\tilde{r}_{(it,j,s)}^{(d)}|$ can be regarded as the (absolute) change of $V_{(i,t)}^{(d)}$ in response to one unit change in $\tilde{\varepsilon}_{(j,s)}^{(d)}$, the summability condition requires that the aggregate response to an innovation be finite. The condition holds trivially if the set $\{\tilde{r}_{(it,j,s)}^{(d)}, i = 1, ..., n \text{ and } t = 1, ..., T\}$ has only a finite number of nonzero elements. In this case, the dependence induced by the innovation $\tilde{\varepsilon}_{(j,s)}^{(d)}$ are limited to a finite number of units. Assumption I3 implies that $\Gamma_{(it,js)}$ decays to zero fast as d_{ij} and d_{ts} increase so that the two summability conditions holds. These conditions hold if

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\sum_{a=1}^{n}\sum_{b=1}^{T}\tilde{r}_{(it,a,b)}^{(c)}\tilde{r}_{(js,a,b)}^{(d)}\right| d_{ij}^{q_1} < \infty, \tag{10}$$

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\sum_{a=1}^{n}\sum_{b=1}^{T}\tilde{r}_{(it,a,b)}^{(c)}\tilde{r}_{(js,a,b)}^{(d)}\right| d_{ts}^{q_2} < \infty \tag{11}$$

for all c and d. (10) and (11) imply that as d_{ij} or d_{ts} increases, the corresponding two row vectors in $\tilde{R}_{(i,t)}$ and $\tilde{R}_{(j,s)}$, $(\tilde{r}_{(it,1,1)}^{(c)}, \ldots, \tilde{r}_{(it,n,T)}^{(c)})$ and $(\tilde{r}_{(js,1,1)}^{(d)}, \ldots, \tilde{r}_{(js,n,T)}^{(d)})$ become nearly orthogonal. As the row vector represents the aggregate response of a unit to all the innovations, this assumption implies the responses of two units become independent as they become spatially or serially distant. Assumption I3 enables us to truncate the sum of $\Gamma_{(it,js)}$ and downweigh the summand without incurring much bias.

As Assumption I3 implies, the key property of d_{ij} is to characterize the decaying pattern of the spatial dependence. In addition, we assume that d_{ij} satisfies the properties of distance in a metric space: (i) $d_{ij} \geq 0$, (ii) $d_{ii} = 0$, (iii) $d_{ij} = d_{ji}$, and (iv) $d_{ij} \leq d_{ik} + d_{kj}$. In practice, nonetheless, the symmetry condition (iii) may not hold for some candidates of economic distance. Conley and Ligon (2000), for example, notice that transportation costs among countries violate this condition if tariff barriers are asymmetric. In such a case adjustment should be made.[4] This adjustment should not affect the asymptotic properties of our estimator from a perspective of the measurement error problem as explained below.

Observed distance data available to empirical researchers usually contain measurement errors, and the results in this paper can be generalized to the case when d_{ij} is error contaminated. Following Kim and Sun (2009), we can show that our asymptotic results are still valid under the following conditions: (i) the measurement error is independent of ε_l; (ii) it is of order $o(d_n)$ as d_n increases; and (iii) the first summability condition in Assumption I3 holds with an error contaminated distance measure. In this paper, however, we do not consider measurement errors for simplicity.

Let

$$\ell_{i,n} = \sum_{j=1}^{n} 1\{d_{ij} \leq d_n\} \text{ and } \ell_n = n^{-1} \sum_{i=1}^{n} \ell_{i,n}.$$

$\ell_{i,n}$ 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 I4 *For all $i = 1, \ldots, n$, $\ell_{i,n} \leq C\ell_n$ for some constant C.*

Assumption I4 allows the units to be irregularly located but rules out the case that they are concentrated only in some limited area while other area is scarce. To be symmetric, we also define

$$\ell_{t,T} = \sum_{s=1}^{T} 1\{d_{ts} \leq d_T\} \text{ and}$$
$$\ell_T = T^{-1} \sum_{t=1}^{T} \ell_{t,T} = 2d_T + 1 - \frac{d_T(d_T+1)}{T},$$

where $-d_T(d_T+1)/T$ is an adjustment coming from the points near the boundary.

In order to obtain the properties of the estimator in Theorem 1 below, it is important to control for the boundary effects. That is, the effects of the units near the boundary should become negligible as the sample size increases, so that the asymptotic properties depend only on the behavior of the units in the interior. We define

$$E_n := \{i : \ell_{i,n} = \ell_n + o(\ell_n)\},\ n_1 = \sum_{i=1}^{n} 1\{i \in E_n\},\ n_2 = n - n_1$$
$$E_T := \{t : \ell_{t,T} = \ell_T + o(\ell_T)\},\ T_1 = \sum_{t=1}^{n} 1\{t \in E_T\} \text{ and } T_2 = T - T_1.$$

E_n and E_T represent the nonboundary sets in the spatial and time dimensions. n_1 and T_1 denote the sizes of E_n and E_T and n_2 and T_2 denote the sizes of the boundary sets. These definitions imply

[4]In Conley and Ligon (2000), the asymmetry of transportation costs are adjusted by using the minimum cost between two countries.

that the size of a boundary set relies on choice of the corresponding bandwidth. We can mitigate the boundary effects by raising d_n and d_T slowly as n and T increase to make the interior large enough. Provided that n_2/n and T_2/T are $o(1)$, the boundary effects are asymptotically negligible. When units are regularly spaced on a lattice in $\mathbb{R}^2$, $n_2/n = o(1)$ if $\ell_n/n = o(1)$. $T_2/T = o(1)$ holds if $\ell_T/T = o(1)$ (or $d_T/T = o(1)$).

3.2 Properties of $\hat{J}_{nT}$ and limiting distribution of Wald statistic under increasing smoothing asymptotics

We present the consistency, the rate of convergence, and the AMSE of the estimator and the limiting distribution of the associated test statistic under increasing smoothing asymptotics. We begin by introducing the assumption on the kernel used in the estimator.

Assumption I5 *(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*

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

(iii) $\ell_n^{-1} \sum_{j=1}^n K^2 \left(\frac{d_{ij}}{d_n} \right) \to \bar{\mathcal{K}}_1$ *for all* $i \in E_n$.

Examples of kernels which satisfy Assumptions I5 (i) and (ii) are the Bartlett, Tukey-Hanning and Parzen kernels. The quadratic spectral (QS) kernel does not satisfy Assumption I5 (i) because it does not truncate. We may generalized our results to include the QS kernel but this requires considerable amount of work. Assumption I5 (iii) is more of an assumption on the distribution of the units. When the observations are located on a 2-dimensional integer lattice and d_{ij} is the Euclidian distance, we have

$$\bar{\mathcal{K}}_1 = \frac{1}{\pi} \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} K^2 \left(\sqrt{x^2 + y^2} \right) dy dx = 2 \int_0^1 r K^2(r) dr.$$

In finite samples, we may use

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

for $\bar{\mathcal{K}}_1$. For the kernel in time dimension, we define

$$\ell_T^{-1} \sum_{s=1}^{T} K^2 \left(\frac{d_{ts}}{d_T} \right) \to \int_0^1 K^2(r) dr := \bar{\mathcal{K}}_2.$$

The asymptotic variance of $\hat{J}_{nT}$ depends on J which is the limit value of J_{nT}.

$$J := \lim_{n,T \to \infty} J_{nT} = \lim_{n,T \to \infty} \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \Gamma_{(it,js)}.$$

Assumption I6 *For $i \in E_n$ and $t \in E_T$,*

$$\lim_{n,T\to\infty} var\left(\frac{1}{\sqrt{\ell_n \ell_T}} \sum_{j:d_{ij}\le d_n} \sum_{s:d_{ts}\le d_T} V_{(j,s)}\right) = J.$$

Assumption I6 states that the covariance matrix defined locally for each nonboundary unit converges to the same limiting value of J_{nT}. This assumption is related to covariance stationarity but weaker. It is implied by covariance stationarity but it can hold even though covariance stationarity is violated. Stationarity seems to be a very strong assumption especially in the spatial dimension because a spatial process is nonstationary simply if each unit has different numbers of neighbors. This assumption is similar to the homogeneity assumption in Bester, Hansen and Conley (2009). They assume that the covariance matrix in each group converges to the same limit.

The asymptotic bias of $\hat{J}_{nT}$ is determined by the smoothness of the kernel at zero and the decaying rates of the spatial and temporal dependence in terms of d_{ij} and d_{ts}. 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 $x = 0$. Under the assumption that $q \le q_i$ with $i = 1, 2$, we define

$$b_1^{(q)} = \lim_{n,T\to\infty} b_n^{(q)}, \text{ where } b_n^{(q)} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} \Gamma_{(it,js)} d_{ij}^{q},$$

$$b_2^{(q)} = \lim_{n,T\to\infty} b_T^{(q)}, \text{ where } b_T^{(q)} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} \Gamma_{(it,js)} d_{ts}^{q}.$$

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

Assumption I7 *(i)* $\sqrt{nT}\left(\hat{\beta}-\beta_0\right) = O_p(1)$. *(ii)* $(nT)^{-\frac{1}{2}} \sum_{i=1}^{n}\sum_{t=1}^{T} u_{it} = O_p(1)$. *(iii)* $(nT)^{-\frac{1}{2}} \sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{X}_{it} u_{it} = O_p(1)$. *(iv)* $\sup_{i,t} E\tilde{X}_{it}^2 < \infty$.

Assumption I7 is rather standard. It excludes the case of strong spatial dependence, which is considered in Gonçalves (2010).

We define the MSE as

$$MSE\left(\frac{nT}{\ell_n\ell_T}, \hat{J}_{nT}, S_{nT}\right) = \frac{nT}{\ell_n\ell_T} E\left[\text{vec}(\hat{J}_{nT}-J_{nT})' S_{nT} \text{vec}(\hat{J}_{nT}-J_{nT})\right],$$

where S_{nT} is some $p^2 \times p^2$ weighting matrix and vec$(\cdot)$ is the column by column vectorization function. We also define $\tilde{J}_{nT}$ as the pseudo-estimator that is identical to $\hat{J}_{nT}$ but is based on the true parameter, β_0, in place of $\hat{\beta}$. That is,

$$\tilde{J}_{nT} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right) K\left(\frac{d_{ts}}{d_T}\right) V_{(i,t)} V'_{(j,s)}.$$

Under the assumptions above, the effect of using $\hat{\beta}$ instead of β_0 on the asymptotic property is $o_p(1)$ as Theorem 1 (c) states below. Therefore, we use $\tilde{J}_{nT}$ to analyze the asymptotic properties of $\hat{J}_{nT}$.

Assumption I8 *For* $i = 1, \ldots, p$ $E|\hat{\beta}_i|^2 < \infty$, *where* $\hat{\beta}_i$ *is the* i^{th} *element of* $\hat{\beta}$.

Assumption I8 rules out the case when $\hat{\beta}$ has an infinite second moment (Mariano, 1972 and Kinal, 1980) which causes the underlying estimation error to dominate the MSE.[5]

Assumption I9 S_{nT} *is positive semidefinite and* $S_{nT} \xrightarrow{p} S$ *for a positive definite matrix* S.

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

Theorem 1 *Suppose that Assumptions I1-I6 hold,* $d_n, d_T \to \infty$, $n_2(d_n) = o(n)$, $d_T = o(T)$ *and* $\ell_n \ell_T = o(nT)$.

(a) $\lim_{n,T\to\infty} \frac{nT}{\ell_n \ell_T} var\left(vec\tilde{J}_{nT}\right) = \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 (I_{pp} + \mathbb{K}_{pp})(J \otimes J)$.

(b) *Let* $k_{nT} = d_T/d_n$ *and* $k_{nT} \to k > 0$ *as* $n, T \to \infty$. *Then,* $\lim_{n,T\to\infty} d_n^q (E\tilde{J}_{nT} - J_{nT}) = -K_q \left(b_1^{(q)} + \frac{1}{k^q} b_2^{(q)}\right)$

(c) *If Assumption I7 holds and* $d_n^{2q} \ell_n \ell_T / nT \to \tau \in (0, \infty)$, *then* $\sqrt{\frac{nT}{\ell_n \ell_T}}\left(\hat{J}_{nT} - J_{nT}\right) = O_p(1)$ *and* $\sqrt{\frac{nT}{\ell_n \ell_T}}\left(\hat{J}_{nT} - \tilde{J}_{nT}\right) = o_p(1)$.

(d) *Under the conditions of part (c), Assumptions I8 and I9,*

$$\begin{aligned}
&\lim_{n,T\to\infty} MSE\left(\frac{nT}{\ell_n \ell_T}, \hat{J}_{nT}, S_{nT}\right) \\
&= \lim_{n,T\to\infty} MSE\left(\frac{nT}{\ell_n \ell_T}, \tilde{J}_{nT}, S\right) \\
&= \frac{1}{\tau} K_q^2 vec\left(b_1^{(q)} + \frac{1}{k^q} b_2^{(q)}\right)' S vec\left(b_1^{(q)} + \frac{1}{k^q} b_2^{(q)}\right) + \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 tr\left(S(I + \mathbb{K}_{pp})(J \otimes J)\right).
\end{aligned}$$

Proofs are given in the appendix. For each element of $\tilde{J}_{nT}$, the asymptotic variance in Theorem 1 (a) is rewritten as

$$\begin{aligned}
&\lim_{n,T\to\infty} \frac{nT}{\ell_n \ell_T} \text{cov}\left(\tilde{J}_{nT}(c_1, d_1), \tilde{J}_{nT}(c_2, d_2)\right) \\
&= \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 \left[J(c_1, c_2) J(d_1, d_2) + J(c_1, d_2) J(d_1, c_2)\right].
\end{aligned}$$

Theorem 1 (a) and (b) show that the asymptotic variance and bias of $\tilde{J}_{nT}$ depend on the choice of d_n and d_T. When we enlarge d_n and/or d_T, the bias decreases while the variance increases and vice versa.

[5]Instead of introducing Assumption I8, we can consider asymptotic truncated MSE as Andrews (1991) and Kim and Sun (2009).

The second part of Theorem 1 (c) states that, in comparison with the variance term in part (a), the effect of using $\hat{V}_{(i,t)}$ instead of $V_{(i,t)}$ in the construction of $\hat{J}_{nT}$ is of a smaller order. Therefore, the convergence rate of $\hat{J}_{nT}$ is obtained by balancing the variance and the squared bias of $\tilde{J}_{nT}$. Accordingly, $\ell_n \ell_T = o(nT)$ is the condition for the consistency and its rate of convergence is $\sqrt{nT/\ell_n\ell_T}$. It is also required that $d_T = o\left(T\right)$ and $n_2\left(d_n\right) = o\left(n\right)$ to control the boundary effects. Let $\eta_T = 1$. If we set $\ell_n = O(d_n^{\eta_n})$ and $\ell_T = O(d_n^{\eta_T})$ for some $\eta_n > 0$, then the rate of convergence under the rate condition $d_n^{2q}\ell_n\ell_T/nT \to \tau \in (0,\infty)$ is $(nT)^{q/(2q+\eta_n+\eta_T)}$.

As $\hat{J}_{nT}$ is consistent, the limiting distribution of the Wald statistic is a χ_g^2 distribution. This is rather standard. Under H_0

$$W_{nT} \xrightarrow{d} \chi_g^2 \text{ and } F_{nT} \xrightarrow{d} \chi_g^2/g.$$

4 Optimal bandwidth selection procedure

This section presents optimal bandwidth choice in the sense of minimizing the upper bound of AMSE of $\hat{J}_{nT}$ and proposes a parametric plug-in procedure for practical implementation. Let

$$B_{11} := \text{vec}\left(b_1^{(q)}\right)' S_{nT} \text{vec}\left(b_1^{(q)}\right),$$
$$B_{22} := \text{vec}\left(b_2^{(q)}\right)' S_{nT} \text{vec}\left(b_2^{(q)}\right),$$
$$B_{12} := \text{vec}\left(b_1^{(q)}\right)' S_{nT} \text{vec}\left(b_2^{(q)}\right).$$

Then, up to smaller order terms

$$AMSE = K_q^2\left(\frac{B_{11}}{d_n^{2q}} + 2\frac{B_{12}}{d_n^q d_T^q} + \frac{B_{22}}{d_T^{2q}}\right) + \frac{\ell_n\ell_T}{nT}\bar{\mathcal{K}}_1\bar{\mathcal{K}}_2 \text{tr}\left[S_{nT}\left(I_{pp} + \mathbb{K}_{pp}\right)(J \otimes J)\right] \tag{12}$$

$$\leq 2K_q^2\left(\frac{B_{11}}{d_n^{2q}} + \frac{B_{22}}{d_T^{2q}}\right) + \frac{\ell_n\ell_T}{nT}\bar{\mathcal{K}}_1\bar{\mathcal{K}}_2 \text{tr}\left[S_{nT}\left(I_{pp} + \mathbb{K}_{pp}\right)(J \otimes J)\right] \tag{13}$$
$$:= AMSE^*,$$

where we have used the Cauchy inequality

$$2\frac{B_{12}}{d_n^q d_T^q} \leq \frac{B_{11}}{d_n^{2q}} + \frac{B_{22}}{d_T^{2q}}.$$

AMSE* can be regarded as AMSE in the worst case:

$$AMSE^* = \max_{(b_1,b_2)\in\mathfrak{B}} AMSE,$$

where

$$\mathfrak{B} = \Big\{(b_1,b_2) : \text{vec}\left(b_1^{(q)}\right)' S_{nT} \text{vec}\left(b_1^{(q)}\right) = B_{11},$$
$$\text{vec}\left(b_2^{(q)}\right)' S_{nT} \text{vec}\left(b_2^{(q)}\right) = B_{22}\Big\}.$$

We select $(d_n^\star, d_T^\star)$ to minimize the AMSE*:

$$
\begin{aligned}
(d_n^\star, d_T^\star) &= \underset{d_n, d_T}{\arg\min}\ AMSE^* \\
&= \underset{d_n, d_T}{\arg\min}\ 2K_q^2 \left(\frac{B_{11}}{d_n^{2q}} + \frac{B_{22}}{d_T^{2q}} \right) + \frac{\ell_n \ell_T}{nT} \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C, \qquad (14)
\end{aligned}
$$

where $C = \operatorname{tr}\left[S_{nT}(I_{pp} + \mathbb{K}_{pp})(J \otimes J)\right]$. Here, we use the AMSE* instead of the AMSE as the criterion. In the HAC estimation literature, it is standard practice to use the AMSE criterion for bandwidth selection, e.g. Andrews (1991) and Newey and West (1994). In our setting, though, this criterion is intractable. Suppose $B_{12} = -\sqrt{B_{11}B_{22}}$. This may occur when we are interested in a single component of β. In this case, bandwidth choices satisfying $d_n^q / d_T^q = \sqrt{B_{11}/B_{22}}$ make the first order bias terms cancel out with each other. Therefore, in theory, the trade-off becomes between the second order bias and the variance term. However, this choice is infeasible in practice. As B_{11}/B_{22} is unknown, we have to estimate this ratio and there is a bias in this estimation. As this bias is of the same order as the first order bias, bandwidth selection based on the trade-off between the second order bias and variance does not work. In contrast, our minimax criterion is simple to implement, as $d_n^\star$ and $d_T^\star$ depend only on two bias terms but not on their interaction. Our criterion also effectively controls for the MSE in terms of the upper bound.

Under the boundary condition in the time dimension, $\ell_T/T \to 0$, $\ell_T = 2d_T + o(d_T)$. In some cases, it is also possible to approximate ℓ_n as a function of d_n. For example, if individuals are on a 2-dimensional lattice and the Euclidean distance is used, $\ell_n = \pi d_n^2$ would be a reasonable approximation. With the specification of $\ell_n = \alpha_n d_n^{\eta_n}$ and $\ell_T = \alpha_T d_T^{\eta_T}$, we obtain explicit formulas of $d_n^\star$ and $d_T^\star$ as follows:

$$
d_n^\star = \left(\frac{4qK_q^2 B_{11}}{\eta_n \alpha_n \alpha_T \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C} nT \right)^{1/(2q+\eta_n+\eta_T)} \left(\frac{\eta_T B_{11}}{\eta_n B_{22}} \right)^{\eta_T/[2q(2q+\eta_n+\eta_T)]}, \qquad (15)
$$

$$
d_T^\star = \left(\frac{4qK_q^2 B_{22}}{\eta_T \alpha_n \alpha_T \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C} nT \right)^{1/(2q+\eta_n+\eta_T)} \left(\frac{\eta_n B_{22}}{\eta_T B_{11}} \right)^{\eta_n/[2q(2q+\eta_n+\eta_T)]}. \qquad (16)
$$

(15) and (16) show that the degree of persistence in one dimension affects both $d_n^\star$ and $d_T^\star$ but in the opposite direction. For example, if a process becomes spatially persistent, $d_n^\star$ is increased to address the increasing bias, which comes from the usage of kernel in the spatial domain. But, the increase of $d_n^\star$, at the same time, magnifies the variance term. Therefore, in order to minimize AMSE*, $d_T^\star$ is decreased to moderate the inflation of the asymptotic variance. Figure 1 illustrates this relation of $d_n^\star$ and $d_T^\star$ with different dependence structure. The two graphs are the level curves of $d_n^\star$ and $d_T^\star$ as functions of λ and ρ, which determine the spatial and temporal persistence respectively in the following DGP:

$$
V_t = \lambda V_{t-1} + u_t,\ u_t = \rho W_n u_t + \varepsilon_t \text{ and } \varepsilon_t \sim (0, I_n),
$$

where V_t, u_t and ε_t are n-vectors such as $V_t = \left(V_{(1,t)}, V_{(2,t)}, \ldots, V_{(n,t)}\right)'$ and W_n is a spatial weight matrix. These two graphs indicate that $d_n^\star$ increases as spatial dependence increases or temporal dependence decreases and that $d_T^\star$ increases as temporal dependence grows or spatial dependence is reduced.

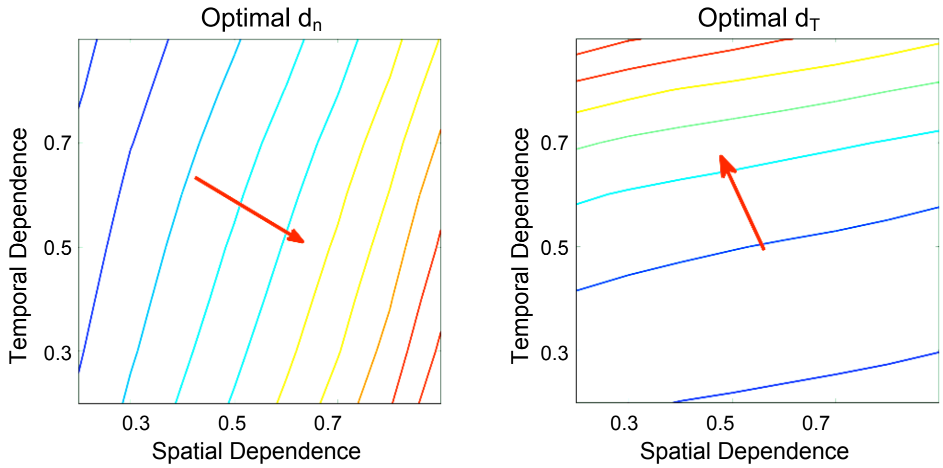

Figure 1 – Level curves of $d_n^\star$ and $d_T^\star$ as functions of spatial and temporal dependences

Corollary 1 *Suppose Assumptions I1-I9 hold. Assume that* $\ell_n = \alpha_n d_n^{\eta_n}$ *and* $\ell_T = \alpha_T d_T^{\eta_T}$ *for some* $\eta_n, \eta_T > 0$, $\alpha_n = \alpha_1 + o(1)$ *and* $\alpha_T = \alpha_2 + o(1)$. *Then, for any sequence of bandwidth parameters* $\{d_n, d_T\}$ *such that* $d_n^{2q+\eta_n} d_T^{\eta_T}/nT \to \tau \in (0, \infty)$, $\{d_n^\star, d_T^\star\}$ *is preferred in the sense that*

$$\lim_{n,T\to\infty} \left[\max_{(b_1,b_2)\in\mathfrak{B}} MSE\left((nT)^{2q/(2q+\eta_n+\eta_T)}, \hat{J}_{nT}(d_n, d_T), S_{nT} \right) - \max_{(b_1,b_2)\in\mathfrak{B}} MSE\left((nT)^{2q/(2q+\eta_n+\eta_T)}, \hat{J}_{nT}(d_n^\star, d_T^\star), S_{nT} \right) \right] \geq 0.$$

The inequality is strict unless $d_n = d_n^\star + o\left((nT)^{1/(2q+\eta_n+\eta_T)} \right)$ *and* $d_T = d_T^\star + o\left((nT)^{1/(2q+\eta_n+\eta_T)} \right)$.

Our bandwidth selection procedure does not apply directly to the rectangular kernel, and more broadly, flat-top kernel estimators because its asymptotic bias is of smaller order than the one in Theorem 1(b). However, it is interesting to consider flat-top kernel estimators because they are higher order accurate (Politis, 2010). In time series HAC estimation, Andrews (1991, footnote on p. 834) and Lin and Shinichi (2009) suggest a practical rule for the rectangular kernel estimator based on the AMSE criterion. Sun and Kaplan (2010) explore this problem rigorously and provide a bandwidth selection procedure that is testing optimal. Both the methods lead the rectangular kernel estimator to better asymptotic properties than any finite order kernel estimator we target. We obtain the optimal bandwidth parameters for the rectangular kernel, $(d_{rec,n}^\star, d_{rec,T}^\star)$ by extending these preceding methods. This is particularly important in our setting because the rectangular kernel is completely compatible with the adaptiveness of our estimator as explained below while finite order kernels yield some discrepancy.

Let $K_{tar}(\cdot)$ be the target kernel and $(d_{tar,n}^\star, d_{tar,T}^\star)$ be its optimal bandwidth parameters. Any finite order kernel can be our target. Given $\ell_n = \alpha_n d_n^{\eta_n}$ and $\ell_T = \alpha_T d_T^{\eta_T}$, if we set

$$d_{rec,n}^\star = d_{tar,n}^\star \left(\frac{\bar{\mathcal{K}}_{tar,1}}{\bar{\mathcal{K}}_{rec,1}} \right)^{1/\eta_n} \text{ and } d_{rec,T}^\star = d_{tar,T}^\star \left(\frac{\bar{\mathcal{K}}_{tar,2}}{\bar{\mathcal{K}}_{rec,2}} \right)^{1/\eta_T}, \tag{17}$$

then the asymptotic variance of the rectangular-kernel estimator is the same as that of the estimator based on the target kernel. However, under some smootherness conditions, the asymptotic bias of the rectangular-kernel estimator is of smaller order. As a result, the rectangular-kernel estimator has smaller AMSE* than the estimator based on the target kernel.

As (14) is the function of unknown values such as B_{11}, B_{22} and C, they need to be estimated for implementation with given data in a parametric (e.g. Andrews, 1991; and Kim and Sun, 2010) or nonparametric way (e.g. Newey and West, 1994). In this paper, we suggest a parametric plug-in method. We consider the following four different spatiotemporal parametric models, which are introduced in Anselin (2001).

$$V_{(i,t)}^{(c)} = \rho_c \left[W_n^{(c)} V_{t-1}^{(c)} \right]_i + \tilde{\varepsilon}_{(i,t)}^{(c)}, \tag{18}$$

$$V_{(i,t)}^{(c)} = \lambda_c V_{(i,t-1)}^{(c)} + \rho_c \left[W_n^{(c)} V_{t-1}^{(c)} \right]_i + \tilde{\varepsilon}_{(i,t)}^{(c)} \tag{19}$$

$$V_{(i,t)}^{(c)} = \lambda_c V_{(i,t-1)}^{(c)} + \phi_c \left[W_n^{(c)} V_{t}^{(c)} \right]_i + \tilde{\varepsilon}_{(i,t)}^{(c)} \tag{20}$$

$$V_{(i,t)}^{(c)} = \lambda_c V_{(i,t-1)}^{(c)} + \phi_c \left[W_n^{(c)} V_{t}^{(c)} \right]_i + \rho_c \left[W_n^{(c)} V_{t-1}^{(c)} \right]_i + \tilde{\varepsilon}_{(i,t)}^{(c)} \tag{21}$$

where $\tilde{\varepsilon}_{(i,t)}^{(c)} \overset{i.i.d}{\sim} (0, \sigma_{cc})$ and $[W_n^{(c)} V_t^{(c)}]_i$ is the i^{th} element of vector $W_n^{(c)} V_t^{(c)}$. The spatial weight matrix $W_n^{(c)}$ is determined a priori and by convention it is row-standardized and its diagonal elements are zeros.

For an illustrative purpose, let's consider the model in (18). It can be rewritten in a recursive way as follows:

$$\begin{aligned}
V_1^{(c)} &= \rho_c W_n^{(c)} V_0^{(c)} + I_n \tilde{\varepsilon}_1^{(c)} \\
V_2^{(c)} &= \rho_c^2 \left(W_n^{(c)} \right)^2 V_0^{(c)} + \rho_c W_n^{(c)} \tilde{\varepsilon}_1^{(c)} + I_n \tilde{\varepsilon}_2^{(c)} \\
&\vdots \\
V_T^{(c)} &= \rho_c^T \left(W_n^{(c)} \right)^T V_0^{(c)} + \rho_c^{T-1} \left(W_n^{(c)} \right)^{T-1} \tilde{\varepsilon}_1^{(c)} + \rho_c^{T-2} \left(W_n^{(c)} \right)^{T-2} \tilde{\varepsilon}_2^{(c)} + \ldots + I_n \tilde{\varepsilon}_T^{(c)}
\end{aligned}$$

Imposing the initial condition of $V_0 = 0$, we can estimator ρ_c by OLS with $\hat{V}_t^{(c)} = (\hat{V}_{(1,t)}^{(c)}, \ldots, \hat{V}_{(n,t)}^{(c)})'$. We define

$$\hat{\tilde{R}}_{ts}^{(c)} = \begin{cases} I_n, & \text{if } t - s = 0 \\ \left(\hat{\rho}_c W_n^{(c)} \right)^{t-s}, & \text{if } t - s > 0 \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\hat{\tilde{R}}_{(i,t)}^{(c)} = \left[\hat{\tilde{R}}_{t1,i}^{(c)}, \hat{\tilde{R}}_{t2,i}^{(c)}, \ldots, \hat{\tilde{R}}_{tT,i}^{(c)} \right],$$

where $\hat{\tilde{R}}_{ts,i}^{(c)}$ denotes the i-th row of $\hat{\tilde{R}}_{ts}^{(c)}$. Consequently, we approximate J, $b_1^{(q)}$ and $b_2^{(q)}$ by

$$\hat{J}\left(c,d\right)=\frac{\hat{\sigma}_{cd}}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{\tilde{R}}_{(i,t)}^{(c)}\right)\left(\hat{\tilde{R}}_{(j,s)}^{(d)}\right)', \tag{22}$$

$$\hat{b}_1^{(q)}\left(c,d\right)=\frac{\hat{\sigma}_{cd}}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{\tilde{R}}_{(i,t)}^{(c)}\right)\left(\hat{\tilde{R}}_{(j,s)}^{(d)}\right)' d_{ij}^{q}, \tag{23}$$

$$\hat{b}_2^{(q)}\left(c,d\right)=\frac{\hat{\sigma}_{cd}}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{\tilde{R}}_{(i,t)}^{(c)}\right)\left(\hat{\tilde{R}}_{(j,s)}^{(d)}\right)' d_{ts}^{q}, \tag{24}$$

where

$$\hat{\sigma}_{cd}=\frac{1}{n(T-1)-1}\left(\hat{\varepsilon}^{(c)}\right)'\left(\hat{\varepsilon}^{(d)}\right),$$

$\hat{\varepsilon}^{(c)}=((\hat{\varepsilon}_1^{(c)})',...,(\hat{\varepsilon}_T^{(c)})')'$, $\hat{\varepsilon}_1^{(c)}=\hat{V}_1^{(c)}$ and $\hat{\varepsilon}_t^{(c)}=\hat{V}_t^{(c)}-\hat{\rho}_c W_n^{(c)}\hat{V}_{t-1.}^{(c)}$ for $t\geq 2$. Substituting these estimators into (14) for the true parameters, we obtain the data dependent bandwidth parameters, $(\hat{d}_n,\hat{d}_T)$ as follows:

$$\left(\hat{d}_n,\hat{d}_T\right)=\underset{d_n,d_T}{\arg\min}\ 2K_q^2\left(\frac{\hat{B}_{11}}{d_n^{2q}}+\frac{\hat{B}_{22}}{d_T^{2q}}\right)+\frac{\ell_n\ell_T}{nT}\hat{C}. \tag{25}$$

where

$$\hat{B}_{11}=\mathrm{vec}\left(\hat{b}_1^{(q)}\right)' S_{nT}\mathrm{vec}\left(\hat{b}_1^{(q)}\right),$$
$$\hat{B}_{22}=\mathrm{vec}\left(\hat{b}_2^{(q)}\right)' S_{nT}\mathrm{vec}\left(\hat{b}_2^{(q)}\right),$$
$$\hat{C}=\mathrm{tr}\left[S_{nT}(I+\mathbb{K}_{pp})(\hat{J}\otimes\hat{J})\right].$$

Correspondingly, using the specification of $\ell_n=\alpha_n d_n^{\eta}$ we obtain

$$\hat{d}_n=\left(\frac{4qK_q^2\hat{B}_{11}}{\eta_n\alpha_n\alpha_T\bar{\mathcal{K}}_1\bar{\mathcal{K}}_2\hat{C}}nT\right)^{1/(2q+\eta_n+\eta_T)}\left(\frac{\eta_T\hat{B}_{11}}{\eta_n\hat{B}_{22}}\right)^{\eta_T/[2q(2q+\eta_n+\eta_T)]}, \tag{26}$$

$$\hat{d}_T=\left(\frac{4qK_q^2\hat{B}_{22}}{\eta_T\alpha_n\alpha_T\bar{\mathcal{K}}_1\bar{\mathcal{K}}_2\hat{C}}nT\right)^{1/(2q+\eta_n+\eta_T)}\left(\frac{\eta_n\hat{B}_{22}}{\eta_T\hat{B}_{11}}\right)^{\eta_n/[2q(2q+\eta_n+\eta_T)]}. \tag{27}$$

It also follows

$$\hat{d}_{rec,n}=\hat{d}_{tar,n}\left(\frac{\bar{\mathcal{K}}_{tar,1}}{\bar{\mathcal{K}}_{rec,1}}\right)^{1/\eta_n}\ \text{ and }\ \hat{d}_{rec,T}=\hat{d}_{tar,T}\left(\frac{\bar{\mathcal{K}}_{tar,2}}{\bar{\mathcal{K}}_{rec,2}}\right)^{1/\eta_T}. \tag{28}$$

Since the models in (19), (20) and (21) can be rewritten as

$$V_{(i,t)}^{(c)} = \left[\left(\lambda_c I_n + \rho_c W_n^{(c)}\right) V_{t-1}^{(c)}\right]_i + \tilde{\varepsilon}_{it}^{(c)},$$
$$V_{(i,t)}^{(c)} = \left[\lambda_c \left(I_n - \phi_c W_n^{(c)}\right)^{-1} V_{t-1}^{(c)}\right]_i + \left[\left(I_n - \phi_c W_n^{(c)}\right)^{-1} \tilde{\varepsilon}_t^{(c)}\right]_i,$$
$$V_{(i,t)}^{(c)} = \left[\left(I_n - \phi_c W_n^{(c)}\right)^{-1} \left(\lambda_c I_n + \rho_c W_n^{(c)}\right) V_{t-1}^{(c)}\right]_i + \left[\left(I_n - \phi_c W_n^{(c)}\right)^{-1} \tilde{\varepsilon}_t^{(c)}\right]_i,$$

we can derive the data dependent bandwidth parameters with these models using the same procedures as (18). While the OLS estimator is consistent for (19), it is not for (20) and (21) due to endogeneity of $[W_n^{(c)} V_t^{(c)}]_i$. For these models, we can have consistent estimators using QMLE as follows:

$$\left(\hat{\lambda}_c, \hat{\phi}_c, \hat{\rho}_c, \hat{\sigma}_{cc}\right) = \underset{\lambda_c, \phi_c, \rho_c, \sigma_{cc}}{\arg\min} \frac{1}{2} \ln \sigma_{cc} - \frac{1}{n} \ln \left| I_n - \phi_c W_n^{(c)} \right| + \frac{1}{2\sigma_{cc}} \frac{1}{nT} \sum_{t=1}^{T} \left(\hat{\varepsilon}_t^{(c)}\right)' \left(\hat{\varepsilon}_t^{(c)}\right).$$

See Yu, de Jong and Lee (2008) for detail. In fact, however, the simple OLS can still be used for (20) and (21). Since the parametric models are most likely to be mis-specified, the QML estimator is not necessarily preferred. In addition, as argued by Andrews (1991), good performance of the estimator only requires $(\hat{d}_n, \hat{d}_T)$ to be near the optimal bandwidth values and not to be precisely equal to them. Furthermore, OLS estimation is computationally much less demanding.

5 Comparison with CCE, DK and KP estimators

For comparison, we examine the asymptotic properties of the CCE, DK and KP estimators based on our data representation in (5) and (6) under the increasing smoothing asymptotics. We also derive the optimal bandwidth parameters for DK and KP estimators using the AMSE criterion.

5.1 CCE

The CCE is defined as

$$\hat{J}_{nT}^{A} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{V}_{(i,t)} \hat{V}_{(i,s)}'.$$

The critical condition for this estimator to be consistent is that each variable from two different individuals (or clusters) is uncorrelated, i.e. $EV_{(i,t)} V_{(j,s)}' = 0$ if $i \neq j$. Under this condition, $\hat{J}_{nT}^{A}$ is robust to heteroskedasticity and arbitrary forms of serial correlation. Our spatiotermporal model accommodates spatial independence by imposing the following restriction.

Assumption I10 $\tilde{r}_{(it,j,s)} = 0$ *if* $i \neq j$.

Under Assumption I10,

$$J_{nT} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[V_{(i,t)} V_{(i,s)}'\right] := J_{nT}^{A}.$$

Assumption I11 *For all i,*

$$\lim_{T\to\infty} var\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T} V_{(i,s)}\right) = J.$$

Assumption I11 implies the homogeneity of $var(T^{-1/2}\sum_{s=1}^{T} V_{(i,s)})$, under which we can derive the asymptotic variance of $\tilde{J}_{nT}^{A}$ in Theorem 2 (a) below.

Theorem 2 *Suppose that Assumptions I1, I2, I10 and I11 hold.*

(a) $\lim_{n,T\to\infty} n \cdot var\left(vec(\tilde{J}_{nT}^{A})\right) = (I_{pp} + \mathbb{K}_{pp})(J \otimes J)$.

(b) *If Assumption I7 holds, then* $\sqrt{n}\left(\hat{J}_{nT}^{A} - J_{nT}^{A}\right) = O_p(1)$ *and* $\sqrt{n}\left(\hat{J}_{nT}^{A} - \tilde{J}_{nT}^{A}\right) = o_p(1)$.

Proofs are given in the appendix. Theorem 2 implies $\sqrt{n}$-convergence of $\hat{J}_{nT}^{A}$ as $n, T \to \infty$, which is consistent with Hansen (2007).

5.2 DK estimator

The DK estimator is based on the time series HAC estimation method with cross-sectional averages. The estimator is defined as

$$\hat{J}_{nT}^{DK} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} K\left(\frac{d_{ts}}{d_T}\right)\hat{V}_{(i,t)}\hat{V}_{(j,s)}', \tag{29}$$

For the asymptotic properties, we introduce the following assumptions in place of Assumptions I3 and I6.

Assumption I12 *There exists $q_d > 0$ such that*

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\|\Gamma_{(it,js)}\right\| d_{ts}^{q_d} < \infty$$

for all n, T.

Assumption I13 *For $t \in E_T$,*

$$\lim_{n,T\to\infty} var\left(\frac{1}{\sqrt{n\ell_T}}\sum_{j=1}^{n}\sum_{s:d_{ts}\le d_T} V_{(j,s)}\right) = J.$$

Compared with Assumption I3, Assumption I12 is sufficient for $\hat{J}_{nT}^{DK}$ because $\hat{J}_{nT}^{DK}$ is not involved with the bias that arises from a kernel in the spatial dimension. Theorem 3 below states the asymptotic properties of $\hat{J}_{T}^{DK}$.

Theorem 3 *Suppose that Assumptions I1, I2, I5(i) and (ii), I12 and I13 hold, and* $d_T \to \infty$, $\ell_T = o(T)$.

(a) $\lim_{n,T\to\infty} \frac{T}{\ell_T} var\left(vec\tilde{J}_{nT}^{DK}\right) = \bar{\mathcal{K}}_2(I_{pp} + \mathbb{K}_{pp})\,(J \otimes J)$.

(b) $\lim_{n,T\to\infty} d_T^q(E\tilde{J}_{nT}^{DK} - J_{nT}) = -K_q b_2^{(q)}$

(c) *If Assumption I7 holds and* $d_T^{2q}\ell_T/T \to \tau \in (0,\infty)$, *then* $\sqrt{\frac{T}{\ell_T}}\left(\hat{J}_{nT}^{DK} - J_{nT}\right) = O_p(1)$ *and* $\sqrt{\frac{T}{\ell_T}}\left(\hat{J}_{nT}^{DK} - \tilde{J}_{nT}^{DK}\right) = o_p(1)$.

(d) *Under the conditions of part (c) and Assumption I9,*

$$\begin{aligned}
&\lim_{n,T\to\infty} MSE\left(\frac{T}{\ell_T}, \hat{J}_{nT}^{DK}, S_{nT}\right) \\
&= \lim_{n,T\to\infty} MSE\left(\frac{T}{\ell_T}, \tilde{J}_{nT}^{DK}, S\right) \\
&= \frac{1}{\tau}K_q^2\left(vec\ b_2^{(q)}\right)' S\left(vec\ b_2^{(q)}\right) + \bar{\mathcal{K}}_2 tr\left[S(I_{pp} + \mathbb{K}_{pp})(J \otimes J)\right].
\end{aligned}$$

Proofs are given in the appendix. Theorem 3 (a) and (b) imply that $\hat{J}_{nT}^{DK}$ is consistent if $d_T \to \infty$ and $\ell_T = o(T)$. The rate of convergence obtained by balancing the variance and the squared bias is $T^{q/(2q+\eta_T)}$. Therefore, $\hat{J}_{nT}$ is more efficient than $\hat{J}_{nT}^{DK}$ if $T = o(n^{(2q+\eta_T)/\eta_n})$.

The optimal bandwidth parameter of $\hat{J}_{nT}^{DK}$ based on the AMSE criterion is

$$d_T^{DK} = \left(\frac{2qK_q^2 B_{22}}{\eta_T \alpha_T \bar{\mathcal{K}}_2 C}T\right)^{1/(2q+\eta_T)}, \tag{30}$$

where $C = \mathrm{tr}\left[S_{nT}(I_{pp} + \mathbb{K}_{pp})(J \otimes J)\right]$. Following Andrews (1991) and Newey and West (1994), we can obtain the data dependent bandwidth parameter.

5.3 KP estimator

Analogous to the DK estimator, we can also consider the usage of spatial HAC estimation with the averages across time especially when n is large. The KP estimator with the serial averages is

$$\hat{J}_{nT}^{KP} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right)\hat{V}_{(i,t)}\hat{V}_{(j,s)}'.$$

Assumption I14 *There exists* $q_1 > 0$ *such that*

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\|\Gamma_{(it,js)}\right\| d_{ij}^{q_1} < \infty$$

for all n, T.

Assumption I15 *For $i \in E_n$,*

$$\lim_{n,T\to\infty} var\left(\frac{1}{\sqrt{\ell_n T}} \sum_{j:d_{ij}\le d_n} \sum_{s=1}^{T} V_{(j,s)}\right) = J.$$

Theorem 4 below states the asymptotic properties of $\hat{J}_{nT}^{KP}$.

Theorem 4 *Suppose that Assumptions I1, I2, I4, I5, I14 and I15 hold, $n_2/n \to 0$, $\ell_n, d_n \to \infty$ and $\ell_n/n \to 0$.*

(a) $\lim_{n,T\to\infty} \frac{n}{\ell_n} var\left(vec\tilde{J}_{nT}^{KP}\right) = \bar{\mathcal{K}}_1 (I_{pp} + \mathbb{K}_{pp})\,(J \otimes J)$.

(b) $\lim_{n,T\to\infty} d_n^q (E\tilde{J}_{nT}^{KP} - J_{nT}) = -K_q b_1^{(q)}$

(c) *If Assumption I7 holds and $d_n^{2q}\ell_n/n \to \tau \in (0,\infty)$, then $\sqrt{\frac{n}{\ell_n}}\left(\hat{J}_{nT}^{KP} - J_{nT}\right) = O_p(1)$ and $\sqrt{\frac{n}{\ell_n}}\left(\hat{J}_{nT}^{KP} - \tilde{J}_{nT}^{KP}\right) = o_p(1)$.*

(d) *Under the conditions of part (c) and Assumption I9,*

$$\begin{aligned}
&\lim_{n,T\to\infty} MSE\left(\frac{n}{\ell_n}, \hat{J}_{nT}^{KP}, S_{nT}\right) \\
&= \lim_{n,T\to\infty} MSE\left(\frac{n}{\ell_n}, \tilde{J}_{nT}^{KP}, S\right) \\
&= \frac{1}{\tau} K_q^2 vec\left(b_1^{(1)}\right)' Svec\left(b_1^{(q)}\right) + \bar{\mathcal{K}}_1 tr\left[S(I_{pp} + \mathbb{K}_{pp})(J \otimes J)\right].
\end{aligned}$$

Proofs are given in the appendix. If we can characterize $\ell_n = \alpha_n d_n^{\eta_n}$, $\hat{J}_{nT}$ achieves the faster convergence rate than $\hat{J}_{nT}^{KP}$ if $n = o(T^{(2q+\eta_n)/\eta_T})$. The optimal bandwidth based on the AMSE criterion is

$$d_n^{\star KP} = \left(\frac{2qK_q^2 B_{11}}{\eta_n \alpha_n \bar{\mathcal{K}}_1 C} n\right)^{1/(2q+\eta_n)}. \tag{31}$$

6 Adaptiveness of $\hat{J}_{nT}$

6.1 Flexibility

$\hat{J}_{nT}$ is flexible in the sense that it includes the estimators in the previous section as special cases, reducing to each of them with certain choice of the bandwidths and kernel function. In order to illustrate the flexibility, we introduce the generalized CCE, $\hat{J}_{nT}^{GA}$:

$$\hat{J}_{nT}^{GA} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} K_{RE}\left(\frac{d_{ts}}{d_T}\right) \hat{V}_{(i,t)} \hat{V}'_{(i,s)},$$

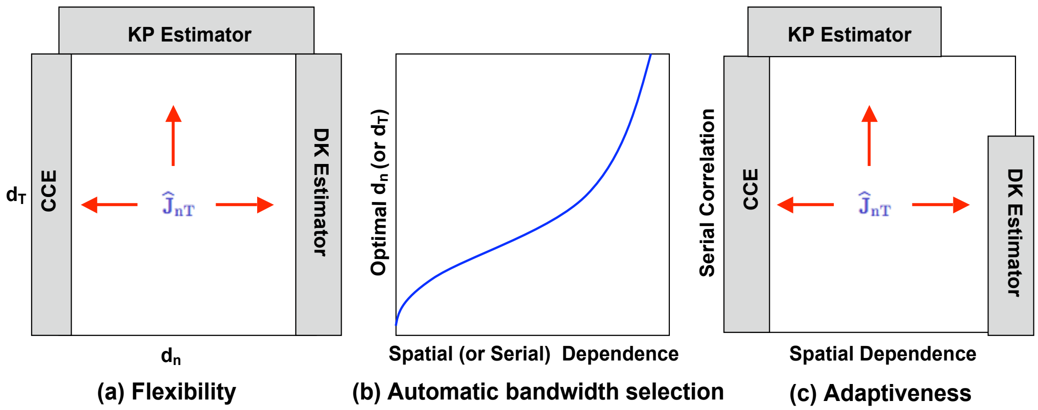

Figure 2 – Adaptiveness of $\hat{J}_{nT}$

where $K_{RE}(\cdot)$ is the rectangular kernel defined as

$$K_{RE}(x) = \begin{cases} 1 & \text{if } |x| \le c, \text{ where } c \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

The rectangular kernel is considered in early HAC estimation literature (e.g. Domowitz and White, 1982) and recently has gained further attention (Lin and Shinichi, 2009; and Sun and Kaplan 2010). It is the special case of flat-top kernels (Politis, 2010).

The following proposition shows the asymptotic equivalence of $\hat{J}_{nT}$ to the existing estimators with certain sequences of d_n and d_T.

Proposition 1 *Let's consider $\hat{J}_{nT}$ with the rectangular kernel.*

(a) *If $d_n \to 0$ as $n \to \infty$, then $\hat{J}_{nT} - \hat{J}_{nT}^{GA} = o_p(1)$.*

(b) *If $\ell_n/n \to 1$ as $n \to \infty$, then $\hat{J}_{nT} - \hat{J}_{nT}^{DK} = o_p(1)$.*

(c) *If $\ell_T/T \to 1$ as $T \to \infty$, then $\hat{J}_{nT} - \hat{J}_{nT}^{KP} = o_p(1)$.*

Proofs are given in the appendix. The flexibility of our estimator relies on the property that the rectangular kernel does not downweigh the covariances between spatially or serially remote units. $\hat{J}_{nT}$ with finite order kernels does not completely reduce to $\hat{J}_{nT}^{DK}$ and $\hat{J}_{nT}^{KP}$ with large d_n and d_T, getting close to them though.

6.2 Adaptiveness

While $\hat{J}_{nT}$ has advantages in terms of robustness over $\hat{J}_{nT}^{A}$ and in terms of efficiency over $\hat{J}_{nT}^{KP}$ and $\hat{J}_{nT}^{DK}$, for certain dependence structure, one of the three estimators $\hat{J}_{nT}^{A}$, $\hat{J}_{nT}^{KP}$ and $\hat{J}_{nT}^{DK}$ is expected to out-perform other estimators. If a process is spatially highly persistent, $\hat{J}_{nT}^{DK}$ is expected to out-perform other estimators in that it is robust to arbitrary form of spatial correlation. For the same reason, $\hat{J}_{nT}^{KP}$ tends to perform well, if a process is serially highly persistent. $\hat{J}_{nT}^{A}$ is asymptotically efficient in the absence of spatial correlation.

The attractivess of our estimator $\hat{J}_{nT}$ is that, with the data driven bandwidth choice, it becomes close to the estimator that is expected to perform the best. This adaptiveness is the novel feature of our estimation method. It practically automates the selection of covariance estimator. As illustrated in Figure 2, adaptiveness arises from the flexibility and automatic bandwidth selection procedure. In case that a process is spatially highly persistent, the automatic bandwidth selection procedure yields large $\hat{d}_n$ so that $\hat{J}_{nT}$ gets close to $\hat{J}_{nT}^{DK}$. Analogously, $\hat{J}_{nT}$ becomes close to $\hat{J}_{nT}^{KP}$ if a process is very persistent in the time dimension. In the absence of spatial dependence, $\hat{J}_{nT}$ become close to $\hat{J}_{nT}^{GA}$ with small $\hat{d}_n$.

It should be pointed out that finite order kernels do not achieve complete adaptiveness because downweighing restricts its flexibility in bridging existing estimators. We can fix this by employing the rectangular kernel. In this case, with appropriate bandwidth choices, $\hat{J}_{nT}$ is asymptoticvally equivalent to the best estimator. The bandwidth selection rule in (17) meets the requirment, as the selected bandwidths from (17) are the proportional to those from (14).[6]

7 Fixed smoothing asymptotics

7.1 Limiting theory for $\hat{J}_{nT}$ under fixed smoothing asymptotics

Following Conley (1999), we assume that, given a distance measure, it is possible to map the individuals onto a 2-dimensional integer lattice so that d_{ij} can be expressed in terms of the lattice indices. Suppose that the locations are indexed by $(i_1, i_2) = [1, 2, \ldots, L_n] \otimes [1, 2, \ldots, M_n]$. We can then rewrite the sample moment condition that defines $\hat{\beta}$ as

$$\frac{1}{L_n M_n T} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^{T} 1_{i_1,i_2} \hat{V}_{(i_1,i_2,t)} = 0,$$

where $\hat{V}_{(i_1,i_2,t)}$ is associated with an observation located at (i_1, i_2) and time t. While our analysis relies on the rectangular lattice structure, it can be potentially generalized to non-lattice case. See BCHV. As we do not assume the presence of an observation at every lattice point, we introduce the indicator function $1_{i_1,i_2}$ to denote the presence of an observation at a particular lattice point (i_1, i_2). Using this indicator function, we define

$$V^*_{(i_1,i_2,t)} = 1_{i_1,i_2} V_{(i_1,i_2,t)},\ \hat{V}^*_{(i_1,i_2,t)} = 1_{i_1,i_2} \hat{V}_{(i_1,i_2,t)} \text{ and } \tilde{X}^*_{(i_1,i_2,t)} = 1_{i_1,i_2} \tilde{X}_{(i_1,i_2,t)}.$$

We maintain the following high level assumptions.

Assumption F1 *The functional central limit theorem*

$$\frac{1}{\sqrt{L_n M_n T}} \sum_{i_1=1}^{[r_1 L_n]} \sum_{i_2=1}^{[r_2 M_n]} \sum_{t=1}^{[\tau T]} V^*_{(i_1,i_2,t)} \xrightarrow{d} \Lambda W_p\left(r_1, r_2, \tau\right)$$

[6]Another issue with flat-top kernel estimators is that they are not positive semi-definite. Politis (2010) and Lin and Sakata (2009) propose simple modification to the estimator to enforce the positive (semi) definiteness without sacrificing efficiency. In our simulation, we use the method suggested by Politis (2010).

holds for all $(r_1, r_2, \tau) \in [0,1]^3$, *where* $\Lambda\Lambda' = J$ *and* $W_p(r_1, r_2, \tau) = \left(W^{(1)}(r_1, r_2, \tau), \ldots, W^{(p)}(r_1, r_2, \tau)\right)'$ *is a p-dimensional independent Wiener process with covariance given by*

$$cov\left(W^{(i)}(r_1, r_2, \tau), W^{(j)}(v_1, v_2, \kappa)\right) = \delta_{ij} \min(r_1, v_1) \min(r_2, v_2) \min(\tau, \kappa)$$

with δ_{ij} *being a Kronecker delta.*

Assumption F2 *For all* $(r_1, r_2, \tau) \in [0,1]^3$,

$$\frac{1}{L_n M_n T} \sum_{i_1=1}^{[r_1 L_n]} \sum_{i_2=1}^{[r_2 M_n]} \sum_{t=1}^{[\tau T]} \tilde{X}^*_{(i_1,i_2,t)} \tilde{X}^{*\prime}_{(i_1,i_2,t)} \xrightarrow{p} r_1 r_2 \tau Q.$$

Assumptions F1 and F2 follow BCHV and Sun and Kim (2010). Under the above assumptions, it is easy to see that

$$\begin{aligned} \sqrt{nT}\left(\hat{\beta} - \beta\right) &= \left(\frac{1}{L_n M_n T} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^{T} \tilde{X}^*_{(i_1,i_2,t)} \tilde{X}^{*\prime}_{(i_1,i_2,t)}\right)^{-1} \frac{1}{\sqrt{L_n M_n T}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^{T} V^*_{(i_1,i_2,t)} \\ &\xrightarrow{d} Q\Lambda W_p(1,1,1) := \Lambda^* W_p(1,1,1). \end{aligned} \tag{32}$$

Therefore,

$$\begin{aligned} &\frac{1}{\sqrt{L_n M_n T}} \sum_{i_1=1}^{[r_1 L_n]} \sum_{i_2=1}^{[r_2 M_n]} \sum_{t=1}^{[\tau T]} \hat{V}^*_{(i_1,i_2,t)} \\ &= \frac{1}{\sqrt{L_n M_n T}} \sum_{i_1=1}^{[r_1 L_n]} \sum_{i_2=1}^{[r_2 M_n]} \sum_{t=1}^{[\tau T]} V^*_{(i_1,i_2,t)} - \frac{1}{L_n M_n T} \sum_{i_1=1}^{[r_1 L_n]} \sum_{i_2=1}^{[r_2 M_n]} \sum_{t=1}^{[\tau T]} \tilde{X}^*_{(i_1,i_2,t)} \tilde{X}^{*\prime}_{(i_1,i_2,t)} \sqrt{L_n M_n T}\left(\hat{\beta} - \beta\right) \\ &\xrightarrow{d} \Lambda\left[W_p(r_1, r_2, \tau) - r_1 r_2 \tau W_p(1,1,1)\right] \\ &:= \Lambda B_p(r_1, r_2, \tau), \end{aligned}$$

where $B_p(r_1, r_2, \tau)$ is a p-dimensional tied-down Brownian sheet. The second term in the equality reflects the estimation uncertainty in $\hat{\beta}$. We introduce the following assumption on the distance measure in the spatial dimension.

Assumption F3 *Let* $d_{(i_1,i_2),(j_1,j_2)}$ *denote* d_{ij} *between two observations at* (i_1, i_2) *and* (j_1, j_2)*. Then,*

$$\frac{d_{(i_1,i_2),(j_1,j_2)}}{d_n} = d\left(\frac{i_1 - j_1}{d_n}, \frac{i_2 - j_2}{d_n}\right).$$

Assumption F3 implies that $d_{(i_1,i_2),(j_1,j_2)}$ is the function of $i_1 - j_1$ and $i_2 - j_2$ and is homogeneous. This is not overly restrictive. p-norm distances that are usually employed in practice satisfy this assumption.

Suppose the level of smoothing is held fixed: $b_1 = d_n/L_n$, $b_2 = d_n/M_n$ and $b_3 = d_T/T$ where $b_2 = b_1\varsigma$ and $L_n/M_n = \varsigma$. Under Assumption F3, we have

$$\begin{aligned}
\hat{J}_{nT} &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right) K\left(\frac{d_{ts}}{d_T}\right)\hat{V}_{(i,t)}\hat{V}_{(j,s)} \\
&= \frac{1}{L_n M_n T}\sum_{i_1=1}^{L_n}\sum_{j_1=1}^{L_n}\sum_{i_2=1}^{M_n}\sum_{j_2=1}^{M_n}\sum_{t=1}^{T}\sum_{s=1}^{T} K\left(d\left(\frac{i_1-j_1}{d_n},\frac{i_2-j_2}{d_n}\right)\right) K\left(\frac{d_{ts}}{d_T}\right)\hat{V}^*_{(i_1,i_2,t)}\hat{V}^*_{(j_1,j_2,s)} \\
&:= \frac{1}{L_n M_n T}\sum_{i_1=1}^{L_n}\sum_{j_1=1}^{L_n}\sum_{i_2=1}^{M_n}\sum_{j_2=1}^{M_n}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathbb{K}_b\left(\frac{i_1-j_1}{L_n},\frac{i_2-j_2}{M_n},\frac{t-s}{T}\right)\hat{V}^*_{(i_1,i_2,t)}\hat{V}^*_{(j_1,j_2,s)} \qquad (33)
\end{aligned}$$

where

$$\mathbb{K}(x,y,z) = K(d(x,y))K(z) = K_n(x,y)K(z)$$

and

$$\mathbb{K}_b(x,y,z) = \mathbb{K}\left(\frac{x}{b_1},\frac{y}{b_2},\frac{z}{b_3}\right) = K\left(d\left(\frac{x}{b_1},\frac{y}{b_2}\right)\right)K\left(\frac{z}{b_3}\right).$$

Define $K_n(x,y) = K(d(x,y))$ and $K_{nb}(x,y) = K(d(x/b_1, y/b_2))$ where the subscript 'n' is used to differentiate K_n, a function of two variables, from K, a function of a single variable. Note that K_n does not depend on the sample size n.

Assumption F4 *(i) $K_n(x,y) : R^2 \to [0,1]$ and $K(z) : R \to [0,1]$ are symmetric with $K(0) = 1$. (ii) $\int_0^\infty\int_0^\infty K_n(x,y)xdxdy < \infty$, $\int_0^\infty\int_0^\infty K_n(x,y)ydxdy < \infty$, $\int_0^\infty\int_0^\infty K_n(x,y)xydxdy < \infty$ and $\int_0^\infty K(z)zdz < \infty$. (iii) The Parzen characteristic exponent is greater than or equal to 1.*

All the commonly used kernels satisfy this assumption. Assumption F4 (ii) enables us to use the Riemann-Lebesgue lemma.

Since $\mathbb{K}_b(\cdot,\cdot,\cdot)$ is square integrable, it has a Fourier series representation:

$$\begin{aligned}
\mathbb{K}_b\left(\frac{i_1-j_1}{L_n},\frac{i_2-j_2}{M_n},\frac{t-s}{T}\right) &= \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}\sum_{m=1}^{\infty}\lambda_{k,\ell,m}\varphi_{b_1,k}\left(\frac{i_1-j_1}{L_n}\right)\varphi_{b_2,\ell}\left(\frac{i_2-j_2}{M_n}\right)\varphi_{b_3,m}\left(\frac{t-s}{T_n}\right) \\
&:= \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}\sum_{m=1}^{\infty}\lambda_{k,\ell,m}\Phi_{b,k\ell m}\left(\frac{i_1}{L_n},\frac{i_2}{M_n},\frac{t}{T_n}\right)\Phi_{b,k\ell m}\left(-\frac{j_1}{L_n},-\frac{j_2}{M_n},-\frac{s}{T_n}\right),
\end{aligned}$$

where $\varphi_{b,k}(x) = \exp\left(i\frac{x}{b}\pi(k-1)\right)$ and $\left\{\Phi_{b,k\ell m}\left(\frac{i_1}{L_n},\frac{i_2}{M_n},\frac{t}{T_n}\right)\Phi_{b,k\ell m}\left(-\frac{j_1}{L_n},-\frac{j_2}{M_n},-\frac{s}{T_n}\right)\right\}$ is an orthonormal basis for $L^2([0,1]^3\times[0,1]^3)$ and the convergence is in the L^2 space.

It follows from Assumption F4 (i) that

$$\sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}\sum_{m=1}^{\infty}\lambda_{k,\ell,m} = 1.$$

Using the representation, we can obtain the following results.

Proposition 2 *Under Assumptions F1 - F3,*

$$\begin{aligned}
&\hat{J}_{nT}\\
&\xrightarrow{d} \Lambda \left[\int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \mathbb{K}\left(\frac{r_1 - v_1}{b_1}, \frac{r_2 - v_2}{b_2}, \frac{\tau - \kappa}{b_3} \right) dB_p\left(r_1, r_2, \tau\right) dB_p'\left(v_1, v_2, \kappa\right) \right] \Lambda' \\
&:= \Lambda \left[\int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \mathbb{K}_b\left(r_1 - v_1, r_2 - v_2, \tau - \kappa\right) dB_p\left(r_1, r_2, \tau\right) dB_p'\left(v_1, v_2, \kappa\right) \right] \Lambda' \qquad (34)
\end{aligned}$$

Proofs are given in the appendix. It is interesting to note that the limiting distribution of $\hat{J}_{nT}$ is exactly analogous to the one in the time series setting. See Sun, Phillips and Jin (2008). Boundary effects does not exist at least under regular lattice structures.

Define the centered version of the kernel function $\mathbb{K}_b^*\left(\cdot, \cdot\right)$ as

$$\begin{aligned}
&\mathbb{K}_b^*\left(\left(r_1, r_2, \tau\right), \left(v_1, v_2, \kappa\right)\right) \\
=&\mathbb{K}_b\left(r_1 - v_1, r_2 - v_2, \tau - \kappa\right) - \int_0^1 \int_0^1 \int_0^1 \mathbb{K}_b\left(x_1 - v_1, y_1 - v_2, z_1 - \kappa\right) dx_1 dy_1 dz_1 \\
&- \int_0^1 \int_0^1 \int_0^1 \mathbb{K}_b\left(r_1 - x_2, r_2 - y_2, \tau - z_2\right) dx_2 dy_2 dz_2 \\
&+ \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \mathbb{K}_b\left(x_1 - x_2, y_1 - y_2, z_1 - z_2\right) dx_1 dy_1 dz_1 dx_2 dy_2 dz_2.
\end{aligned}$$

While $\mathbb{K}_b\left(r_1 - v_1, r_2 - v_2, \tau - \kappa\right)$ depends only on the differences $\left(r_1 - v_1, r_2 - v_2, \tau - \kappa\right)$, $\mathbb{K}_b^*\left(\left(r_1, r_2, \tau\right), \left(v_1, v_2, \kappa\right)\right)$ is in general not a function of the differences. The centered kernel function captures the estimation uncertainty in $\hat{\beta}$. Using $\mathbb{K}_b^*\left(\cdot, \cdot\right)$, (34) is equivalent to

$$\Lambda \left[\int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \mathbb{K}_b^*\left(\left(r_1, r_2, \tau\right), \left(v_1, v_2, \kappa\right)\right) dW_g\left(r_1, r_2, \tau\right) dW_p'\left(v_1, v_2, \kappa\right) \right] \Lambda'. \qquad (35)$$

With (32) and (35), we can show that under H_0,

$$\begin{aligned}
F_{nT} =& \sqrt{nT} \left[R\left(\hat{\beta} - \beta_0\right) \right]' \left(R\hat{Q}_{nT} \hat{J}_{nT} \hat{Q}_{nT}' R' \right)^{-1} \sqrt{nT} \left[R\left(\hat{\beta} - \beta_0\right) \right] / g \\
\xrightarrow{d}& \left(R\Lambda^* W_p\left(1, 1, 1\right)\right)' \\
&\times \left(R\Lambda^* \left[\int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \mathbb{K}_b^*\left(\left(r_1, r_2, \tau\right), \left(v_1, v_2, \kappa\right)\right) dW_p\left(r_1, r_2, \tau\right) dW_p'\left(v_1, v_2, \kappa\right) \right] \Lambda^{*\prime} R' \right)^{-1} \\
&\times \left(R\Lambda^* W_p\left(1, 1, 1\right)\right) / g \\
\overset{d}{=}& W_g'\left(1, 1, 1\right) \left[\int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \mathbb{K}_b^*\left(\left(r_1, r_2, \tau\right), \left(v_1, v_2, \kappa\right)\right) dW_g\left(r_1, r_2, \tau\right) dW_g'\left(v_1, v_2, \kappa\right) \right]^{-1} \\
&\times W_g\left(1, 1, 1\right) / g \\
:=& F_\infty(g, b), \qquad (36)
\end{aligned}$$

where the equality in distribution holds because

$$R\Lambda^* W_p(x,y,z) \overset{d}{=} DW_g(x,y,z)$$

for a Wiener process $W_g(x,y,z)$ and some $g \times g$ matrix D such that $DD' = RQJQ'R'$.

7.2 Expansion of the limiting distribution and F-approximation

We present the asymptotic expansion of the distribution of $F_\infty(g,b)$ in (36) and establish the validity of a standard F-approximation.

The distribution of $F_\infty(g,b)$ is nonstandard because of the random limit of $\hat{J}_{nT}$ with fixed b_1, b_2 and b_3 as $n,T \to \infty$. As b_1, b_2 and $b_3 \to 0$, however, the effect of this randomness diminishes and $gF_\infty(g,b)$ converges in distribution to the χ_g^2 distribution. Therefore, we can develop an asymptotic expansion of the distribution of $gF_\infty(g,b)$ as b_1, b_2 and $b_3 \to 0$ to examine its difference from the χ_g^2 distribution.

Let

$$\int_0^1\int_0^1\int_0^1\int_0^1\int_0^1\int_0^1 \mathbb{K}_b^*\left(\left(r_1,r_2,\tau\right),\left(v_1,v_2,\kappa\right)\right) dW_g\left(r_1,r_2,\tau\right) dW_g'\left(v_1,v_2,\kappa\right) = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}$$

and

$$\left(\int_0^1\int_0^1\int_0^1\int_0^1\int_0^1\int_0^1 \mathbb{K}_b^*\left(\left(r_1,r_2,\tau\right),\left(v_1,v_2,\kappa\right)\right) dW_g\left(r_1,r_2,\tau\right) dW_g'\left(v_1,v_2,\kappa\right)\right)^{-1} = \begin{pmatrix} v^{11} & v^{12} \\ v^{21} & v^{22} \end{pmatrix},$$

where v_{11} and v^{11} are scalars. Following Sun (2010), we can show that

$$P\left\{gF_\infty\left(g,b\right) \le z\right\} = EG_g\left(z\left(v_{11} - v_{12}v_{22}^{-1}v_{21}\right)\right) = EG_g\left(zv_{11.2}\right),$$

where $G_g(\cdot)$ is the cdf of a central χ_g^2 variate and $v_{11.2} = v_{11} - v_{12}v_{22}^{-1}v_{21}$.

As b_1, b_2 and $b_3 \to 0$, we expect $v_{11.2}$ to be concentrated around 1. By taking a Taylor expansion $G_g\left(zv_{11.2}\right)$ around $G_g\left(z\right)$ and computing the moments of $v_{11.2}$, we can prove the following theorem.

Theorem 5 *Suppose Assumptions F1 - F4 hold and $L_n/M_n \to \varsigma$. As b_1, b_2 and $b_3 \to 0$, we have*

$$P\left\{gF_\infty\left(g,b\right) \le z\right\} = G_g\left(z\right) + A\left(z\right) b_1b_2b_3 + o\left(b_1b_2b_3\right)$$

where

$$A\left(z\right) = G_g''\left(z\right) z^2 c_2 - G_g'\left(z\right) z\left[c_1 + \left(g-1\right)c_2\right],$$

$$c_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{K}_b\left(x,y,z\right) dxdydz \text{ and } c_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{K}_b^2\left(x,y,z\right) dxdydz.$$

Proofs are given in the appendix. Theorem 5 implies that the deviation of the fixed smoothing limiting distribution of the Wald statistic from the χ_g^2 distribution depends on the smoothing parameters, kernel function and the number of restrictions being tested.

While this kernel-based fixed smoothing asymptotic approximation is more accurate in size, it is not as easy to use as the conventional χ^2-approximation because the limiting distribution of F_{nT},

$F_\infty(g,b)$, is nonstandard. This contrasts with the testing procedures by Bester, Conley and Hansen (2009), Ibragimov and Müller (2010) and Sun and Kim (2010), which use t and F distributions as the reference distributions. We fill this gap in the literature by establishing the validity of an F-approximation to this distribution.

Since $\mathbb{K}_b^*((r_1,r_2,\tau),(v_1,v_2,\kappa)) \in L^2([0,1]^6)$, it has a Fourier series representation:

$$\begin{aligned}
&\mathbb{K}_b^*((r_1,r_2,\tau),(v_1,v_2,\kappa)) \\
&= \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}\sum_{m=1}^{\infty}\sum_{k'=1}^{\infty}\sum_{\ell'=1}^{\infty}\sum_{m'=1}^{\infty} \lambda_{k\ell mk'\ell'm'}\psi_{b_1,k}(r_1)\psi_{b_2,\ell}(r_2)\psi_{b_3,m}(\tau)\psi_{b_1,k'}(v_1)\psi_{b_2,\ell'}(v_2)\psi_{b_3,m'}(\kappa) \\
&:= \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}\sum_{m=1}^{\infty}\sum_{k'=1}^{\infty}\sum_{\ell'=1}^{\infty}\sum_{m'=1}^{\infty} \lambda_{k\ell mk'\ell'm'} g_{b,k\ell m}(r_1,r_2,\tau) g_{b,k'\ell'm'}(v_1,v_2,\kappa),
\end{aligned}$$

where $\{g_{b,k\ell m}(r_1,r_2,\tau)g_{b,k'\ell'm'}(v_1,v_2,\kappa)\}$ is an orthonormal basis for $L^2([0,1]^3\times[0,1]^3)$ and the convergence is in the L^2 space. As $\int_0^1\int_0^1\int_0^1 \mathbb{K}_b^*((x,y,z)(v_1,v_2,\kappa))\,dxdydz = 0$ by definition, $g_{b,k\ell m}(x,y,z)$ has the property of zero mean, i.e.

$$\int_0^1\int_0^1\int_0^1 g_{b,k\ell m}(x,y,z)dxdydz = 0.$$

Plugging this Fourier series representation, we have

$$\begin{aligned}
&\int_0^1\int_0^1\int_0^1\int_0^1\int_0^1\int_0^1 \mathbb{K}_b^*((r_1,r_2,\tau),(v_1,v_2,\kappa))\,dW_g(r_1,r_2,\tau)\,dW_g'(v_1,v_2,\kappa) \\
&= \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}\sum_{m=1}^{\infty}\sum_{k'=1}^{\infty}\sum_{\ell'=1}^{\infty}\sum_{m'=1}^{\infty} \lambda_{k\ell mk'\ell'm'}\xi_{b,k\ell m}\xi_{b,k'\ell'm'}',
\end{aligned} \quad (37)$$

where $\xi_{b,k\ell m} = \int_0^1\int_0^1\int_0^1 g_{b,k\ell m}(x,y,z)dW_g(x,y,z) \overset{i.i.d.}{\sim} N(0,I_g)$.

We can simplify the above representation. First, using the Cantor tuple function we can encode (h_1,h_2,h_3) into a single natural number h. That is,

$$h = \pi^{(3)}(h_1,h_2,h_3) := \pi^{(2)}(\pi^{(2)}(h_1,h_2),h_3),$$

where

$$\pi^{(2)}(h_1,h_2) = \frac{1}{2}(h_1+h_2)(h_1+h_2+1)+h_2.$$

The map between (h_1,h_2,h_3) and h is one-to-one and onto. With this definition, we abuse the notation a little and write

$$\lambda_{h_1h_2h_3h_1'h_2'h_3'} = \lambda_{hh'} \text{ and } \xi_{b,h_1h_2h_3} = \xi_{b,h}.$$

Then, (37) can be rewritten as

$$\sum_{h=1}^{\infty}\sum_{h'=1}^{\infty}\lambda_{hh'}\xi_{b,h}\xi_{b,h'}'. \quad (38)$$

With (38), we can follow Sun and Kaplan (2010). Note that

$$\sum_{h=1}^{\infty}\sum_{h'=1}^{\infty}\lambda_{hh'}\xi_{b,h}\xi'_{b,h'} = \lim_{N\to\infty}\sum_{h=1}^{N}\sum_{h'=1}^{N}\lambda_{hh'}\xi_{b,h}\xi'_{b,h'} = \lim_{N\to\infty}\xi'\Delta_N\xi,$$

where $\xi_{N\times g} = (\xi_1, \ldots, \xi_N)'$ and $\Delta_N = (\lambda_{hh'})$ is a $N \times N$ symmetric matrix due to $\lambda_{hh'} = \lambda_{h'h}$. Let $\Delta_N = HD_NH'$ be the spectral decomposition of $\Delta_N = diag(\lambda_1, \ldots, \lambda_N)$ and $H'H = HH' = I_N$. Then,

$$\lim_{N\to\infty}\xi'\Delta_N\xi = \lim_{N\to\infty}\left(\xi'H\right)D_N\left(H'\xi\right) = \sum_{k=1}^{\infty}\lambda_k\zeta_k\zeta'_k,$$

where $\zeta := (\zeta_1, \ldots, \zeta_N)' = H'\xi$. It is easy to see that $\zeta_k \overset{i.i.d.}{\sim} N(0, I_g)$. By definition, $\zeta_k\zeta'_k$ is a Wishart distribution $\mathbb{W}_g(I_g, 1)$, so $\sum_{k=1}^{\infty}\lambda_k\zeta_k\zeta'_k$ is an infinite weighted sum of independent Wishart distributions.

Let $\phi = W_g(1, 1, 1)$. Then, we have

$$gF_{\infty}(g, b) = \phi'\left[\sum_{k=1}^{\infty}\lambda_k\zeta_k\zeta'_k\right]^{-1}\phi, \tag{39}$$

where ζ_k is independent of ϕ for all k.

Motivated from the spectral density estimation literature (e.g. Priestley, 1981, p. 467), Sun (2010) shows that the infinite weighted sum of Wishart distributions can be approximated with a scaled single Wishart distribution by matching the first two moments. We adapt this in the panel setting. Let

$$\Phi = \mu_1^{-1}\sum_{k=1}^{\infty}\lambda_k\zeta_k\zeta'_k.$$

where

$$\mu_1 = \sum_{k=1}^{\infty}\lambda_k = \int_0^1\int_0^1\int_0^1 \mathbb{K}_b^*\left((r_1, r_2, \tau), (r_1, r_2, \tau)\right)dr_1dr_2d\tau.$$

We approximate the distribution of Φ by $\Psi \sim \mathbb{W}_g\left(I_g, K\right)/K$ for some integer $K > 0$. By the properties of the Wishart distribution, we have

$$E\Psi = I_g \tag{40}$$

$$E\Psi D\Psi = \frac{1}{K}tr\left(D\right)I_g + \left(1 + \frac{1}{K}\right)D \tag{41}$$

for any symmetric matrix D. By definition, $E\Phi = E\Psi$. Following Example 7.1 in Bilodeau and Brenner (1999), we can show that

$$E\Phi D\Phi = \frac{\mu_2}{\mu_1^2}tr\left(D\right)I_g + \left(1 + \frac{\mu_2}{\mu_1^2}\right)D \tag{42}$$

where

$$\mu_2 = \sum_{k=1}^{\infty} \lambda_k^2 = \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \left[\mathbb{K}_b^* \left((r_1, r_2, \tau), (v_1, v_2, \kappa) \right) \right]^2 dr_1 dr_2 d\tau dv_1 dv_2 d\kappa.$$

From (41) and (42), it is reasonable to approximate the distribution of Φ by $\mathbb{W}_g(I_g, K)/K$ with

$$K = \frac{\mu_1^2}{\mu_2}.$$

K is called equivalent degree of freedom (EDF) of $\hat{J}_{nT}$. Combining this results with (39), we have

$$\mu_1 g F_\infty(g, b) \overset{d}{\approx} \phi' \Psi^{-1} \phi.$$

By definition, $\phi' \Psi^{-1} \phi$ follows Hotelling's $T^2(g, K)$ distribution. As

$$\frac{K - g + 1}{gK} T^2(g, K) \sim F_{g, K-g+1},$$

we can use an F distribution as the reference distribution based on the following approximation:

$$\frac{\mu_1 (K - g + 1)}{K} F_\infty(g, b) \overset{d}{\approx} F_{g, K-g+1}. \tag{43}$$

Lemma 1 *As b_1, b_2 and $b_3 \to 0$, we have*

$$(a) \;\; \mu_1 = 1 - b_1 b_2 b_3 c_1 + o(b_1 b_2 b_3); \quad (b) \;\; \mu_2 = b_1 b_2 b_3 c_2 + o(b_1 b_2 b_3).$$

Proofs are given in the supplementary appendix.[7] Lemma 1 implies

$$\frac{\mu_1 (K - g + 1)}{K} = \frac{1}{1 + b_1 b_2 b_3 (c_1 + (g-1) c_2)} + o(b_1 b_2 b_3). \tag{44}$$

Hence, we can use $1/(1 + b_1 b_2 b_3 (c_1 + (g-1) c_2))$ as the correction factor. This is always between zero and one. The following theorem summarizes the F-approximation.

Theorem 6 *Suppose Assumptions F1 - F4 hold and $F_\infty^*(g, b)$ is defined by*

$$F_\infty^*(g, b) = F_\infty(g, b)/\nu$$

where

$$\nu = 1 + (c_1 + (g-1) c_2) b_1 b_2 b_3.$$

As b_1, b_2 and $b_3 \to 0$, we have

$$P\{F_\infty^*(g, b) \le z\} = P\{F_{g, K^*} \le z\} + o(b_1 b_2 b_3)$$

where $K^* = \max(5, \lceil 1/(b_1 b_2 b_3 c_2) \rceil)$ and $\lceil \cdot \rceil$ denotes the integer part.

[7] The supplementary appendix is available at the author's homepage.

Proofs are given in the appendix. Some comments on Theorem 6 are in order. We use K^* in place of $K-g+1$ for the second degree of freedom in the F-approximation. This modification ensures that the variance of the F distribution exists.

Let $F_\infty^\alpha(g,b)$ and F_{g,K^*}^α denote the $1-\alpha$ quantiles of the distribution of $F_\infty(g,b)$ and the F distribution with the degrees of freedom g and K^*. Theorem 6 suggests that for the F-test version of Wald statistic, F_{nT}, we use

$$\mathcal{F}_{g,b}^\alpha := \nu F_{g,K^*}^\alpha$$

as the critical value for the test with nominal size α.

As the EDF is proportional to $1/(b_1b_2b_3)$, we may want to choose the large EDF by increasing the degree of smoothing for more powerful inference. However, for a given sample size, small $b_1b_2b_3$ introduces large bias in $\hat{J}_{nT}$, which will cause F_{nT}/ν to deviate substantially from F_{g,K^*} distribution. At the same time, there can be a significant loss in power if we choose too large bandwidths.

8 Monte Carlo simulation

In this section, we provide some simulation evidence on the finite sample performance of our estimator and testing procedure under both the increasing smoothing asymptotics and fixed smoothing asymptotics. We choose the bandwidths based on the AMSE* criterion and consider the rectangular kernel as well as the Parzen kernel to construct $\hat{J}_{nT}$. We compare the performance of $\hat{J}_{nT}$ with $\hat{J}_{nT}^{DK}$, $\hat{J}_{nT}^{A}$ and $\hat{J}_{nT}^{KP}$. We evaluate the estimators using the RMSE criterion and the coverage accuracy of the associated confidence intervals (CIs). We examine the robustness to the measurement errors in economic distance. It is also investigated how the number of restrictions being tested affect the performance of the Wald tests under the two different smoothing asymptotics.

We assume a lattice structure, in which each individual is located on a square grid of integers over time and use the Euclidean distance for d_{ij}. The data generating processes we consider here are:

$$\begin{aligned}
\text{DGP1:}\quad & Y_{it} = \beta_0 + u_{it} \qquad \beta_0 = 0;\\
& u_t = \lambda u_{t-1} + \varepsilon_t, \quad \varepsilon_t = \theta(I-\tilde{W}_n)^{-1}v_t,\ v_t \overset{i.i.d.}{\sim} N(0, I_n);
\end{aligned}$$

$$\begin{aligned}
\text{DGP2:}\quad & Y_{it} = X_{it}^{(1)}\beta_{10} + \ldots + X_{it}^{(p)}\beta_{p0} + \alpha_i + f_t + u_{it},\\
& \beta_{10} = \ldots = \beta_{p0} = 0, \quad \alpha_i = f_t = 0;\\
& X_t = \lambda X_{t-1} + \nu_t, \qquad \nu_t = \theta(I-\tilde{W}_n)^{-1}\eta_t,\ \eta_t \overset{i.i.d.}{\sim} N(0, I_n);\\
& u_t = \lambda u_{t-1} + \varepsilon_t, \qquad \varepsilon_t = \theta(I-\tilde{W}_n)^{-1}v_t,\ v_t \overset{i.i.d.}{\sim} N(0, I_n),
\end{aligned}$$

where X_{it} is a p-vector. $X_t = (X_{1t}, \ldots, X_{nt})'$ and $u_t = (u_{1t}, \ldots, u_{nt})'$ are modeled as AR(1) processes with the autoregressive parameter λ to generate the temporal correlation of data. X_t and u_t are also spatially dependent as $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{nt})'$ and $\nu_t = (\nu_{1t}, \ldots, \nu_{nt})'$ follow spatial AR(1) processes with parameter θ. The values considered for the model parameters λ and θ are 0, 0.3, 0.6 and 0.9. $\tilde{W}_n$ is a contiguity matrix and individuals i and j are neighbors if $d_{ij} = 1$. Following convention, it is row-standardized and its diagonal elements are zero.

DGP1 is used for the RMSE criterion and the DGP2 is for the coverage accuracy of the associated CIs. DGP2 includes the individual and time effects and β_0 is estimated with the fixed effects estimator. In contrast, these effects are absent in DGP1 for easy calculation of the RMSE. We estimate β_0 in DGP1 with the simple average $(= (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} y_{it})$.

For bandwidth selection, the formulas in (30) and (31) are considered with the AR(1) and spatial AR(1) models for $\hat{J}_{nT}^{DK}$ and $\hat{J}_{nT}^{KP}$ respectively. For $\hat{J}_{nT}$ with the Parzen kernel, we choose the bandwidths based on (15) and (16) using the spatiotemporal parametric model in (20). We use the Parzen kernel as the target kernel to obtain the bandwidths for $\hat{J}_{nT}$ with the rectangular kernel. One concern of the rectangular kernel estimator is that it is not positive semi-definite. However, we can attain positive semi-definiteness with a simple modification suggested by Politis (2010). As $\hat{J}_{nT}$ is symmetry, $\hat{J}_{nT} = \hat{U}\hat{\Lambda}\hat{U}'$, 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 $\hat{J}_{nT}$. 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 $\hat{J}_{nT}^+ = \hat{U}\hat{\Lambda}^+\hat{U}'$. As each eigenvalue of $\hat{J}_{nT}$ is nonnegative, it is positive semi-definite.

W_n is the contiguity matrix in which individuals i and j are neighbors if $d_{ij} = 1$. We set $\eta = 2$ and $\ell_n = \pi d_n^2$. Note that the approximating parametric models for $\hat{J}_{nT}^{KP}$ and $\hat{J}_{nT}$ are mis-specified whereas the AR(1) model for $\hat{J}_{nT}^{DK}$ is correctly specified. We estimate parameters in (20) and (31) with the QMLE.

We consider three different sample sizes; (i) small T and n; $T = 15, n = 49$ (7×7), (ii) large T and small n; $T = 50, n = 49$, and (iii) small T and large n; $T = 15, n = 196$ (14×14). The Parzen kernel, K_{PA}, and the rectangular kernel, K_{RE}, are defined as:

$$K_{PA}(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & \text{for } |x| \leq 1/2 \\ 2(1 - |x|)^3, & \text{for } 1/2 < |x| \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad K_{RE}(x) = \begin{cases} 1, & \text{for } |x| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

respectively. The following values are used for each kernel.

	$\bar{\mathcal{K}}_1$	$\bar{\mathcal{K}}_2$	c_1	c_2	K_q
Parzen	0.2889	0.5393	0.4123	0.1558	-6
Rectangular	1	2	6.2926	6.2926	

We also allow for the case with measurement errors in the distance measure. The error contaminated distance, d_{ij}^* is generated as follows. If $d_{ij} < 2$, then d_{ij} is observed without a measurement error. But if $d_{ij} \geq 2$, then we observe d_{ij}^* which is generated from the following process:

$$d_{ij}^* = d_{ij} + e_{ij}, \qquad e_{ij} = \begin{cases} -1 & w.p.\ 1/3 \\ 0 & w.p.\ 1/3 \\ 1 & w.p.\ 1/3. \end{cases}$$

PHAC, CCE, DK and KP denote the test statistics based on $\hat{J}_{nT}$, $\hat{J}_{nT}^A$, $\hat{J}_{nT}^{DK}$, and $\hat{J}_{nT}^{KP}$ respectively. We use the F-approximation for fixed smoothing asymptotics.

Table 1 presents the ratio of the RMSEs to J_{nT} for $\hat{J}_{nT}$ and $\hat{J}_{nT}^{DK}$ evaluated at the data dependent bandwidth estimators and at infeasible optimal bandwidth parameters. First, we find that $\hat{J}_{nT}$ performs better than $\hat{J}_{nT}^{DK}$ in almost all the cases. When spatial dependence is absent or weak, $\hat{J}_{nT}$

has substantially smaller ratio than $\hat{J}_{nT}^{DK}$. This is because the DK estimator loses information in the spatial dimension with the cross-sectional averaging. When $\theta = 0.9$, both the estimators are not much different. Especially $\hat{J}_{nT}$ with the rectangular kernel is as accurate as and sometimes better than $\hat{J}_{nT}^{DK}$. This implies that adaptiveness works quite well in this setting. Second, the increase of n reduces only the ratio of $\hat{J}_{nT}$ while increasing T improves the performance of both the estimators. Finally, the AMSE* criterion tends to effectively controls the RMSE of $\hat{J}_{nT}$.

Table 2 reports the empirical coverage probabilities (ECPs) of 95% CIs associated with the different covariance estimators; $\hat{J}_{nT}$, $\hat{J}_{nT}^{DK}$, $\hat{J}_{nT}^{A}$, and $\hat{J}_{nT}^{KP}$. The null hypothesis we consider is

$$H_0 : \ \beta_1 = 0.$$

DGP2 is used with a univariate regressor ($p = 1$). For the testing with $\hat{J}_{nT}$, we use both the fixed smoothing asymptotics and increasing smoothing asymptotics. The simulation results verify our theoretical results. First, we compare $\hat{J}_{nT}$ with the other estimators under increasing smoothing asymptotics. PHAC performs as well as CCE when $\theta = 0$ and is significantly more accurate when there exists spatial dependence. Compared with the KP, the CIs associated with PHAC is more precise unless the process is temporally highly persistent. Even under strong temporal persistence ($\lambda = 0.9$) PHAC is almost as good as KP especially if n is small. Both the PHAC and KP become more accurate with the size of n, but only the performance of PHAC improves with the size of T. Comparison with DK is quite similar to the RMSE criterion case. The PHAC tends to perform better than DK except in some cases with $\theta = 0.9$. Even when $\theta = 0.9$, the ECP of PHAC especially with the rectangular kernel is very close to that of DK.

Second, Table 2 compares the performance of PHAC using increasing smoothing asymptotics and fixed smoothing asymptotics. The results indicate that fixed smoothing asymptotic approximation is substantially more accurate than increasing smoothing asymptotics. The difference increases as the process becomes more persistent. When $\theta = 0.9$, $\lambda = 0.9$ and $T = 15, n = 49$, the ECP of the PHAC with the Parzen kernel under fixed smoothing asymptotics is 79.6% but it is 64.4% under increasing smoothing asymptotics. This result shows that fixed smoothing asymptotics captures the demeaning bias arising from the estimation errors of $\hat{\beta}$ while it is ignored under increasing smoothing asymptotics.

Third, Table 2 provides a strong evidence that the rectangular kernel performs better than the finite order kernel under fixed smoothing asymtoptics. Especially in our simulation, PHAC with the rectangular kernel is very robust to spatial dependence so that size distortion does not increase with the spatial dependence. This size advantage of the rectangular kernel may arise from its bias reducing property and adaptiveness.

Finally, Table 2 shows that our testing procedure based on fixed smoothing asymptotics is reasonably robust to measurement errors. As economic distance data is likely to be error contaminated in practice, the robustness to measurement errors is a highly desirable property. Comparing PHAC with PHAC_e, we see that the performance of PHAC_e is quite close to that of PHAC in most cases.

Table 3 compares the two different asymptotics with the different number of restrictions being tested. The DGP2 is used with $p = 3$. We consider a single restriction ($g = 1$) and joint restrictions ($g = 3$)

$$H_0 : \ \beta_1 = 0, \quad H_0 : \ \beta_1 = \beta_2 = \beta_3 = 0$$

respectively. The table evidently indicates that under increasing smoothing asymptotics the size distortion increases with the number of restrictions being tested. This is especially severe when the process is highly persistent. When $g = 3$ and $\theta = \lambda = 0.9$, the ECP of PHAC with the Parzen kernel is only 27.1% under increasing smoothing asymptotics. The size distortion of PHAC also increases under fixed smoothing asymptotics with the number of restrictions being tested but much lesser. This is consistent with our asymptotic expansion in Theorem 5. The theorem shows that fixed smoothing asymptotics allows for the number of restrictions being tested and the critical value of PHAC_F is the increasing function of g.

9 Conclusion

In this paper we study robust inference for linear panel models with fixed effects in the presence of heteroskedasticity and spatiotemporal dependence of unknown forms. We consider a bivariate kernel covariance matrix estimator and examine the properties of the covariance estimator and the associated test statistics under both the increasing smoothing asymptotics and fixed smoothing asymptotics. We also derive the optimal selection procedures based on the upper bound of AMSE. For the fixed smoothing asymptotic distribution, we establish the validity of F-approximation. The adaptiveness of our estimator enables our method to be safely used without the knowledge of dependence structure.

Instead of using the upper bound of asymptotic MSE criterion, we can study the optimal bandwidth selection based on a criterion which is most suitable for hypothesis testing and CI construction. It is interesting to extend the bandwidth selection methods by Sun, Phillips and Jin (2008), Sun (2010) and Sun and Kaplan (2010) on time series HAC estimation to the panel setting.

Table 1: RMSE/Estimand with $\hat{J}_{nT}$ and $\hat{J}_{nT}^{DK}$ - DGP1

λ		θ								
		0.0	0.3	0.6	0.9		0.0	0.3	0.6	0.9
					T=15, n=49					
0.0		0.09	0.20	0.27	0.46		0.42	0.21	0.25	0.43
0.3	$\hat{J}_{nT}$	0.16	0.34	0.45	0.65	$\hat{J}_{nT}$	0.13	0.33	0.43	0.60
0.6	$(\hat{d}_n, \hat{d}_T)_{PA}$	0.22	0.41	0.55	0.72	$(d_n^\star, d_T^\star)_{PA}$	0.17	0.40	0.53	0.71
0.9		0.35	0.53	0.67	0.84		0.31	0.52	0.65	0.83
0.0		0.13	0.23	0.29	0.38		1.00	0.36	0.38	0.36
0.3	$\hat{J}_{nT}$	0.21	0.36	0.51	0.62	$\hat{J}_{nT}$	0.19	0.31	0.41	0.50
0.6	$(\hat{d}_n, \hat{d}_T)_{RE}$	0.29	0.47	0.62	0.68	$(d_n^\star, d_T^\star)_{RE}$	0.20	0.42	0.49	0.64
0.9		0.38	0.56	0.70	0.83		0.19	0.48	0.63	0.79
0.0		0.48	0.46	0.48	0.47		0.36	0.36	0.38	0.36
0.3	$\hat{J}_{nT}^{DK}$	0.54	0.56	0.57	0.54	$\hat{J}_{nT}^{DK}$	0.56	0.56	0.58	0.54
0.6	$(\hat{d}_T^{DK})$	0.68	0.70	0.69	0.70	(d_T^{DK})	0.70	0.71	0.71	0.72
0.9		0.89	0.88	0.88	0.88		0.89	0.89	0.88	0.88
					T=50, n=49					
0.0		0.05	0.13	0.18	0.40		0.42	0.12	0.17	0.39
0.3	$\hat{J}_{nT}$	0.10	0.24	0.34	0.55	$\hat{J}_{nT}$	0.14	0.24	0.33	0.50
0.6	$(\hat{d}_n, \hat{d}_T)_{PA}$	0.14	0.33	0.50	0.64	$(d_n^\star, d_T^\star)_{PA}$	0.13	0.31	0.42	0.58
0.9		0.26	0.48	0.60	0.83		0.21	0.43	0.57	0.76
0.0		0.08	0.14	0.18	0.21		1.00	0.20	0.19	0.20
0.3	$\hat{J}_{nT}$	0.13	0.26	0.37	0.57	$\hat{J}_{nT}$	0.21	0.22	0.29	0.32
0.6	$(\hat{d}_n, \hat{d}_T)_{RE}$	0.19	0.41	0.67	0.58	$(d_n^\star, d_T^\star)_{RE}$	0.20	0.28	0.36	0.46
0.9		0.34	0.56	0.68	0.81		0.20	0.40	0.54	0.70
0.0		0.28	0.29	0.27	0.28		0.21	0.20	0.19	0.20
0.3	$\hat{J}_{nT}^{DK}$	0.40	0.41	0.40	0.40	$\hat{J}_{nT}^{DK}$	0.38	0.38	0.38	0.37
0.6	$(\hat{d}_T^{DK})$	0.53	0.54	0.55	0.56	(d_T^{DK})	0.52	0.52	0.53	0.52
0.9		0.77	0.76	0.77	0.78		0.77	0.76	0.77	0.78
					T=15, n=196					
0.0		0.05	0.13	0.18	0.29		0.43	0.20	0.21	0.27
0.3	$\hat{J}_{nT}$	0.09	0.24	0.33	0.54	$\hat{J}_{nT}$	0.07	0.24	0.32	0.47
0.6	$(\hat{d}_n, \hat{d}_T)_{PA}$	0.13	0.30	0.42	0.57	$(d_n^\star, d_T^\star)_{PA}$	0.12	0.29	0.39	0.56
0.9		0.29	0.43	0.52	0.72		0.28	0.43	0.51	0.69
0.0		0.07	0.15	0.21	0.30		1.00	0.34	0.36	0.35
0.3	$\hat{J}_{nT}$	0.11	0.27	0.37	0.62	$\hat{J}_{nT}$	0.09	0.23	0.28	0.41
0.6	$(\hat{d}_n, \hat{d}_T)_{RE}$	0.15	0.36	0.51	0.62	$(d_n^\star, d_T^\star)_{RE}$	0.10	0.26	0.37	0.50
0.9		0.22	0.43	0.55	0.74		0.10	0.35	0.45	0.66
0.0		0.47	0.43	0.48	0.47		0.37	0.34	0.36	0.35
0.3	$\hat{J}_{nT}^{DK}$	0.53	0.56	0.55	0.55	$\hat{J}_{nT}^{DK}$	0.54	0.56	0.56	0.55
0.6	$(\hat{d}_T^{DK})$	0.68	0.70	0.69	0.69	(d_T^{DK})	0.70	0.72	0.71	0.70
0.9		0.88	0.87	0.88	0.88		0.89	0.88	0.89	0.89

The subscripts 'PA' and 'RE' denote the Parzen and rectangular kernels respectively.

Table 2: Empirical Coverage Probabilities of Nominal 95% CIs Constructed
Using Alternative Covariance Estiamtors - DGP2

λ		θ													
		0.0	0.3	0.6	0.9		0.0	0.3	0.6	0.9		0.0	0.3	0.6	0.9
T=15, n=49															
0.0	PHAC (PA,F)	93.9	94.2	91.8	88.0	DK	89.4	89.1	88.6	90.5	PHAC (RE,F)	94.2	94.6	93.3	94.6
0.3		91.4	90.3	90.9	83.7		87.0	83.7	88.3	86.1		91.5	90.4	92.6	94.3
0.6		87.5	88.2	85.1	79.2		77.4	79.0	76.1	77.0		88.6	89.4	88.9	87.8
0.9		87.5	84.2	83.4	79.6		64.8	64.1	62.2	62.9		86.1	84.7	87.2	79.3
0.0	$PHAC_e$ (PA,F)	93.7	94.2	91.3	87.9	CCE	94.9	93.6	86.6	56.6	$PHAC_e$ (RE,F)	93.6	94.3	92.8	92.0
0.3		91.0	90.0	89.8	83.1		93.0	91.3	86.6	54.2		91.0	89.9	91.4	88.6
0.6		86.7	87.1	82.7	77.2		92.9	91.9	83.9	53.7		87.6	88.1	87.0	85.4
0.9		85.9	82.6	79.5	74.8		92.8	90.1	83.6	55.1		85.4	84.1	85.1	82.9
0.0	PHAC (PA,I)	93.7	94.0	91.5	87.5	KP	91.3	90.2	86.9	67.1	PHAC (RE,I)	93.7	94.0	91.6	90.3
0.3		91.0	90.0	90.3	82.3		89.5	88.0	86.2	64.2		90.6	89.4	90.9	85.1
0.6		86.8	87.6	82.6	71.8		88.1	88.9	84.4	62.1		86.7	87.9	83.0	73.1
0.9		86.1	83.0	77.6	64.4		90.0	86.7	83.1	65.2		83.9	81.6	75.5	58.0
T=50, n=49															
0.0	PHAC (PA,F)	94.7	92.7	91.5	88.0	DK	92.6	93.1	92.7	93.9	PHAC (RE,F)	94.8	93.4	92.9	95.5
0.3		92.9	93.3	89.6	83.9		90.5	92.1	90.1	90.1		92.9	93.8	91.1	93.7
0.6		93.1	91.5	90.2	84.4		87.6	87.4	88.5	87.3		93.6	92.6	92.6	95.4
0.9		88.3	87.4	88.3	75.5		69.4	70.1	71.6	69.7		88.1	88.7	90.7	85.4
0.0	$PHAC_e$ (PA,F)	94.8	92.8	91.0	87.8	CCE	93.9	91.6	83.7	55.0	$PHAC_e$ (RE,F)	95.0	93.4	92.2	93.6
0.3		92.5	92.9	88.6	83.8		93.5	92.3	81.9	54.1		92.4	93.2	89.4	89.0
0.6		93.0	91.0	88.8	83.8		94.3	91.9	85.9	55.5		93.2	92.5	91.7	92.9
0.9		87.1	85.1	84.3	72.2		93.5	91.5	84.7	53.9		88.3	88.2	88.9	84.9
0.0	PHAC (PA,I)	94.6	92.7	91.3	88.7	KP	90.2	88.1	83.3	66.7	PHAC (RE,I)	94.7	93.0	92.2	94.4
0.3		92.4	93.1	88.8	84.8		88.5	89.6	82.7	65.3		92.7	93.2	89.8	89.6
0.6		93.1	91.2	88.6	82.5		90.5	88.5	85.1	65.5		93.4	91.8	89.6	85.0
0.9		87.1	85.4	81.3	67.4		88.9	87.6	85.6	63.5		86.4	85.0	79.5	67.2
T=15, n=196															
0.0	PHAC (PA,F)	93.6	92.4	93.2	90.8	DK	86.1	87.7	88.8	89.4	PHAC (RE,F)	93.6	93.0	94.2	92.9
0.3		92.1	92.6	92.0	89.4		85.0	86.1	85.0	87.3		92.2	91.2	91.1	92.3
0.6		91.0	89.9	88.2	88.2		80.1	82.7	80.1	75.1		89.9	92.1	90.2	90.7
0.9		88.6	90.7	86.9	89.2		62.9	65.5	64.4	61.5		85.8	89.0	88.0	82.5
0.0	$PHAC_e$ (PA,F)	93.5	92.3	92.7	89.0	CCE	94.5	92.7	84.9	50.3	$PHAC_e$ (RE,F)	93.4	92.8	94.0	91.9
0.3		92.1	91.6	89.8	88.0		94.7	92.2	85.0	50.3		91.4	90.4	90.9	90.1
0.6		90.4	88.7	86.9	83.7		94.4	94.2	86.5	47.7		89.8	91.5	89.3	89.5
0.9		88.4	88.9	83.3	82.5		93.1	93.0	85.3	47.8		84.5	88.8	86.7	87.0
0.0	PHAC (PA,I)	93.6	92.3	92.7	88.9	KP	93.3	92.9	90.0	79.4	PHAC (RE,I)	93.4	92.8	93.3	90.7
0.3		92.3	90.8	89.8	86.5		93.9	92.1	90.0	78.8		92.1	90.9	90.3	87.0
0.6		89.9	91.0	87.9	76.5		93.6	94.2	89.6	74.6		89.4	91.3	88.0	76.5
0.9		86.1	88.9	83.2	71.6		92.3	93.3	89.8	76.3		85.0	87.4	81.2	62.5

'PA' and 'RE' denote the Parzen and rectangular kernels respectively.
'F' and 'I' denote fixed smoothing and increasing smoothing respectively.
The superscript 'e' denotes measurement errors.

Table 3: Empirical Coverage Probabilities of Nominal 95% CIs
Constructed with Different Number of Restrictions - DGP2

	λ	g=1				g=3			
		θ				θ			
		0.0	0.3	0.6	0.9	0.0	0.3	0.6	0.9
PHAC (PA,F)	0.0	93.3	91.9	92.2	85.7	92.6	91.2	88.1	77.8
	0.3	92.3	91.0	90.2	82.6	91.9	88.0	83.9	72.2
	0.6	89.8	88.1	85.3	78.6	82.4	82.1	78.4	69.0
	0.9	86.9	84.1	81.4	80.2	81.2	77.1	73.3	68.7
PHAC (PA,I)	0.0	93.1	91.5	91.5	84.8	92.4	90.3	85.9	72.2
	0.3	92.1	90.6	89.5	80.9	91.4	86.9	81.5	65.5
	0.6	89.0	87.1	83.2	70.6	80.7	79.7	70.6	46.8
	0.9	84.7	82.5	75.8	62.5	77.4	70.8	55.4	27.1
PHAC (RE,F)	0.0	93.7	92.2	93.3	92.7	93.0	92.5	91.4	93.0
	0.3	92.4	90.8	92.2	91.7	91.6	89.2	88.9	92.7
	0.6	89.7	89.6	89.6	88.2	84.9	85.1	87.9	87.8
	0.9	85.8	85.1	85.8	83.0	82.0	78.5	83.4	83.8
PHAC (RE,I)	0.0	93.5	91.7	92.5	86.9	92.0	91.1	87.1	76.5
	0.3	91.7	90.5	89.7	83.8	89.9	86.6	82.2	68.1
	0.6	88.7	88.3	84.9	72.6	80.1	79.5	72.3	47.9
	0.9	82.5	81.6	74.2	62.5	71.2	66.2	49.2	44.1

See note to Table 2.

APPENDIX

Proof of Theorem 1

For notational simplicity, we re-order the individuals and time and make new indices. For $i_{(j)} = 1, ..., \ell_{j,n}$, $d_{i_{(j)}j} \leq d_n$, and for $i_{(j)} = \ell_{j+1,n}, \ldots, n$, $d_{i_{(j)}j} > d_n$. For $t_{(s)} = 1, ..., \ell_{s,T}$, $d_{t_{(s)}s} \leq d_T$, and for $t_{(s)} = \ell_{s,T}+1, \ldots, T$, $d_{t_{(s)}s} > d_T$.

(a) Asymptotic Variance

Let $\varphi_{lkcd} = \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} r_{(i,t),l}^{(c)} r_{(j,s),k}^{(d)} K\left(\frac{d_{ij}}{d_n}\right) K\left(\frac{d_{ts}}{d_T}\right)$. We have

$$
\begin{aligned}
&\frac{nT}{\ell_n \ell_T} cov\left(\tilde{J}_{nT}(c_1, d_1), \tilde{J}_{nT}(c_2, d_2)\right) \\
&= \frac{1}{nT\ell_n\ell_T} E\left[\sum_{l,k=1}^{nTp} \varphi_{lkc_1d_1}(\varepsilon_l\varepsilon_k - E\varepsilon_l\varepsilon_k) \sum_{e,f=1}^{nTp} \varphi_{efc_2d_2}(\varepsilon_e\varepsilon_f - E\varepsilon_e\varepsilon_f)\right] \\
&= \frac{1}{nT\ell_n\ell_T} E\left[\sum_{l,k,e,f=1}^{nTp} \varphi_{lkc_1d_1}\varphi_{efc_2d_2}(\varepsilon_l\varepsilon_k\varepsilon_e\varepsilon_f - \varepsilon_l\varepsilon_k E\varepsilon_e\varepsilon_f - \varepsilon_e\varepsilon_f E\varepsilon_l\varepsilon_k + E\varepsilon_l\varepsilon_k E\varepsilon_e\varepsilon_f)\right] \\
&= \frac{1}{nT\ell_n\ell_T} E\left[\sum_{l=1}^{nTp} \varphi_{llc_1d_1}\varphi_{llc_2d_2}\left(E\varepsilon_l^4 - 3\right) + \sum_{l,k=1}^{nTp} \varphi_{lkc_1d_1}\varphi_{lkc_2d_2} + \sum_{l,k=1}^{nTp} \varphi_{lkc_1d_1}\varphi_{klc_2d_2}\right] \\
&:= C_{1nT} + C_{2nT} + C_{3nT},
\end{aligned}
$$

where

$$
C_{1nT} = \frac{1}{nT\ell_n\ell_T} \sum_{l=1}^{nTp} \left(E\varepsilon_l^4 - 3\right) \sum_{i,j,a,b=1}^{n} \sum_{t,s,u,v=1}^{T} K\left(\frac{d_{ij}}{d_n}\right) K\left(\frac{d_{ts}}{d_T}\right) K\left(\frac{d_{ab}}{d_n}\right) K\left(\frac{d_{uv}}{d_T}\right) r_{(i,t),l}^{(c_1)} r_{(j,s),l}^{(d_1)} r_{(a,u),l}^{(c_2)} r_{(b,v),l}^{(d_2)},
$$

$$
C_{2nT} = \frac{1}{nT\ell_n\ell_T} \sum_{l,k=1}^{nTp} \sum_{i,j,a,b=1}^{n} \sum_{t,s.u,v=1}^{T} K\left(\frac{d_{ij}}{d_n}\right) K\left(\frac{d_{ts}}{d_T}\right) K\left(\frac{d_{ab}}{d_n}\right) K\left(\frac{d_{uv}}{d_T}\right) r_{(i,t),l}^{(c_1)} r_{(j,s),k}^{(d_1)} r_{(a,u),l}^{(c_2)} r_{(b,v),k}^{(d_2)},
$$

$$
C_{3nT} = \frac{1}{nT\ell_n\ell_T} \sum_{l,k=1}^{nTp} \sum_{i,j,a,b=1}^{n} \sum_{t,s,u,v=1}^{T} K\left(\frac{d_{ij}}{d_n}\right) K\left(\frac{d_{ts}}{d_T}\right) K\left(\frac{d_{ab}}{d_n}\right) K\left(\frac{d_{uv}}{d_T}\right) r_{(i,t),l}^{(c_1)} r_{(j,s),k}^{(d_1)} r_{(a,u),k}^{(c_2)} r_{(b,v),l}^{(d_2)}.
$$

For C_{1nT},

$$
\begin{aligned}
|C_{1nT}| &\leq \frac{1}{nT\ell_n\ell_T} \sum_{l=1}^{nTp} \left|E\varepsilon_l^4 - 3\right| \sum_{i,j,a,b=1}^{n} \sum_{t,s,u,v=1}^{T} K\left(\frac{d_{ij}}{d_n}\right) K\left(\frac{d_{ab}}{d_n}\right) K\left(\frac{d_{ts}}{d_T}\right) K\left(\frac{d_{uv}}{d_T}\right) \\
&\quad \times \left| r_{(i,t),l}^{(c_1)} r_{(j,s),l}^{(d_1)} r_{(a,u),l}^{(c_2)} r_{(b,v),l}^{(d_2)} \right| \\
&\leq \frac{c_R^4}{\ell_n\ell_T} \frac{1}{nT} \sum_{l=1}^{nTp} \left|E\varepsilon_l^4 - 3\right| \leq \frac{c_R^4 c_E p}{\ell_n\ell_T} = o(1) \qquad \text{(A.1)}
\end{aligned}
$$

using Assumptions I1 and I2.

For C_{2nT}, we can decompose as follows to consider boundary effects:

$$
\begin{aligned}
C_{2nT} &= \frac{1}{nT\ell_n\ell_T}\sum_{i,a=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u=1}^{T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\\
&\times \gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\\
&= \frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\\
&\times \gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\\
&+ \frac{1}{nT\ell_n\ell_T}\sum_{i,a=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t\notin E_T}\sum_{u=1}^{T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\\
&\times \gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\\
&+ \frac{1}{nT\ell_n\ell_T}\sum_{i,a=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t\in E_T}\sum_{u\notin E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\\
&\times \gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\\
&+ \frac{1}{nT\ell_n\ell_T}\sum_{i\notin E_n}\sum_{a=1}^{n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\\
&\times \gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\\
&+ \frac{1}{nT\ell_n\ell_T}\sum_{i\in E_n}\sum_{a\notin E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\\
&\times \gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\\
&= D_{1nT}+D_{2nT}+D_{3nT}+D_{4nT}+D_{5,nT} \qquad \text{(A.2)}
\end{aligned}
$$

D_{1nT} is based on nonboundary units whereas the others are on boundary ones. In the following, we show that D_{1nT} converges to $\bar{\mathcal{K}}_1\bar{\mathcal{K}}_2 J(c_1,c_2)\, J(d_1,d_2)$ and the other terms become negligible as n and T increase.

For D_{1nT}, the first step is to show that

$$
\begin{aligned}
&\lim_{n,T\to\infty}\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}} K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right)K^2\left(\frac{d_{ts_{(t)}}}{d_T}\right)\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\\
&= \bar{\mathcal{K}}_1\bar{\mathcal{K}}_2 J(c_1,c_2)\, J(d_1,d_2), \qquad \text{(A.3)}
\end{aligned}
$$

and the next step is to prove that

$$
\begin{aligned}
&\lim_{n,T\to\infty} D_{1nT}\\
&= \lim_{n,T\to\infty}\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}} K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right)K^2\left(\frac{d_{ts_{(t)}}}{d_T}\right)\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}
\end{aligned}
\qquad \text{(A.4)}
$$

For (A.3), let $\gamma^{(d_1 d_2)}_{(\bar{i}\bar{t},b_{(a)}v_{(u)})} = (\ell_n \ell_T)^{-1} \sum_{h_{(i)}=1}^{\ell_{i,n}} \sum_{w_{(t)}=1}^{\ell_{t,T}} \gamma^{(d_1 d_2)}_{(h_{(i)}w_{(t)},b_{(a)}v_{(u)})}$. Then,

$$
\begin{aligned}
&\frac{1}{nT\ell_n\ell_T} \sum_{i,a\in E_n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t,u\in E_T} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right) K^2\left(\frac{d_{ts_{(t)}}}{d_T}\right) \gamma^{(c_1c_2)}_{(it,au)} \gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})} \\
&= \frac{1}{nT\ell_n\ell_T} \sum_{i,a\in E_n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t,u\in E_T} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right) K^2\left(\frac{d_{ts_{(t)}}}{d_T}\right) \gamma^{(c_1c_2)}_{(it,au)} \gamma^{(d_1d_2)}_{(\bar{i}\bar{t},b_{(a)}v_{(u)})} \\
&+ \frac{1}{nT\ell_n\ell_T} \sum_{i,a\in E_n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t,u\in E_T} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right) K^2\left(\frac{d_{ts_{(t)}}}{d_T}\right) \gamma^{(c_1c_2)}_{(it,au)} \\
&\times \left(\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})} - \gamma^{(d_1d_2)}_{(\bar{i}\bar{t},b_{(a)}v_{(u)})}\right) \\
&= L_{1nT} + L_{2nT}. \qquad \text{(A.5)}
\end{aligned}
$$

L_{1nT} is rewritten as

$$
\begin{aligned}
&\frac{1}{nT} \sum_{i,a\in E_n} \sum_{t,u\in E_T} \gamma^{(c_1c_2)}_{(it,au)} \left(\frac{1}{\ell_n\ell_T} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} \gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right) \left(\frac{1}{\ell_n}\sum_{j=1}^{n} K^2\left(\frac{d_{ij}}{d_n}\right)\right) \left(\frac{1}{\ell_T}\sum_{s=1}^{T} K^2\left(\frac{d_{ts}}{d_T}\right)\right) \\
&= \frac{1}{nT} \sum_{i,a\in E_n} \sum_{t,u\in E_T} \gamma^{(c_1c_2)}_{(it,au)} \frac{1}{\ell_n\ell_T} cov\left(\sum_{j:d_{ij}\le d_n} \sum_{s:d_{ts}\le d_T} V^{(d_1)}_{(j,s)}, \sum_{b:d_{ab}\le d_n} \sum_{v:d_{uv}\le d_T} V^{(d_2)}_{(b,v)}\right) \left(\frac{1}{\ell_n}\sum_{j=1}^{n} K^2\left(\frac{d_{ij}}{d_n}\right)\right) \\
&\times \left(\frac{1}{\ell_T}\sum_{s=1}^{T} K^2\left(\frac{d_{ts}}{d_T}\right)\right) \\
&:= G_{1nT} + G_{2nT}
\end{aligned}
$$

where

$$
\begin{aligned}
G_{1nT} = &\frac{1}{nT} \sum_{i,a\in E_n} \sum_{t,u\in E_T} \gamma^{(c_1c_2)}_{(it,au)} 1\{d_{ia}\le c_n, d_{ut}\le c_T\} \left[\frac{1}{\ell_n\ell_T} cov\left(\sum_{j:d_{ij}\le d_n} \sum_{s:d_{ts}\le d_T} V^{(d_1)}_{(j,s)}, \sum_{b:d_{ab}\le d_n} \sum_{v:d_{uv}\le d_T} V^{(d_2)}_{(b,v)}\right)\right] \\
&\times \left(\frac{1}{\ell_n}\sum_{j=1}^{n} K^2\left(\frac{d_{ij}}{d_n}\right)\right) \left(\frac{1}{\ell_T}\sum_{s=1}^{T} K^2\left(\frac{d_{ts}}{d_T}\right)\right)
\end{aligned}
$$

and

$$
\begin{aligned}
G_{2nT} = &\frac{1}{nT} \sum_{i,a\in E_n} \sum_{t,u\in E_T} \gamma^{(c_1c_2)}_{(it,au)} \left[1\{d_{ia}>c_n, d_{ut}>c_T\} + 1\{d_{ia}>c_n, d_{ut}\le c_T\} + 1\{d_{ia}\le c_n, d_{ut}>c_T\}\right] \\
&\times \frac{1}{\ell_n\ell_n} cov\left(\sum_{j:d_{ij}\le d_n} \sum_{s:d_{ts}\le d_T} V^{(d_1)}_{(j,s)}, \sum_{b:d_{ab}\le d_n} \sum_{v:d_{uv}\le d_T} V^{(d_2)}_{(b,v)}\right) \left(\frac{1}{\ell_n}\sum_{j=1}^{n} K^2\left(\frac{d_{ij}}{d_n}\right)\right) \left(\frac{1}{\ell_T}\sum_{s=1}^{T} K^2\left(\frac{d_{ts}}{d_T}\right)\right) \\
&= o(1)
\end{aligned}
$$

as $c_n, c_T \to \infty$.

It suffices to consider G_{1nT}. When $d_{ia} \leq c_n$ and $d_{tu} \leq c_T$, we have

$$\begin{aligned}
& cov\left(\sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T} V_{(j,s)}^{(d_1)}, \sum_{b:d_{ab}\leq d_n}\sum_{v:d_{uv}\leq d_T} V_{(b,v)}^{(d_2)}\right) \\
&= cov\left(\sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T} V_{(j,s)}^{(d_1)}, \sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T} V_{(j,s)}^{(d_2)}\right) \\
&+ cov\left(\sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T} V_{(j,s)}^{(d_1)}, \sum_{b:d_{ab}\leq d_n}\sum_{v:d_{uv}\leq d_T} V_{(b,v)}^{(d_2)} - \sum_{b:d_{ib}\leq d_n}\sum_{v:d_{tv}\leq d_T} V_{(b,v)}^{(d_2)}\right)
\end{aligned}$$

but

$$\begin{aligned}
& \sum_{b:d_{ab}\leq d_n}\sum_{v:d_{uv}\leq d_T} V_{(b,v)}^{(d_2)} - \sum_{b:d_{ib}\leq d_n}\sum_{v:d_{tv}\leq d_T} V_{(b,v)}^{(d_2)} \\
&= \sum_{b:d_{ab}\leq d_n}\sum_{v:d_{uv}\leq d_T} V_{(b,v)}^{(d_2)} - \sum_{b:d_{ab}\leq d_n}\sum_{v:d_{tv}\leq d_T} V_{(b,v)}^{(d_2)} + \sum_{b:d_{ab}\leq d_n}\sum_{v:d_{tv}\leq d_T} V_{(b,v)}^{(d_2)} - \sum_{b:d_{ib}\leq d_n}\sum_{v:d_{tv}\leq d_T} V_{(b,v)}^{(d_2)} \\
&= \sum_{b:d_{ab}\leq d_n}\sum_{v:d_{uv}\leq d_T, d_{tv}>d_T} V_{(b,v)}^{(d_2)} + \sum_{b:d_{ab}\leq d_n, d_{ib}>d_n}\sum_{v:d_{tv}\leq d_T} V_{(b,v)}^{(d_2)}
\end{aligned}$$

Now $d_{ab} \leq d_n$ and $d_{ia} \leq c_n$ implies that $d_{bi} \leq d_n + c_n$. As the result,

$$\begin{aligned}
& \frac{1}{\ell_n \ell_T}\left| cov\left(\sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T} V_{(j,s)}^{(d_1)}, \sum_{b:d_{ab}\leq d_n, d_{ib}>d_n}\sum_{v:d_{tv}\leq d_T} V_{(b,v)}^{(d_2)}\right)\right| \\
&\leq \frac{1}{\ell_n \ell_T}\sum_{j:d_{ij}\leq d_n}\sum_{b:d_n<d_{ib}\leq d_n+c_n}\sum_{s:d_{ts}\leq d_T}\sum_{v:d_{tv}\leq d_T}\left| EV_{(j,s)}^{(d_1)} V_{(b,v)}^{(d_2)}\right| \\
&= o(1),
\end{aligned}$$

by choosing c_n such that $\sum_{b=1}^{n} 1\{d_{ib} \leq d_n + c_n\} \leq \bar{C}\ell_n$ for all i and for some constant $\bar{C}$. Similarly,

$$\frac{1}{\ell_n \ell_T}\left| cov\left(\sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T} V_{(j,s)}^{(d_1)}, \sum_{b:d_{ab}\leq d_n}\sum_{v:d_{uv}\leq d_T, d_{tv}>d_T} V_{(b,v)}^{(d_2)}\right)\right| = o(1).$$

Hence

$$\begin{aligned}
& \frac{1}{\ell_n \ell_T} cov\left(\sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T} V_{(j,s)}^{(d_1)}, \sum_{b:d_{ab}\leq d_n}\sum_{v:d_{uv}\leq d_T} V_{(b,v)}^{(d_2)}\right) \\
&= \frac{1}{\ell_n \ell_T} cov\left(\sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T} V_{(j,s)}^{(d_1)}, \sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T} V_{(j,s)}^{(d_2)}\right) + o\left(1\right) \qquad \text{(A.6)}
\end{aligned}$$

where $o(1)$ term holds uniformly over i and t.

Now under Assumption I6, we have

$$
\begin{aligned}
G_{1nT} &= \frac{1}{nT}\sum_{i,a\in E_n}\sum_{t,u\in E_T}\gamma_{(it,au)}^{(c_1c_2)}\frac{1}{\ell_n\ell_T}cov\left(\sum_{j:d_{ij}\le d_n}\sum_{s:d_{ts}\le d_T}V_{(j,s)}^{(d_1)},\sum_{j:d_{ij}\le d_n}\sum_{s:d_{ts}\le d_T}V_{(j,s)}^{(d_2)}\right)\\
&\quad\times\left(\frac{1}{\ell_n}\sum_{j=1}^{n}K^2\left(\frac{d_{ij}}{d_n}\right)\right)\left(\frac{1}{\ell_T}\sum_{s_{(t)}=1}^{T}K^2\left(\frac{d_{ts}}{d_T}\right)\right)(1+o(1))\\
&\quad\to\bar{\mathcal{K}}_1\bar{\mathcal{K}}_2J\left(c_1,c_2\right)J(d_1,d_2).
\end{aligned}
$$

by choosing d_n and d_T such that $n_1/n\to 1$ and $T_1/T\to 1$.

For L_{2nT} in (A.5), the first step is to show

$$
\begin{aligned}
&\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left(K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right)K^2\left(\frac{d_{ts_{(t)}}}{d_T}\right)\right.\\
&\left.-K^2\left(\frac{d_{ia}}{d_n}\right)K^2\left(\frac{d_{tu}}{d_T}\right)\right)\gamma_{(it,au)}^{(c_1c_2)}\left(\gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)}-\gamma_{(\bar{\imath}\bar{t},b_{(a)}v_{(u)})}^{(d_1d_2)}\right)=o(1),
\end{aligned}
\tag{A.7}
$$

and the second step is to prove

$$
\begin{aligned}
&\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}K^2\left(\frac{d_{ia}}{d_n}\right)K^2\left(\frac{d_{tu}}{d_T}\right)\\
&\times\gamma_{(it,au)}^{(c_1c_2)}\left(\gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)}-\gamma_{(\bar{\imath}\bar{t},b_{(a)}v_{(u)})}^{(d_1d_2)}\right)=o(1).
\end{aligned}
\tag{A.8}
$$

For (A.7),

$$
\begin{aligned}
&\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left(K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right)K^2\left(\frac{d_{ts_{(t)}}}{d_T}\right)-K^2\left(\frac{d_{ia}}{d_n}\right)K^2\left(\frac{d_{tu}}{d_T}\right)\right)\\
&\gamma_{(it,au)}^{(c_1c_2)}\left(\gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)}-\gamma_{(\bar{\imath}\bar{t},b_{(a)}v_{(u)})}^{(d_1d_2)}\right)\\
&\le\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left|\gamma_{(it,au)}^{(c_1c_2)}\right|\left|\gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)}-\gamma_{(\bar{\imath}\bar{t},b_{(a)}v_{(u)})}^{(d_1d_2)}\right|\\
&=\frac{1}{nT}\sum_{(i,a)\in\mathcal{F}_1}\sum_{(t,u)\in\mathcal{G}_1}\left|\gamma_{(it,au)}^{(c_1c_2)}\right|\left(\frac{1}{\ell_n\ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left|\gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)}-\gamma_{(\bar{\imath}\bar{t},b_{(a)}v_{(u)})}^{(d_1d_2)}\right|\right)\\
&+\frac{1}{nT}\sum_{(i,a)\in\mathcal{F}_2}\sum_{(t,u)\in\mathcal{G}_1}\left|\gamma_{(it,au)}^{(c_1c_2)}\right|\left(\frac{1}{\ell_n\ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left|\gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)}-\gamma_{(\bar{\imath}\bar{t},b_{(a)}v_{(u)})}^{(d_1d_2)}\right|\right)\\
&+\frac{1}{nT}\sum_{(i,a)\in\mathcal{F}_1}\sum_{(t,u)\in\mathcal{G}_2}\left|\gamma_{(it,au)}^{(c_1c_2)}\right|\left(\frac{1}{\ell_n\ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left|\gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)}-\gamma_{(\bar{\imath}\bar{t},b_{(a)}v_{(u)})}^{(d_1d_2)}\right|\right)\\
&+\frac{1}{nT}\sum_{(i,a)\in\mathcal{F}_2}\sum_{(t,u)\in\mathcal{G}_2}\left|\gamma_{(it,au)}^{(c_1c_2)}\right|\left(\frac{1}{\ell_n\ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left|\gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)}-\gamma_{(\bar{\imath}\bar{t},b_{(a)}v_{(u)})}^{(d_1d_2)}\right|\right)\\
&=M_{1nT}+M_{2nT}+M_{3nT}+M_{4nT},
\end{aligned}
$$

where

$$\mathcal{F}_1 = \{(i,a) : d_{ia} \le f_n \ \&\ i,a \in E_n\}, \quad \mathcal{F}_2 = \{(i,a) : d_{ia} > f_n \ \&\ i,a \in E_n\},$$
$$\mathcal{G}_1 = \{(t,u) : d_{tu} \le g_T \ \&\ t,u \in E_T\}, \quad \mathcal{G}_2 = \{(t,u) : d_{tu} > g_T \ \&\ t,u \in E_T\},$$

in which $f_n/d_n = O(1)$ and $g_T/d_T = O(1)$.

For M_{1nT}, we obtain

$$\begin{aligned}M_{1nT} &\le \left(\frac{\ell_n\ell_T}{nT}\right)\frac{1}{\ell_n\ell_T}\sum_{(i,a)\in\mathcal{F}_1}\sum_{(t,u)\in\mathcal{G}_1}\left|\gamma^{(c_1c_2)}_{(it,au)}\right|\left(\frac{1}{\ell_n\ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left|\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right|\right.\\
&\quad\left.+\frac{1}{\ell_n^2\ell_T^2}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{h_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{w_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left|\gamma^{(d_1d_2)}_{(h_{(i)}w_{(t)},b_{(a)}v_{(u)})}\right|\right)\\
&= O\left(\frac{\ell_n\ell_T}{nT}\right).\end{aligned}$$

For M_{2nT},

$$\begin{aligned}M_{2nT} &\le \left(\frac{\ell_T}{T}\right)\frac{1}{nd_T}\sum_{(i,a)\in\mathcal{F}_2}\sum_{(t,u)\in\mathcal{G}_1}\left|\gamma^{(c_1c_2)}_{(it,au)}\right|\left(\frac{1}{\ell_n\ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left|\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right|\right.\\
&\quad\left.+\frac{1}{\ell_n^2\ell_T^2}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{h_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{w_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left|\gamma^{(d_1d_2)}_{(h_{(i)}w_{(t)},b_{(a)}v_{(u)})}\right|\right)\\
&= O\left(\frac{\ell_T}{T}\right).\end{aligned}$$

It is straightforward that $M_{3nT} = O\left(\ell_n/n\right)$.

For M_{4nT},

$$\begin{aligned}M_{4nT} &\le \frac{1}{f_n^q}\frac{1}{nT}\sum_{(i,a)\in\mathcal{F}_2}\sum_{(t,u)\in\mathcal{G}_2}\left|\gamma^{(c_1c_2)}_{(it,au)}\right|d_{ia}^q\left(\frac{1}{\ell_n\ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left|\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right|\right.\\
&\quad\left.+\frac{1}{\ell_n^2\ell_T^2}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{h_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{w_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left|\gamma^{(d_1d_2)}_{(h_{(i)}w_{(t)},b_{(a)}v_{(u)})}\right|\right)\\
&= O\left(f_n^{-q}\right).\end{aligned}$$

By choosing f_n and g_T such that $f_n = O(d_n)$ and $g_T = O(d_T)$, we obtain

$$M_{1nT} = o(1),\ M_{2nT} = o(1),\ M_{3nT} = o(1),\ \text{and } M_{4nT} = o(1).$$

Therefore, (A.7) holds.

The next step is to show (A.8).

$$\begin{aligned}&\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}K^2\left(\frac{d_{ia}}{d_n}\right)K^2\left(\frac{d_{tu}}{d_T}\right)\gamma^{(c_1c_2)}_{(it,au)}\left(\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}-\gamma^{(d_1d_2)}_{(\bar{i}\bar{t},b_{(a)}v_{(u)})}\right)\\
&= \frac{1}{nT}\sum_{i,a\in E_n}\sum_{t,u\in E_T}K^2\left(\frac{d_{ia}}{d_n}\right)K^2\left(\frac{d_{tu}}{d_T}\right)\gamma^{(c_1c_2)}_{(it,au)}\left(\frac{1}{\ell_n\ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left(\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}-\gamma^{(d_1d_2)}_{(\bar{i}\bar{t},b_{(a)}v_{(u)})}\right)\right)\\
&= o(1),\end{aligned}$$

because

$$\frac{1}{\ell_n \ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left(\gamma^{(d_1 d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}-\gamma^{(d_1 d_2)}_{\left(\bar{\imath}\bar{t},b_{(a)}v_{(u)}\right)}\right)$$

$$=\left[\frac{1}{\ell_n \ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\gamma^{(d_1 d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}-\left(\frac{1}{\ell_n \ell_T}\right)^2\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\left(\sum_{h_{(i)}=1}^{\ell_{i,n}}\sum_{w_{(t)}=1}^{\ell_{t,T}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\gamma^{(d_1 d_2)}_{(h_{(i)}w_{(t)},b_{(a)}v_{(u)})}\right)\right]$$

$$=\left[\frac{1}{\ell_n \ell_T}\sum_{j:d_{ij}\leq d_n}\sum_{b:d_{ab}\leq d_n}\sum_{s:d_{ts}\leq d_T}\sum_{v:d_{uv}\leq d_T}\gamma^{(d_1 d_2)}_{(js,bv)}-\left(\frac{1}{\ell_n \ell_T}+o(1)\right)\sum_{h:d_{ih}\leq d_n}\sum_{b:d_{ab}\leq d_n}\sum_{w:d_{tw}\leq d_T}\sum_{v:d_{uv}\leq d_T}\gamma^{(d_1 d_2)}_{(hw,bv)}\right]$$

$$=\left[\frac{1}{\ell_n \ell_T}\sum_{j:d_{ij}\leq d_n}\sum_{b:d_{ib}\leq d_n}\sum_{s:d_{ts}\leq d_T}\sum_{v:d_{tv}\leq d_T}\gamma^{(d_1 d_2)}_{(js,bv)}-\left(\frac{1}{\ell_n \ell_T}+o(1)\right)\sum_{h:d_{ih}\leq d_n}\sum_{b:d_{ib}\leq d_n}\sum_{w:d_{tw}\leq d_T}\sum_{v:d_{tv}\leq d_T}\gamma^{(d_1 d_2)}_{(hw,bv)}+o(1)\right]$$

$$\to 0$$

by (A.6) and Assumption I6. Therefore, $L_{1nT}=o(1)$ and $L_{2nT}=(1)$, which complete the proof of (A.3).

By the same argument, we obtain the result that

$$\lim_{n,T\to\infty}\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}K^2\left(\frac{d_{ab_{(a)}}}{d_n}\right)K^2\left(\frac{d_{uv}}{d_T}\right)\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1 d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}$$
$$=\bar{\mathcal{K}}_1\bar{\mathcal{K}}_2 J(c_1,c_2)\,J(d_1,d_2),$$

which completes the proof of (A.3).

The next step is to prove (A.4). In view of previous derivations, it suffices to show that

$$\lim_{n,T\to\infty}\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{u,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)-K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right]^2$$
$$\times\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1 d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}=0. \tag{A.9}$$

$$\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)-K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right]^2$$
$$\times\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1 d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}$$
$$=\frac{1}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_1}\sum_{(t,s,u,v)\in\mathcal{J}_1}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)-K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right]^2\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1 d_2)}_{\left(j_{(i)}s_{(t)},b_{(a)}v_{(u)}\right)}$$
$$+\frac{1}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_2}\sum_{(t,s,u,v)\in\mathcal{J}_1}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)-K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right]^2\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1 d_2)}_{\left(j_{(i)}s_{(t)},b_{(a)}v_{(u)}\right)}$$
$$+\frac{1}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_1}\sum_{(t,s,u,v)\in\mathcal{J}_2}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)-K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right]^2\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1 d_2)}_{\left(j_{(i)}s_{(t)},b_{(a)}v_{(u)}\right)}$$
$$+\frac{1}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_2}\sum_{(t,s,u,v)\in\mathcal{J}_2}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)-K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right]^2\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1 d_2)}_{\left(j_{(i)}s_{(t)},b_{(a)}v_{(u)}\right)}$$
$$:=F_{1nT}+F_{2nT}+F_{3nT}+F_{4nT},$$

where

$$\mathcal{I}_1 = \left\{(i,j,a,b) : \left|d_{ij_{(i)}} - d_{ab_{(a)}}\right| \leq 2c_n \ \&\ i,a \in E_n\right\},$$
$$\mathcal{I}_2 = \left\{(i,j,a,b) : \left|d_{ij_{(i)}} - d_{ab_{(a)}}\right| > 2c_n \ \&\ i,a \in E_n\right\},$$
$$\mathcal{J}_1 = \left\{(t,s,u,v) : \left|d_{ts_{(t)}} - d_{uv_{(u)}}\right| \leq 2c_T \ \&\ t,u \in E_T\right\},$$
$$\mathcal{J}_2 = \left\{(t,s,u,v) : \left|d_{ts_{(t)}} - d_{uv_{(u)}}\right| > 2c_T \ \&\ t,u \in E_T\right\}.$$

For F_{1nT}, we have

$$\begin{aligned}
&F_{1nT} \\
&\leq \left|\frac{1}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_1}\sum_{(t,s,u,v)\in\mathcal{J}_1}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right) - K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right]^2\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right| \\
&= \left|\frac{1}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_1}\sum_{(t,s,u,v)\in\mathcal{J}_1}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)\left(K\left(\frac{d_{ts_{(t)}}}{d_T}\right) - K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right)\right.\right. \\
&\left.\left.+K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\left(K\left(\frac{d_{ij_{(i)}}}{d_n}\right) - K\left(\frac{d_{ab_{(a)}}}{d_n}\right)\right)\right]^2\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right| \\
&\leq \left|\frac{2}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_1}\sum_{(t,s,u,v)\in\mathcal{J}_1}\left[K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right)\left(K\left(\frac{d_{ts_{(t)}}}{d_T}\right) - K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right)^2\right.\right. \\
&\left.\left.+K^2\left(\frac{d_{uv_{(u)}}}{d_T}\right)\left(K\left(\frac{d_{ij_{(i)}}}{d_n}\right) - K\left(\frac{d_{ab_{(a)}}}{d_n}\right)\right)^2\right]\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right| \\
&\leq \frac{2c_L^2}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_1}\sum_{(t,s,u,v)\in\mathcal{J}_1}\left[\left(\frac{d_{ts_{(t)}}}{d_T} - \frac{d_{uv_{(u)}}}{d_T}\right)^2 + \left(\frac{d_{ij_{(i)}}}{d_n} - \frac{d_{ab_{(a)}}}{d_n}\right)^2\right]\left|\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(c_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right| \\
&\leq 8c_L^2\left(\frac{c_T^2}{d_T^2} + \frac{c_n^2}{d_n^2}\right)\left(\frac{1}{nT}\sum_{i,a\in E_n}\sum_{t,u\in E_T}\left|\gamma^{(c_1c_2)}_{(it,au)}\right|\right)\left(\frac{1}{\ell_n\ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left|\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right|\right) \\
&= O\left(\frac{c_T^2}{d_T^2}\right) + O\left(\frac{c_n^2}{d_n^2}\right), \qquad \text{(A.10)}
\end{aligned}$$

because under Assumption I3

$$\begin{aligned}
\frac{1}{\ell_n\ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left|\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right| &= \frac{1}{\ell_n\ell_T}\sum_{j:d_{ij}\leq d_n}\sum_{b:d_{ab}\leq d_n}\sum_{s:d_{ts}\leq d_T}\sum_{v:d_{uv}\leq d_T}\left|\gamma^{(d_1d_2)}_{(js,bv)}\right| \\
&\leq \frac{1}{\ell_n\ell_T}\sum_{\substack{j:d_{ij}\leq d_n \\ \text{or } d_{aj}\leq d_n}}\sum_{\substack{b:d_{ib}\leq d_n \\ \text{or } d_{ab}\leq d_n}}\sum_{\substack{s:d_{ts}\leq d_T \\ \text{or } d_{us}\leq d_T}}\sum_{\substack{v:d_{tv}\leq d_T \\ \text{or } d_{uv}\leq d_T}}\left|\gamma^{(d_1d_2)}_{(js,bv)}\right| \\
&= O(1). \qquad \text{(A.11)}
\end{aligned}$$

For F_{2nT} we note that if $\left|d_{ij_{(i)}} - d_{ab_{(a)}}\right| > 2c_n$, then either $d_{ia} > c_n$ or $d_{j_{(i)}b_{(a)}} > c_n$. Otherwise, if both

$d_{ia} \leq c_n$ and $d_{j_{(i)}b_{(a)}} \leq c_n$, then

$$\begin{aligned} d_{ij_{(i)}} - d_{ab_{(a)}} &\leq d_{ia} + d_{ab_{(a)}} + d_{b_{(a)}j_{(i)}} - d_{ab_{(a)}} \leq 2c_n, \\ d_{ij_{(i)}} - d_{ab_{(a)}} &\geq d_{ij_{(i)}} - d_{ia} - d_{ij_{(i)}} - d_{j_{(i)}b_{(a)}} \geq -2c_n. \end{aligned}$$

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

$$\begin{aligned} & F_{2nT} \\ \leq & \left| \frac{1}{nT\ell_n\ell_T} \sum_{(i,j,a,b)\in\mathcal{I}_2} \sum_{(t,s,u,v)\in\mathcal{J}_1} \left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right) K\left(\frac{d_{ts_{(t)}}}{d_T}\right) - K\left(\frac{d_{ab_{(a)}}}{d_n}\right) K\left(\frac{d_{uv_{(u)}}}{d_T}\right) \right]^2 \gamma_{(it,au)}^{(c_1c_2)} \gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)} \right| \\ \leq & \left| \frac{2}{nT\ell_n\ell_T} \sum_{(i,j,a,b)\in\mathcal{I}_2} \sum_{(t,s,u,v)\in\mathcal{J}_1} \left[K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right) \left(K\left(\frac{d_{ts_{(t)}}}{d_T}\right) - K\left(\frac{d_{uv_{(u)}}}{d_T}\right) \right)^2 \right.\right. \\ & \left.\left. + K^2\left(\frac{d_{uv_{(u)}}}{d_T}\right) \left(K\left(\frac{d_{ij_{(i)}}}{d_n}\right) - K\left(\frac{d_{ab_{(a)}}}{d_n}\right) \right)^2 \right] \gamma_{(it,au)}^{(c_1c_2)} \gamma_{(j_{(i)}s,b_{(a)}v)}^{(d_1d_2)} \right| \\ \leq & \frac{8c_L^2}{nT\ell_n\ell_T} \sum_{(i,j,a,b)\in\mathcal{I}_2} \sum_{(t,s,u,v)\in\mathcal{J}_1} \left(\frac{c_T}{d_T}\right)^2 \left| \gamma_{(it,au)}^{(c_1c_2)} \gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)} \right| \\ & + \frac{8}{nT\ell_n\ell_T} \sum_{(i,j,a,b)\in\mathcal{I}_2} \sum_{(t,s,u,v)\in\mathcal{J}_1} (d_{ia})^q \left| \gamma_{(it,au)}^{(c_1c_2)} \gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)} \right| (d_{ia})^{-q} \\ \leq & \frac{8}{(c_n)^q} \left(\frac{1}{nT} \sum_{i,a\in E_n} \sum_{t,u\in E_T} (d_{ia})^q \left| \gamma_{(it,au)}^{(c_1c_2)} \right| \right) \left(\frac{1}{\ell_n\ell_T} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} \left| \gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)} \right| \right) + O\left(\frac{c_T^2}{d_T^2}\right) \\ = & O\left(\frac{1}{c_n^q}\right) + O\left(\frac{c_T^2}{d_T^2}\right). \end{aligned}$$

With the similar procedure, we can show that $F_{3nT} = O\left(c_T^{-q}\right) + O\left(c_n^2/d_n^2\right)$.

For F_{4nT},

$$\begin{aligned} & F_{4nT} \\ \leq & \left| \frac{1}{nT\ell_n\ell_T} \sum_{(i,j,a,b)\in\mathcal{I}_2} \sum_{(t,s,u,v)\in\mathcal{J}_2} \left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right) K\left(\frac{d_{ts_{(t)}}}{d_T}\right) - K\left(\frac{d_{ab_{(a)}}}{d_n}\right) K\left(\frac{d_{uv_{(u)}}}{d_T}\right) \right]^2 \gamma_{(it,au)}^{(c_1c_2)} \gamma_{(j_{(i)}s,b_{(a)}v)}^{(d_1d_2)} \right| \\ \leq & \frac{4}{nT\ell_n\ell_T} \sum_{(i,j,a,b)\in\mathcal{I}_2} \sum_{(t,s,u,v)\in\mathcal{J}_2} (d_{ia})^q \left| \gamma_{(it,au)}^{(c_1c_2)} \gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)} \right| (d_{ia})^{-q} \\ \leq & \frac{4}{(c_n)^q} \left(\frac{1}{nT} \sum_{i,a\in E_n} \sum_{t,u\in E_T} (d_{ia})^q \left| \gamma_{(it,au)}^{(c_1c_2)} \right| \right) \left(\frac{1}{\ell_n\ell_T} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} \left| \gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)} \right| \right) \\ = & O\left(\frac{1}{c_n^q}\right). \end{aligned}$$

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

$$F_{1nT} = o(1)\ , F_{2nT} = o(1),\ F_{3nT} = o(1) \text{ and } F_{4nT} = o(1)$$

and (A.4) is proved.

Next, we show that D_{2nT} is $o(1)$. For D_{2nT},

$$D_{2nT} \leq \frac{1}{nT} \sum_{i,a=1}^{n} \sum_{t \notin E_T} \sum_{u=1}^{T} \left| \gamma_{(it,au)}^{(c_1 c_2)} \right| \left(\frac{1}{\ell_n \ell_T} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} \left| \gamma_{(j_{(i)} s_{(t)}, b_{(a)} v_{(u)})}^{(d_1 d_2)} \right| \right)$$
$$= o(1), \tag{A.12}$$

as $T_2/T \to 0$. With the same procedure, we can show D_{3nT} is $o(1)$.

For D_{4nT},

$$D_{4nT} \leq \frac{1}{nT} \sum_{i \notin E_n} \sum_{a=1}^{n} \sum_{t \in E_T} \sum_{u \in E_T} \left| \gamma_{(it,au)}^{(c_1 c_2)} \right| \left(\frac{1}{\ell_n \ell_T} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} \left| \gamma_{(j_{(i)} s_{(t)}, b_{(a)} v_{(u)})}^{(d_1 d_2)} \right| \right)$$
$$= o(1) \tag{A.13}$$

by choosing the sequence of d_n in a way that $n_2/n \to 0$ as $n \to \infty$. We can also show $D_{5,nT}$ in the symmetric way. By symmetry, it is straightforward that

$$\lim_{n,T\to\infty} C_{3nT} = \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 J(c_1, d_2) J(c_2, d_1).$$

Therefore,

$$\lim_{n,T\to\infty} \frac{nT}{\ell_n \ell_T} \text{cov}\left(\tilde{J}_{nT}(c_1, d_1), \tilde{J}_{nT}(c_2, d_2) \right) = \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 (J(c_1, c_2) J(d_1, d_2) + J(c_1, d_2) J(c_2, d_1)).$$

In terms of matrix form,

$$\lim_{n,T\to\infty} \frac{nT}{\ell_n \ell_T} \text{var}\left(\text{vec}\left(\tilde{J}_{nT} \right) \right) = \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 (I_{pp} + \mathbb{K}_{pp}) (J \otimes J),$$

where $J = [J(c,d)]$, $c, d = 1, \ldots, p$.

(b) Asymptotic Bias

Let $d_T = k_{nT} d_n$ and $k_{nT} = k + o(1)$ where $k > 0$. We have

$$d_n^q \left(E\tilde{J}_{nT} - J_{nT} \right) = \frac{1}{nT} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} \Gamma_{(it,js)} \left[(d_{ij})^q \frac{K\left(\frac{d_{ij}}{d_n}\right) - 1}{\left(\frac{d_{ij}}{d_n}\right)^q} + \left(\frac{d_{ts}}{k_{nT}}\right)^q \frac{K\left(\frac{d_{ts}}{d_T}\right) - 1}{\left(\frac{d_{ts}}{d_T}\right)^q} \right.$$
$$\left. + (d_{ij})^q \left(\frac{d_{ts}}{d_T}\right)^q \frac{\left(K\left(\frac{d_{ij}}{d_n}\right) - 1\right)\left(K\left(\frac{d_{ts}}{d_T}\right) - 1\right)}{\left(\frac{d_{ij}}{d_n}\right)^q \left(\frac{d_{ts}}{d_T}\right)^q} \right]$$
$$= -K_q b_1^{(q)} - \frac{1}{k^q} K_q b_2^{(q)} + o(1).$$

Therefore, $\lim_{n,T\to\infty} d_n^q (\tilde{J}_{nT} - J_{nT}) = -K_q b_1^{(q)} - \frac{1}{k^q} K_q b_2^{(q)}$.

(c) $\sqrt{\frac{nT}{\ell_n\ell_T}}\left(\hat{J}_{nT}-J_{nT}\right)=O_p(1)$ **and** $\sqrt{\frac{nT}{\ell_n\ell_T}}\left(\hat{J}_{nT}-\tilde{J}_{nT}\right)=o_p(1)$

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

$$
\begin{aligned}
\sqrt{\frac{nT}{\ell_n\ell_T}}\left(\hat{J}_{nT}-\tilde{J}_{nT}\right)=&\sqrt{\frac{nT}{\ell_n\ell_T}}\frac{1}{nT}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\left[\hat{V}_{(i,t)}\hat{V}'_{(j,s)}-V_{(i,t)}V'_{(j,s)}\right]\\
=&\left(\sqrt{nT}\left(\hat{\beta}-\beta_0\right)\right)^2\sqrt{\frac{\ell_n\ell_T}{nT}}\frac{1}{\ell_n\ell_T nT}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{it}^2\tilde{X}_{js}^2\\
&-2\sqrt{nT}\left(\hat{\beta}-\beta_0\right)\sqrt{\frac{\ell_n\ell_T}{nT}}\frac{1}{\ell_n\ell_T\sqrt{nT}}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{js}^2\tilde{X}_{it}\tilde{u}_{it}\\
&-\frac{2}{\sqrt{\ell_n\ell_T nT}}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{it}u_{it}\tilde{X}_{js}\left(\bar{u}_j+\bar{u}_s-\bar{u}\right)\\
&+\frac{1}{\sqrt{\ell_n\ell_T nT}}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{it}\tilde{X}_{js}\left(\bar{u}_i+\bar{u}_t-\bar{u}\right)\left(\bar{u}_j+\bar{u}_s-\bar{u}\right)\\
=&H_{1nT}+H_{2nT}+H_{3nT}+H_{4nT}.
\end{aligned}
$$

For H_{1nT},

$$
\begin{aligned}
\frac{1}{\ell_n\ell_T nT}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{it}^2\tilde{X}_{js}^2=&\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^2\left(\frac{1}{\ell_n\ell_T}\sum_{j=1}^{n}\sum_{s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{js}^2\right)\\
\leq&\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^2\left(\frac{\ell_{i,n}\ell_{t,T}}{\ell_n\ell_T}\right)\left(\frac{1}{\ell_{i,n}\ell_{t,T}}\sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T}\tilde{X}_{js}^2\right)\\
=&O_p(1)
\end{aligned}
$$

by Assumptions I4 and I7 (iv). Therefore $H_{1nT}=o_p(1)$ by Assumptions I7 (i).

For H_{2nT}, it can be rewritten as

$$
H_{2nT}=-2\sqrt{nT}\left(\hat{\beta}-\beta_0\right)\sqrt{\frac{\ell_n\ell_T}{nT}}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\left(\frac{1}{\ell_n\ell_T}\sum_{j=1}^{n}\sum_{s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{js}^2\right)
$$

and the part in the parenthesis is $O_p\left(1\right)$ uniformly by Assumption I7 (iv). As

$$
\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}u_{it}=O_p\left(1\right)
$$

by Assumption I7(iii), $H_{2nT}=o_p(1)$.

For $H_{3nT,}$ we need to show that for all i and t

$$
\frac{1}{\sqrt{\ell_n\ell_T}}\sum_{j=1}^{n}\sum_{s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{js}\left(\bar{u}_j+\bar{u}_s-\bar{u}\right)=o_p\left(1\right). \tag{A.14}
$$

First, $\bar{u}_j + \bar{u}_s - \bar{u} = o_p(1)$ uniformly. Second, by Assumption I7 (iv)

$$P\left(\left|\frac{1}{\sqrt{\ell_n \ell_T}} \sum_{j=1}^{n} \sum_{s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right) K\left(\frac{d_{ts}}{d_T}\right) \tilde{X}_{js}\right| > \Delta\right) \leq \frac{1}{\Delta^2} \frac{2}{\ell_n \ell_T} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{s_{(t)}=1}^{\ell_{t,T}} E\left[\tilde{X}_{j_{(i)} s_{(t)}}\right]^2 \to 0,$$

as $\Delta \to \infty$. Therefore, $H_{3nT} = o_p(1)$. With the similar procedures, we can show that H_{4nT} is $o_p(1)$.

As a result,

$$\sqrt{\frac{nT}{\ell_n \ell_T}} \left(\hat{J}_{nT} - \tilde{J}_{nT}\right) = o_p(1).$$

(d) AMSE

The first equality holds by Theorem 1(c). For the last equality of Theorem 1(d), since

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

we have

$$\begin{aligned}
&\lim_{n,T\to\infty} MSE\left(\frac{nT}{\ell_n \ell_T}, \tilde{J}_{nT}, S_{nT}\right) \\
&= \lim_{n,T\to\infty} \frac{nT}{\ell_n \ell_T} \text{ vec }\left(E\tilde{J}_{nT} - J_{nT}\right)' S_{nT} \text{ vec}\left(E\tilde{J}_{nT} - J_{nT}\right) + \lim_{n,T\to\infty} \frac{nT}{\ell_n \ell_T} \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 \text{tr}\left(S_{nT} \text{var}(\text{vec } \tilde{J}_{nT})\right) \\
&= \frac{1}{\tau} K_q^2 \text{ vec}\left(b_1^{(q)} + \frac{1}{k^q} b_2^{(q)}\right)' S \text{ vec}\left(b_1^{(q)} + \frac{1}{k^q} b_2^{(q)}\right) + \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 tr\left[S(I_{pp} + K_{pp})(J \otimes J)\right],
\end{aligned}$$

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

Proof of Corollary 1

Letting $k_{nT} = d_T/d_n$ and $k_{nT} \to k$ as $n, T \to \infty$. By Theorem 1(d), we obtain

$$\begin{aligned}
&\lim_{n,T\to\infty} \max_{(b_1,b_2)\in\mathfrak{B}} MSE\left((nT)^{2q/(2q+\eta_n+\eta_T)}, \hat{J}_{nT}(d_n, d_T), S_{nT}\right) \\
&= \lim_{n,T\to\infty} (nT)^{2q/(2q+\eta_n+\eta_T)} \left(\frac{2K_q^2}{d_n^{2q}}\left(B_{11} + \frac{B_{22}}{k_{nT}^{2q}}\right) + \frac{\ell_n \ell_T}{nT} \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C\right) \\
&= \lim_{n,T\to\infty} (\alpha_n \alpha_T k_{nT}^{\eta_T})^{2q/(2q+\eta_n+\eta_T)} \left(\frac{d_n^{2q} \ell_n \ell_T}{nT}\right)^{(\eta_n+\eta_T)/(2q+\eta_n+\eta_T)} \\
&\times \left(\frac{2K_q^2}{d_n^{2q} \ell_n \ell_T / nT}\left(B_{11} + \frac{B_{22}}{k_{nT}^{2q}}\right) + \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C\right) \\
&= (\alpha_1 \alpha_2 k^{\eta_T})^{2q/(2q+\eta_n+\eta_T)} \tau^{(\eta_n+\eta_T)/(2q+\eta_n+\eta_T)} \left(\frac{2K_q^2}{\tau}\left(B_{11} + \frac{B_{22}}{k^{2q}}\right) + \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C\right),
\end{aligned}$$

It is straightforward to show that this is uniquely minimized over $\tau \in (0, \infty)$ by

$$\tau^\star = \frac{4qK_q^2\left(B_{11} + \frac{B_{22}}{(k^\star)^{2q}}\right)}{(\eta_n + \eta_T) \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C} \text{ and } k^\star = \left(\frac{2(2q + \eta_n) K_q^2 B_{22}}{\eta_T \left(2K_q^2 B_{11} + \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C \tau^\star\right)}\right)^{1/(2q)},$$

since S is pd. Therefore,

$$\tau^{\star} = \frac{4qK_q^2 B_{11}}{\eta_n \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C} \text{ and } k^{\star} = \left(\frac{\eta_n B_{22}}{\eta_T B_{11}}\right)^{\frac{1}{2q}}$$

and that a sequence $\{(d_n, d_T)\}$ satisfies $d_n^{2q}\ell_n\ell_T/nT \to \tau^{\star}$ if and only if $d_n = d_n^{\star} + o\left((nT)^{1/(2q+\eta+1)}\right)$ and $d_T = d_T^{\star} + o\left((nT)^{1/(2q+\eta+1)}\right)$.

Proof of Theorem 2

(a) Asymptotic variance

Let $\varphi_{lkcd} = \sum_{i=1}^{n}\sum_{t,s=1}^{T} r_{(i,t),l}^{(c)} r_{(i,s),k}^{(d)}$. As in the proof of Theorem 1 (a), we have

$$n \cdot cov\left(\tilde{J}_{nT}^{A}(c_1, d_1), \tilde{J}_{nT}^{A}(c_2, d_2)\right) = C_{1nT} + C_{2nT} + C_{3nT},$$

where

$$C_{1nT} = \frac{1}{nT^2}\sum_{l=1}^{nTp}\left(E\varepsilon_l^4 - 3\right)\left[\sum_{i=1}^{n}\sum_{t,s=1}^{T} r_{(i,t),l}^{(c_1)} r_{(i,s),l}^{(d_1)} \sum_{a=1}^{n}\sum_{u,v=1}^{T} r_{(a,u),l}^{(c_2)} r_{(a,v),l}^{(d_2)}\right],$$

$$C_{2nT} = \frac{1}{nT^2}\sum_{l,k=1}^{nTp}\left[\sum_{i=1}^{n}\sum_{t,s=1}^{T} r_{(i,t),l}^{(c_1)} r_{(i,s),k}^{(d_1)} \sum_{a=1}^{n}\sum_{u,v=1}^{T} r_{(a,u),l}^{(c_2)} r_{(a,v),k}^{(d_2)}\right],$$

$$C_{3nT} = \frac{1}{nT^2}\sum_{l,k=1}^{nTp}\left[\sum_{i=1}^{n}\sum_{t,s=1}^{T} r_{(i,t),l}^{(c_1)} r_{(j,s),k}^{(d_1)} \sum_{a=1}^{n}\sum_{u,v=1}^{T} r_{(a,u),k}^{(c_2)} r_{(a,v),l}^{(d_2)}\right].$$

We can show $C_{1nT} = o(1)$ in a similar way to the proof of Theorem 1 (a).

For C_{2nT}, we have

$$\begin{aligned} C_{2nT} &= \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t,s,u,v=1}^{T}\gamma_{(it,iu)}^{(c_1c_2)}\gamma_{(\bar{\imath}s,\bar{\imath}v)}^{(d_1d_2)} + \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t,s,u,v=1}^{T}\gamma_{(it,iu)}^{(c_1c_2)}\left(\gamma_{(is,iv)}^{(d_1d_2)} - \gamma_{(\bar{\imath}s,\bar{\imath}v)}^{(d_1d_2)}\right) \\ &= L_{1nT} + L_{2nT}, \end{aligned} \tag{A.15}$$

where $\gamma_{(\bar{\imath}s,\bar{\imath}v)}^{(d_1d_2)} = n^{-1}\sum_{j=1}^{n}\gamma_{(js,jv)}^{(d_1d_2)}$ by Assumption I10. For L_{1nT},

$$L_{1nT} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t,u=1}^{T}\gamma_{(it,iu)}^{(c_1c_2)}\left(\frac{1}{nT}\sum_{j=1}^{n}\sum_{s,v=1}^{T}\gamma_{(js,kv)}^{(d_1d_2)}\right) \to J(c_1,c_2)\,J(d_1,d_2). \tag{A.16}$$

as $n, T \to \infty$. For L_{2nT} in (A.15),

$$L_{2nT} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t,u=1}^{T}\gamma_{(it,iu)}^{(c_1c_2)}\frac{1}{T}\sum_{s,v=1}^{T}\left(\gamma_{(is,iv)}^{(d_1d_2)} - \gamma_{(\bar{\imath}s,\bar{\imath}v)}^{(d_1d_2)}\right) = o(1) \tag{A.17}$$

by Assumption I11. From (A.16) and (A.17)

$$\lim_{n,T\to\infty} C_{2nT} = J(c_1,c_2)\,J(d_1,d_2).$$

With the same procedure, it is straightforward that

$$\lim_{n,T\to\infty} C_{3nT} = J\left(c_1, d_2\right) J\left(c_2, d_1\right).$$

Therefore,

$$\lim_{n\to\infty}\lim_{T\to\infty} n\cdot \text{var}\left(\text{vec}\left(\tilde{J}^{A}_{nT}\right)\right) = \left(I_{pp} + \mathbb{K}_{pp}\right)\left(J\otimes J\right).$$

(b) $\sqrt{\frac{nT}{\ell_n d_T}}\left(\hat{J}^{A}_{nT} - J_{nT}\right) = O_p(1)$ **and** $\sqrt{\frac{nT}{\ell_n d_T}}\left(\hat{J}^{A}_{nT} - \tilde{J}^{A}_{nT}\right) = o_p(1)$

The proof is analogous to the proof of Theorem 1 (c).

Proof of Theorem 3

(a) Asymptotic Variance

Let $\varphi_{lkcd} = \sum_{i,j=1}^{n}\sum_{t,s=1}^{T} r^{(c)}_{(i,t),l} r^{(d)}_{(j,s),k} K\left(\frac{d_{ts}}{d_T}\right)$. As in the proof of Theorem 1 (a), We have

$$\frac{T}{\ell_T} cov\left(\tilde{J}^{DK}_{nT}\left(c_1, d_1\right), \tilde{J}^{DK}_{nT}\left(c_2, d_2\right)\right) := C_{1nT} + C_{2nT} + C_{3nT},$$

where

$$\begin{aligned}
C_{1nT} &= \frac{1}{n^2 T\ell_T}\sum_{l=1}^{nTp}\left(E\varepsilon_l^4 - 3\right)\left[\sum_{i,j=1}^{n}\sum_{t,s=1}^{T} r^{(c_1)}_{(i,t),l} r^{(d_1)}_{(j,s),l} K\left(\frac{d_{ts}}{d_T}\right)\sum_{a,b1}^{n}\sum_{u,v=1}^{T} r^{(c_2)}_{(a,u),l} r^{(d_2)}_{(b,v),l} K\left(\frac{d_{uv}}{d_T}\right)\right],\\
C_{2nT} &= \frac{1}{n^2 T\ell_T}\sum_{l,k=1}^{nTp}\left[\sum_{i,j=1}^{n}\sum_{t,s=1}^{T} r^{(c_1)}_{(i,t),l} r^{(d_1)}_{(j,s),k} K\left(\frac{d_{ts}}{d_T}\right)\sum_{a,b=1}^{n}\sum_{u,v=1}^{T} r^{(c_2)}_{(a,u),l} r^{(d_2)}_{(b,v),k} K\left(\frac{d_{uv}}{d_T}\right)\right],\\
C_{3nT} &= \frac{1}{n^2 T\ell_T}\sum_{l,k=1}^{nTp}\left[\sum_{i,j=1}^{n}\sum_{t,s=1}^{T} r^{(c_1)}_{(i,t),l} r^{(d_1)}_{(j,s),k} K\left(\frac{d_{ts}}{d_T}\right)\sum_{a,b=1}^{n}\sum_{u,v=1}^{T} r^{(c_2)}_{(a,u),k} r^{(d_2)}_{(b,v),l} K\left(\frac{d_{uv}}{d_T}\right)\right].
\end{aligned}$$

We can show $C_{1nT} = o\left(1\right)$ in a similar way to the proof of Theorem 1 (a).
For C_{2nT}, we can decompose this to control the boundary effects as (A.2) in the proof of Theorem 1 (a).

$$\begin{aligned}
C_{2nT} = &\frac{1}{n^2 T\ell_T}\sum_{i,j,a,b=1}^{n}\sum_{t\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{u\in E_T}\sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ts_{(t)}}}{\ell_T}\right) K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(js_{(t)},bv_{(u)})}\\
&+\frac{1}{n^2 T\ell_T}\sum_{i,j,a,b=1}^{n}\sum_{t\notin E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{u=1}^{T}\sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ts_{(t)}}}{d_T}\right) K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(js_{(t)},bv_{(u)})}\\
&+\frac{1}{n^2 T\ell_T}\sum_{i,j,a,b=1}^{n}\sum_{t\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{u\notin E_T}\sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ts_{(t)}}}{d_T}\right) K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(js_{(t)},bv_{(u)})}\\
&= D_{1nT} + D_{2nT} + D_{3nT}. \qquad \text{(A.18)}
\end{aligned}$$

Following the similar procedures in the proofs of (A.3) and (A.4), we can show

$$D_{1nT} \to \bar{\mathcal{K}}_2 J\left(c_1, c_2\right) J\left(d_1, d_2\right), \tag{A.19}$$

as $n, T \to \infty$. We can also show that D_{2nT} and D_{3nT} are $o(1)$ based on (A.12).

With the symmetric procedure, it is straightforward that

$$\lim_{n,T\to\infty} C_{3nT} = \bar{\mathcal{K}}_2 J(c_1, d_2) J(c2, d_1).$$

As a consequence,

$$\lim_{n,T\to\infty} \frac{T}{\ell_T} \text{var}\left(\text{vec}\left(\tilde{J}_{nT}^{DK}\right)\right) = \bar{\mathcal{K}}_2 (I_{pp} + \mathbb{K}_{pp}) \left(J \otimes J\right).$$

(b) Asymptotic Bias

We have

$$\begin{aligned} d_T^q \left(E\tilde{J}_{nT}^{DK} - J_{nT}\right) &= \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \Gamma_{(it,js)} \left(d_{ts}\right)^q \frac{K\left(\frac{d_{ts}}{d_T}\right) - 1}{\left(\frac{d_{ts}}{d_T}\right)^q} \\ &= -K_q b_2^{(q)} + o(1). \end{aligned}$$

Therefore, $\lim_{n,T\to\infty} d_T^q (\tilde{J}_{nT}^{DK} - J_{nT}) = -K_q b_2^{(q)}$.

(c) $\sqrt{\frac{T}{\ell_T}}\left(\hat{J}_{nT}^{DK} - J_{nT}\right) = O_p(1)$ **and** $\sqrt{\frac{T}{\ell_T}}\left(\hat{J}_{nT}^{DK} - \tilde{J}_{nT}^{DK}\right) = o_p(1)$

The proof is analogous to the proof of Theorem 1 (c).

(d) Asymptotic MSE

The first equality holds by Theorem 2 (c). For the last equality of Theorem 2(d), since

$$\frac{T}{\ell_T} = \frac{d_T^{2q}}{d_T^{2q} \ell_T / T} = \frac{d_T^{2q}}{\tau + o(1)},$$

we have

$$\begin{aligned} &\lim_{n,T\to\infty} MSE\left(\frac{T}{\ell_T}, \tilde{J}_{nT}^{DK}, S_{nT}\right) \\ &= \lim_{n,T\to\infty} \frac{T}{\ell_T} \text{ vec } \left(E\tilde{J}_{nT}^{DK} - J_{nT}\right)' S_{nT} \text{ vec}\left(E\tilde{J}_{nT}^{DK} - J_{nT}\right) + \lim_{n,T\to\infty} \frac{T}{\ell_T} \bar{\mathcal{K}}_2 \text{tr}\left(S_{nT} \text{var}(\text{vec } \tilde{J}_{nT}^{DK})\right) \\ &= \frac{1}{\tau} K_q^2 \text{ vec}\left(b_2^{(q)}\right)' S \text{ vec}\left(b_2^{(q)}\right) + \bar{\mathcal{K}}_2 tr\left[S(I_{pp} + K_{pp})(J \otimes J)\right], \end{aligned} \tag{A.20}$$

where the last equality holds by Theorem 2 (a) and (b).

Proof of Theorem 4

(a) Asymptotic Variance

Let $\varphi_{lkcd} = \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} r_{(i,t),l}^{(c)} r_{(j,s),k}^{(d)} K\left(\frac{d_{ij}}{d_n}\right)$. As in the proof of Theorem 1 (a), We have

$$\frac{n}{\ell_n} cov\left(\tilde{J}_{nT}^{KP}(c_1, d_1), \tilde{J}_{nT}^{KP}(c_2, d_2)\right) = C_{1nT} + C_{2nT} + C_{3nT},$$

where

$$C_{1nT} = \frac{1}{nT^2\ell_n} \sum_{l=1}^{nTp} \left(E\varepsilon_l^4 - 3\right) \left[\sum_{i,j=1}^{n} \sum_{t,s=1}^{T} r_{(i,t),l}^{(c_1)} r_{(j,s),l}^{(d_1)} K\left(\frac{d_{ij}}{d_n}\right) \sum_{a,b=1}^{n} \sum_{u,v=1}^{T} r_{(a,u),l}^{(c_2)} r_{(b,v),l}^{(d_2)} K\left(\frac{d_{ab}}{d_n}\right)\right],$$

$$C_{2nT} = \frac{1}{nT^2\ell_n} \sum_{l,k=1}^{nTp} \left[\sum_{i,j=1}^{n} \sum_{t,s=1}^{T} r_{(i,t),l}^{(c_1)} r_{(j,s),k}^{(d_1)} K\left(\frac{d_{ij}}{d_n}\right) \sum_{a,b=1}^{n} \sum_{u,v=1}^{T} r_{(a,u),l}^{(c_2)} r_{(b,v),k}^{(d_2)} K\left(\frac{d_{ab}}{d_n}\right)\right],$$

$$C_{3nT} = \frac{1}{nT^2\ell_n} \sum_{l,k=1}^{nTp} \left[\sum_{i,j=1}^{n} \sum_{t,s=1}^{T} r_{(i,t),l}^{(c_1)} r_{(j,s),k}^{(d_1)} K\left(\frac{d_{ij}}{d_n}\right) \sum_{a,b=1}^{n} \sum_{u,v=1}^{T} r_{(a,u),k}^{(c_2)} r_{(b,v),l}^{(d_2)} K\left(\frac{d_{ab}}{d_n}\right)\right].$$

We can show $C_{1nT} = o(1)$ in a similar way to the proof of Theorem 1 (a).

For C_{2nT}, we can decompose this to control the boundary effects as in (A.2) and (A.18).

$$\begin{aligned} C_{2nT} &= \frac{1}{nT^2\ell_n} \sum_{i\in E_n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{a\in E_n} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t,s,u,v=1}^{T} K\left(\frac{d_{ij_{(i)}}}{d_n}\right) K\left(\frac{d_{ab_{(a)}}}{d_n}\right) \gamma_{(it,au)}^{(c_1c_2)} \gamma_{(j_{(i)}s,b_{(a)}v)}^{(d_1d_2)} \\ &+ \frac{1}{nT^2\ell_n} \sum_{i\notin E_n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{a=1}^{n} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t,s,u,v=1}^{T} K\left(\frac{d_{ij_{(i)}}}{d_n}\right) K\left(\frac{d_{ab_{(a)}}}{d_n}\right) \gamma_{(it,au)}^{(c_1c_2)} \gamma_{(j_{(i)}s,b_{(a)}v)}^{(d_1d_2)} \\ &+ \frac{1}{nT^2\ell_n} \sum_{i\in E_n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{a\notin E_n} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t,s,u,v=1}^{T} K\left(\frac{d_{ij_{(i)}}}{d_n}\right) K\left(\frac{d_{ab_{(a)}}}{d_n}\right) \gamma_{(it,au)}^{(c_1c_2)} \gamma_{(j_{(i)}s,b_{(a)}v)}^{(d_1d_2)} \\ &= D_{1nT} + D_{2nT} + D_{3nT} \end{aligned} \tag{A.21}$$

Based on the proofs of (A.3) and (A.4), it is easy to show

$$D_{1nT} \to \bar{\mathcal{K}}_1 J(c_1, c_2) J(d_1, d_2), \tag{A.22}$$

as $n, T \to \infty$.

Based on (A.13), it is straightforward that D_{2nT} and D_{3nT} are $o(1)$ with the sequence of d_n satisfying $n_2/n \to 0$ as $n \to \infty$.

With the same procedure, it is easy to show

$$\lim_{n,T\to\infty} C_{3nT} = \bar{\mathcal{K}}_1 J(c_1, d_2) J(c_2, d_1).$$

As a consequence

$$\lim_{n,T\to\infty} \frac{n}{\ell_n} \mathrm{var}\left(\mathrm{vec}\left(\tilde{J}_{nT}^{KP}\right)\right) = \bar{\mathcal{K}}_1 (I_{pp} + \mathbb{K}_{pp}) (J \otimes J).$$

(b) Asymptotic Bias

The proof is analogous to the proof of Theorem 2 (b).

(c) $\sqrt{\frac{n}{\ell_n}}\left(\hat{J}_{nT}^{KP}-J_{nT}\right)=O_p(1)$ **and** $\sqrt{\frac{n}{\ell_n}}\left(\hat{J}_{nT}^{KP}-\tilde{J}_{nT}^{KP}\right)=o_p(1)$

The proof is analogous to the proof of Theorem 1 (c).

(d) Asymptotic MSE

The proof is analogous to the proof of Theorem 2 (d).

Proof of Proposition 5

(a) $\hat{J}_{nT}-\hat{J}_{nT}^{GA}=o_p(1)$ **if** $d_n\to 0$ **as** $n\to\infty$.

From Theorem 1 (c), $\hat{J}_{nT}-\tilde{J}_{nT}=o_p(1)$ and similarly $\hat{J}_{nT}^{GA}-\tilde{J}_{nT}^{GA}=o_p(1)$. Therefore, it is enough to show that

$$\tilde{J}_{nT}(c,d)-\tilde{J}_{nT}^{GA}(c,d)=o_p(1), \tag{A.23}$$

if $d_n\to 0$ as $n\to\infty$.

Recall $V_{i,t}^{(c)}=\sum_{l=1}^{nTp}r_{(i,t),l}^{(c)}\varepsilon_l$. By Chebyshev's inequality, for any $\Delta>0$,

$$\begin{aligned}
&P\left(\left|\tilde{J}_{nT}(c,d)-\tilde{J}_{nT}^{GA}(c,d)\right|>\Delta\right)\\
&=P\left(\left|\frac{1}{nT}\sum_{i\neq j}\sum_{t,s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)V_{(i,t)}^{(c)}V_{(j,s)}^{(d)}\right|>\Delta\right)\\
&=\frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{i\neq j}\sum_{t,s=1}^{T}\sum_{a\neq b}\sum_{u,v=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ab}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)K\left(\frac{d_{uv}}{d_T}\right)E\left[V_{(i,t)}^{(c)}V_{(j,s)}^{(d)}V_{(a,u)}^{(c)}V_{(b,v)}^{(d)}\right]\\
&=\tilde{C}_{1nT}+\tilde{C}_{2nT}+\tilde{C}_{3nT}+\tilde{C}_{4nT},
\end{aligned}$$

where

$$\tilde{C}_{1nT}=\frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{t,s,u,v=1}^{T}\sum_{l=1}^{nTp}\sum_{i\neq j}\sum_{a\neq b}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ab}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)K\left(\frac{d_{uv}}{d_T}\right)r_{(i,t),l}^{(c)}r_{(j,s),l}^{(d)}r_{(a,u),l}^{(c)}r_{(b,v),l}^{(d)}\left(E\varepsilon_l^4-3\right)$$

$$\tilde{C}_{2nT}=\frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{t,s,u,v=1}^{T}\sum_{i\neq j}\sum_{a\neq b}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ab}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)K\left(\frac{d_{uv}}{d_T}\right)\gamma_{(it,js)}^{(cd)}\gamma_{(au,bv)}^{(cd)}$$

$$\tilde{C}_{3nT}=\frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{t,s,u,v=1}^{T}\sum_{i\neq j}\sum_{a\neq b}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ab}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)K\left(\frac{d_{uv}}{d_T}\right)\gamma_{(it,js)}^{(cd)}\gamma_{(au,bv)}^{(cd)}$$

$$\tilde{C}_{4nT}=\frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{t,s,u,v=1}^{T}\sum_{i\neq j}\sum_{a\neq b}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ab}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)K\left(\frac{d_{uv}}{d_T}\right)\gamma_{(it,js)}^{(cd)}\gamma_{(au,bv)}^{(cd)}.$$

Following (A.1), we can show $\tilde{C}_{1nT}=o(1)$.

For $\tilde{C}_{2nT}$,

$$\tilde{C}_{2nT}\leq\frac{1}{\Delta^2}\left(\frac{1}{nT}\sum_{t,s=1}^{T}\sum_{i\neq j}K\left(\frac{d_{ij}}{d_n}\right)\left|\gamma_{(it,js)}^{(cd)}\right|\right)^2\to 0$$

as $d_n\to 0$ because $K(d_{ij}/d_n)=0$ for all $i\neq j$ provided $d_n<\min_{i,j}d_{ij}$.

With the similar procedures, we can show that $\tilde{C}_{3nT}\to 0$ and $\tilde{C}_{4nT}\to 0$. Therefore, (A.23) holds.

(b) $\hat{J}_{nT} - \hat{J}_{nT}^{DK} = o_p(1)$ **if** $\ell_n^{(c)}/n \to 1$ **as** $n \to \infty$.

From Theorem 3 (c), $\hat{J}_{nT}^{DK} - \tilde{J}_{nT}^{DK} = o_p(1)$. Therefore, it is enough to show that

$$\tilde{J}_{nT}(c,d) - \tilde{J}_{nT}^{DK}(c,d) = o_p(1), \tag{A.24}$$

if $\ell_n^{(c)}/n \to 1$ as $n \to \infty$.

By Chebyshev's inequality, for any Δ,

$$\begin{aligned}
&P\left(\left|\tilde{J}_{nT}(c,d) - \tilde{J}_{nT}^{DK}(c,d)\right| > \Delta\right) \\
&\le \frac{1}{\Delta^2} E\left(\tilde{J}_{nT}(c,d) - \tilde{J}_{nT}^{DK}(c,d)\right)^2 \\
&= \frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{i,j,a,b=1}^{n}\sum_{t,s,u,v=1}^{T}\left(K\left(\frac{d_{ij}}{d_n}\right)-1\right)\left(K\left(\frac{d_{ab}}{d_n}\right)-1\right)K\left(\frac{d_{ts}}{d_T}\right)K\left(\frac{d_{uv}}{d_T}\right)E\left[V_{(i,t)}^{(c)}V_{(j,s)}^{(d)}V_{(a,u)}^{(c)}V_{(b,v)}^{(d)}\right] \\
&= \check{C}_{1nT} + \check{C}_{2nT} + \check{C}_{3nT} + \check{C}_{4nT},
\end{aligned}$$

where

$$\begin{aligned}
\check{C}_{1nT} &= \frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{i,j,a,b=1}^{n}\sum_{t,s,u,v=1}^{T}\sum_{l=1}^{nTp}\left(K\left(\frac{d_{ij}}{d_n}\right)-1\right)\left(K\left(\frac{d_{ab}}{d_n}\right)-1\right)K\left(\frac{d_{ts}}{d_T}\right)K\left(\frac{d_{uv}}{d_T}\right) \\
&\quad \times r_{(i,t),l}^{(c)} r_{(j,s),l}^{(d)} r_{(a,u),l}^{(c)} r_{(b,v),l}^{(d)}\left(E\varepsilon_l^4 - 3\right), \\
\check{C}_{2nT} &= \frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{i,j,a,b=1}^{n}\sum_{t,s,u,v=1}^{T}\left(K\left(\frac{d_{ij}}{d_n}\right)-1\right)\left(K\left(\frac{d_{ab}}{d_n}\right)-1\right)K\left(\frac{d_{ts}}{d_T}\right)K\left(\frac{d_{uv}}{d_T}\right)\gamma_{(it,js)}^{(cd)}\gamma_{(au,bv)}^{(cd)}, \\
\check{C}_{3nT} &= \frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{i,j,a,b=1}^{n}\sum_{t,s,u,v=1}^{T}\left(K\left(\frac{d_{ij}}{d_n}\right)-1\right)\left(K\left(\frac{d_{ab}}{d_n}\right)-1\right)K\left(\frac{d_{ts}}{d_T}\right)K\left(\frac{d_{uv}}{d_T}\right)\gamma_{(it,js)}^{(cd)}\gamma_{(au,bv)}^{(cd)}, \\
\check{C}_{4nT} &= \frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{i,j,a,b=1}^{n}\sum_{t,s,u,v=1}^{T}\left(K\left(\frac{d_{ij}}{d_n}\right)-1\right)\left(K\left(\frac{d_{ab}}{d_n}\right)-1\right)K\left(\frac{d_{ts}}{d_T}\right)K\left(\frac{d_{uv}}{d_T}\right)\gamma_{(it,js)}^{(cd)}\gamma_{(au,bv)}^{(cd)}.
\end{aligned}$$

We can show that $\check{C}_{1nT} = o(1)$ using the procedure in (A.1).

For $\check{C}_{2nT}$,

$$\begin{aligned}
\check{C}_{2nT} &= \frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{i,j,a,b=1}^{n}\sum_{t,s,u,v=1}^{T} 1\left\{\frac{d_{ij}}{d_n} > c\right\} 1\left\{\frac{d_{ab}}{d_n} > c\right\} K\left(\frac{d_{ts}}{d_T}\right)K\left(\frac{d_{uv}}{d_T}\right)\gamma_{(it,js)}^{(cd)}\gamma_{(au,bv)}^{(cd)} \\
&\le \frac{1}{\Delta^2}\left(\frac{1}{nT}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T} 1\left\{\frac{d_{ij}}{d_n} > c\right\} d_{ij}^{-q}\left|\gamma_{(it,js)}^{(cd)}\right| d_{ij}^{q}\right)^2 \\
&\le \left(\frac{1}{c \cdot d_n}\right)^{2q}\frac{1}{\Delta^2}\left(\frac{1}{nT}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}\left|\gamma_{(it,js)}^{(cd)}\right| d_{ij}^{q}\right)^2 \to 0,
\end{aligned}$$

as $d_n \to \infty$.

For $\check{C}_{3nT}$,

$$\check{C}_{3nT} \leq \frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{i,j,a,b=1}^{n}\sum_{t,s,u,v=1}^{T} 1\left\{\frac{d_{ij}}{d_n} > c\right\} 1\left\{\frac{d_{ab}}{d_n} > c\right\} 1\left\{\frac{d_{ia}}{d_n} \leq c\right\} 1\left\{\frac{d_{jb}}{d_n} \leq c\right\} \left|\gamma_{(it,au)}^{(cc)}\gamma_{(js,bv)}^{(dd)}\right|$$

$$+ \frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{i,j,a,b=1}^{n}\sum_{t,s,u,v=1}^{T} 1\left\{\frac{d_{ij}}{d_n} > c\right\} 1\left\{\frac{d_{ab}}{d_n} > c\right\} 1\left\{\frac{d_{ia}}{d_n} > c \text{ or } \frac{d_{jb}}{d_n} > c\right\} \left|\gamma_{(it,au)}^{(cc)}\gamma_{(js,bv)}^{(dd)}\right|$$

$$= \frac{1}{\Delta^2}\frac{1}{n^2T^2}\sum_{i,j,a,b=1}^{n}\sum_{t,s,u,v=1}^{T} 1\left\{\frac{d_{ij}}{d_n} > c\right\} 1\left\{\frac{d_{ab}}{d_n} > c\right\} 1\left\{\frac{d_{ia}}{d_n} \leq c\right\} 1\left\{\frac{d_{jb}}{d_n} \leq c\right\} \left|\gamma_{(it,au)}^{(cc)}\gamma_{(js,bv)}^{(dd)}\right| + o(1)$$

$$= \frac{1}{\Delta^2}\frac{1}{nT}\sum_{i=1}^{n}\sum_{\{a:d_{ia}/d_n\leq c\}}\sum_{t,u=1}^{T}\left|\gamma_{(it,au)}^{(cc)}\right|\left(\frac{1}{nT}\sum_{\{j:d_{ij}/d_n>c\}}\sum_{\{b:d_{jb}/d_n\leq c,\ d_{ab}/d_n>c\}}\sum_{s,v=1}^{T}\left|\gamma_{(js,bv)}^{(dd)}\right|\right) + o\left(1\right).$$

As $\ell_{i,n}^{(c)} \leq C\ell_n^{(c)}$ with some constant C, if $\ell_n^{(c)}/n \to 1$, then

$$\begin{aligned}
&\frac{1}{nT}\sum_{\{j:d_{ij}/d_n>c\}}\sum_{\{b:d_{jb}/d_n\leq c,\ d_{ab}/d_n>c\}}\sum_{s,v=1}^{T}\left|\gamma_{(js,bv)}^{(dd)}\right| \\
&= \frac{n-\ell_n^{(c)}}{n}\frac{1}{\left(n-\ell_n^{(c)}\right)T}\sum_{\{j:d_{ij}/d_n>c\}}\sum_{\{b:d_{jb}/d_n\leq c,\ d_{ab}/d_n>c\}}\sum_{s,v=1}^{T}\left|\gamma_{(js,bv)}^{(dd)}\right| \\
&\to 0,
\end{aligned}$$

which implies $\check{C}_{3nT} \to 0$ as $n, T \to \infty$. With the same procedure, we can show that $\check{C}_{4nT} = o\left(1\right)$. Therefore, (A.24) holds.

(c) $\hat{J}_{nT} - \hat{J}_{nT}^{KP} = o_p(1)$ **if** $\ell_T^{(c)}/T \to 1$ **as** $T \to \infty$.

The proof is analogous to that of (b).

Lemma 2 *Let*

$$X = \frac{1}{\sqrt{L_nM_nT}}\sum_{i_1=1}^{L_n}\sum_{i_2=1}^{M_n}\sum_{t=1}^{T}\Phi_{b,k\ell m}\left(\frac{i_1}{L_n},\frac{i_2}{M_n},\frac{t}{T}\right)\hat{V}_{(i_1,i_2,t)}^{*}.$$

Then, under F1 - F2

$$X \xrightarrow{d} \Lambda\int_0^1\int_0^1\int_0^1\Phi_{b,k\ell m}\left(r_1,r_2,\tau\right)dB_p\left(r_1,r_2,\tau\right).$$

Proof of Lemma 2

Proofs are in the supplementary appendix.

Proof of Proposition 6

Let

$$\check{J}_{nT} = \frac{1}{L_nM_nT}\sum_{i_1=1}^{L_n}\sum_{i_2=1}^{M_n}\sum_{t=1}^{T}\sum_{j_1=1}^{L_n}\sum_{j_2=1}^{M_n}\sum_{s=1}^{T}\sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}\sum_{m=1}^{\infty}\lambda_{k,\ell,m}\Phi_{b,k\ell m}\left(\frac{i_1}{L_n},\frac{i_2}{M_n},\frac{t}{T}\right)$$
$$\times\,\Phi_{b,k\ell m}\left(-\frac{j_1}{L_n},-\frac{j_2}{M_n},-\frac{s}{T}\right)\hat{V}_{(i_1,i_2,t)}^{*}\hat{V}_{(j_1,j_2,s)}^{*\prime}.$$

Then, for any given $\Delta > 0$

$$P(\|\hat{J}_{nT} - \check{J}_{nT}\| \geq \Delta) \leq \frac{1}{\Delta} E\|\hat{J}_{nT} - \check{J}_{nT}\| \to 0, \quad n, T \to \infty$$

because by Assumption I8

$$E\|\hat{V}^*_{(i_1,i_2,t)} \hat{V}^{*\prime}_{(j_1,j_2,s)}\| < \infty$$

and

$$\left| \mathbb{K}_b(x_1 - x_2, y_1 - y_2, z_1 - z_2) - \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \sum_{m=1}^{\infty} \lambda_{k,\ell,m} \Phi_{b,k\ell m}(x_1, y_1, z_1) \Phi_{b,k\ell m}(-x_2, -y_2, -z_2) \right| = 0$$

by the Fourier series representation. This implies

$$\hat{J}_{nT} - \check{J}_{nT} = o_p(1). \tag{A.25}$$

Hence, we can derive the limiting random matrix of $\check{J}_{nT}$ for that of $\hat{J}_{nT}$.

$$\begin{aligned}
\check{J}_{nT} &= \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \sum_{m=1}^{\infty} \lambda_{k,\ell,m} \frac{1}{\sqrt{L_n M_n T}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^{T} \Phi_{b,k\ell m}\left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T}\right) \hat{V}^*_{(i_1,i_2,t)} \\
&\times \frac{1}{\sqrt{L_n M_n T}} \sum_{j_1=1}^{L_n} \sum_{j_2=1}^{M_n} \sum_{s=1}^{T} \left(\Phi_{b,k\ell m}\left(\frac{j_1}{L_n}, \frac{j_2}{M_n}, \frac{s}{T}\right) \hat{V}^*_{(j_1,j_2,s)} \right)^H \\
&:= \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \sum_{m=1}^{\infty} \lambda_{k,\ell,m} X X^H,
\end{aligned}$$

where superscript 'H' denotes the conjugate transpose.

From Lemma 2 and (A.25), we have

$$\begin{aligned}
\hat{J}_{nT} &\xrightarrow{d} \Lambda \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \sum_{m=1}^{\infty} \lambda_{k,\ell,m} \Phi_{b,k\ell m}(r_1, r_2, \tau) \\
&\times \Phi_{b,k\ell m}(-v_1, -v_2, -\kappa) \, dB_p(r_1, r_2, \tau) \, dB_p'(v_1, v_2, \kappa) \, \Lambda' \\
&\overset{d}{=} \Lambda \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \mathbb{K}_b(r_1 - v_1, r_2 - v_2, \tau - \kappa) \, dB_p(r_1, r_2, \tau) \, dB_p'(v_1, v_2, \kappa) \, \Lambda',
\end{aligned}$$

where the equality in distribution holds because

$$\begin{aligned}
&P\Bigg(\Bigg\| \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \Bigg(\sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \sum_{m=1}^{\infty} \lambda_{k,\ell,m} \Phi_{b,k\ell m}(r_1, r_2, \tau) \Phi_{b,k\ell m}(-v_1, -v_2, -\kappa) \\
&- \mathbb{K}_b(r_1 - v_1, r_2 - v_2, \tau - \kappa) \Bigg) dB_p(r_1, r_2, \tau) \, dB_p'(v_1, v_2, \kappa) \Bigg\| \geq \Delta \Bigg) \\
&\leq \int_0^1 \int_0^1 \int_0^1 \left(\sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \sum_{m=1}^{\infty} \lambda_{k,\ell,m} \Phi_{b,k\ell m}(r_1, r_2, \tau) \Phi_{b,k\ell m}(r_1, r_2, \tau) - 1 \right) I_p dr_1 dr_2 d\tau \\
&= 0.
\end{aligned}$$

Lemma 3 *As b_1, b_2 and $b_3 \to 0$, we have*

(a) $E\left(v_{11} - v_{12} v_{22}^{-1} v_{21}\right) = 1 - b_1 b_2 b_3 c_1 - (g-1) b_1 b_2 b_3 c_2 + o(b_1 b_2 b_3)$,

(b) $E\left(v_{11} - v_{12} v_{22}^{-1} v_{21}\right)^2 = 1 - 2 b_1 b_2 b_3 (c_1 + (g-2) c_2) + o(b_1 b_2 b_3)$,

(c) $E\left[\left(v_{11} - v_{12} v_{22}^{-1} v_{21}\right) - 1\right]^2 = 2 b_1 b_2 b_3 c_2 + o(b_1 b_2 b_3)$.

Proof of Lemma 3

This is a direct application of Lemma 3 in Sun (2010).

Proof of Theorem 5

Taking a Taylor expansion, we have

$$
\begin{aligned}
& P\left\{gF_{\infty}(g,b)\le z\right\} \\
& = EG_g\left(z\left(v_{11}-v_{12}v_{22}^{-1}v_{21}\right)\right) \\
& = G_g(z)+G_g'(z)\,zE\left[\left(v_{11}-v_{12}v_{22}^{-1}v_{21}\right)-1\right]+\frac{1}{2}G_g''(z)\,z^2E\left[\left(v_{11}-v_{12}v_{22}^{-1}v_{21}\right)-1\right]^2 \\
& +\frac{1}{2}E\left[G_g''(\tilde{z})-G_g''(z)\right]z^2\left[\left(v_{11}-v_{12}v_{22}^{-1}v_{21}\right)-1\right]^2
\end{aligned}
$$

where $\tilde{z}$ is between z and $z\left(v_{11}-v_{12}v_{22}^{-1}v_{21}\right)$. Using Lemma 3, we have

$$
\begin{aligned}
& P\left\{gF_{\infty}(g,b)\le z\right\} \\
& = G_g(z)-G_g'(z)\,z\left[b_1b_2b_3c_1+(g-1)\,b_1b_2b_3c_2\right]+G_g''(z)\,z^2b_1b_2b_3c_2+o\left(b_1b_2b_3\right) \\
& = G_g(z)+\left[G_g''(z)\,z^2c_2-G_g'(z)\,z\left(c_1+(g-1)\,c_2\right)\right]b_1b_2b_3+o\left(b_1b_2b_3\right) \\
& = G_g(z)+A(z)\,b_1b_2b_3+o\left(b_1b_2b_3\right).
\end{aligned}
$$

Proof of Theorem 6

It follows from Theorem 5 that

$$
\begin{aligned}
& P\left\{F_{\infty}^{*}(g,b)\le z\right\} \\
& = P\left\{gF_{\infty}(g,b)\le gz\left[1+b_1b_2b_3\left(c_1+(g-1)\,c_2\right)\right]\right\} \\
& = G_g\left(gz\left[1+b_1b_2b_3\left(c_1+(g-1)\,c_2\right)\right]\right) \\
& + A\left(gz\left[1+b_1b_2b_3\left(c_1+(g-1)\,c_2\right)\right]\right)b_1b_2b_3+o\left(b_1b_2b_3\right) \\
& = G_g(gz)+G_g'(gz)\,gz\left[c_1+(g-1)\,c_2\right]b_1b_2b_3+A(gz)\,b_1b_2b_3+o\left(b_1b_2b_3\right) \\
& = G_g(gz)+G_p''(gz)\,g^2z^2c_2b_1b_2b_3+o\left(b_1b_2b_3\right).
\end{aligned}
$$

By definition,

$$
\begin{aligned}
P\left\{F_{g,K}\le z\right\} & = P\left\{\chi_g^2\le gz\frac{\chi_K^2}{K}\right\}=EG_g\left(gz\frac{\chi_K^2}{K}\right) \\
& = G_g(gz)+G_g'(gz)gzE\left(\frac{\chi_K^2}{K}-1\right)+\frac{1}{2}G_g''(gz)\left(\frac{gz}{K}\right)^2E\left(\chi_K^2-K\right)^2+o\left(\frac{1}{K}\right) \\
& = G_g(gz)+\frac{1}{K}G_g''(gz)g^2z^2+o\left(\frac{1}{K}\right) \\
& = G_g(gz)+G_g''(gz)\,g^2z^2c_2b_1b_2b_3+o\left(b_1b_2b_3\right).
\end{aligned}
$$

Hence

$$
P\left\{F_{\infty}^{*}(g,b)\le z\right\}=P\left\{F_{g,K}\le z\right\}+o\left(b_1b_2b_3\right).
$$

References

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Anselin, L. (2001). Spatial econometrics. In: Baltagi, Badi H. (Ed.), A companion to theoretical econometrics.

Arellano, M. (1987). Computing Robust Standard Errors for Within Group Estimators. *Oxford bulletin of Economics and Statistics*, 49:431–434.

Arellano, M. and Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. *ECONOMETRIC SOCIETY MONOGRAPHS*, 43:381.

Barrios, T., Diamond, R., Imbens, G., and Kolesar, M. (2010). Clustering, Spatial Correlations and Randomization Inference. *NBER Working Paper.*

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates?*. *Quarterly Journal of Economics*, 119(1):249–275.

Bester, A., Conley, T., Hansen, C., and Vogelsang, T. (2008). Fixed-b asymptotics for spatially dependent robust nonparametric covariance matrix estimators. *Working paper, Michigan State University.*

Bester, C., Conley, T., and Hansen, C. (2009). Inference with dependent data using cluster covariance estimators. *Working paper, University of Chicago.*

Cameron, A., Gelbach, J., and Miller, D. (2009). Robust Inference with Multi-way Clustering. *Working Papers, UC Davis.*

Chen, X. and Conley, T. (2001). A new semiparametric spatial model for panel time series. *Journal of Econometrics*, 105(1):59–83.

Conley, T. (1999). GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics*, 92(1):1–45.

Conley, T. and Dupor, B. (2003). A spatial analysis of sectoral complementarity. *Journal of Political Economy*, 111(2):311–352.

Conley, T. and Ligon, E. (2002). Economic distance and cross-country spillovers. *Journal of Economic Growth*, 7(2):157–187.

Domowitz, I. and White, H. (1982). Misspecified models with dependent observations. *Journal of Econometrics*, 20(1):35–58.

Driscoll, J. and Kraay, A. (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics*, 80(4):549–560.

Fama, E. and MacBeth, J. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3).

Gonçalves, S. (2008). The moving blocks bootstrap for panel linear regression models with individual fixed effects. *Working paper, Universite de Montreal.*

Hannan, E. (1970). *Multiple Time Series.* Wiley New York.

Hansen, C. (2007). Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *Journal of Econometrics*, 141(2):597–620.

Ibragimov, R. and Müller, U. (2010). t-Statistic Based Correlation and Heterogeneity Robust Inference. *Journal of Business and Economic Statistics*, (0):1–16.

Jansson, M. (2004). The error in rejection probability of simple autocorrelation robust tests. *Econometrica*, 72(3):937–946.

Kelejian, H. and Prucha, I. (2007). HAC Estimation in a Spatial Framework. *Journal of Econometrics*, 140(1):131–154.

Kézdi, G. Robust Standard Error Estimation In Fixed-Effects Panel Models. *Mimeo.*

Kiefer, N. and Vogelsang, T. (2002a). Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size. *Econometric Theory*, 18(06):1350–1366.

Kiefer, N. and Vogelsang, T. (2002b). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica*, 70(5):2093–2095.

Kiefer, N. and Vogelsang, T. (2005). A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests. *Econometric Theory*, 21(06):1130–1164.

Kiefer, N., Vogelsang, T., and Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, 68(3):695–714.

Kim, M. and Sun, Y. (2010). Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Working paper, UCSD.*

Kinal, T. (1980). The existence of moments of k-class estimators. *Econometrica*, 48(1):241–249.

Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13.

Lin, C. and Sakata, S. (2009). On Long-Run Covariance Matrix Estimation with the Truncated Flat Kernel. *Working paper, University of British Columbia.*

Mariano, R. (1972). The existence of moments of the ordinary least squares and two-stage least squares estimators. *Econometrica*, 40(4):643–652.

Newey, W. and West, K. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55:703–708.

Newey, W. and West, K. (1994). Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, 61:631–631.

Neyman, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, 16(1):1–32.

Percival, D. and Walden, A. (1993). *Spectral analysis for physical applications: multitaper and conventional univariate techniques.* Cambridge Univ Press.

Petersen, M. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, 22(1):435.

Phillips, P. (2005). HAC estimation by automated regression. *Econometric Theory*, 21(01):116–142.

Politis, D. (2001). On nonparametric function estimation with infinite-order flat-top kernels. *Probability and Statistical Models with applications*, pages 469–483.

Politis, D. (2010). Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econometric Theory.*

Priestley, M. (1981). *Spectral analysis and time series.* Academic Press.

Sun, Y. (2010). Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference. *Working paper, USCD.*

Sun, Y. and Kaplan, D. (2010). A New Asymptotic Theory for Vector Autoregressive Long-run Variance Estimation and Autocorrelation Robust Testing. *Working paper, USCD.*

Sun, Y. and Kim, M. (2010). Asymptotic F Test in the Presence of Nonparametric Spatial Dependence. *Working paper, USCD.*

Sun, Y. and Phillips, P. (2009). Bandwidth Choice for Interval Estimation in GMM Regression. *Working paper, USCD.*

Sun, Y., Phillips, P., and Jin, S. (2008). Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing. *Econometrica*, 76(1):175–194.

Thompson, S. (2006). Simple formulas for standard errors that cluster by both firm and time. *Working paper, Harvard University.*

Vogelsang, T. (2008). Heteroskedasticity, Autocorrelation, and Spatial Correlation Robust Inference in Linear Panel Models with Fixed-Effects. *Working paper, Michigan State University.*

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 48(4):817–838.

White, H. (1984). Asymptotic theory for econometricians.

Wooldridge, J. (2003). Cluster-sample methods in applied econometrics. *American Economic Review*, 93(2):133–138.

Yu, J., de Jong, R., and Lee, L. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, 146(1):118–134.