Heteroskedasticity and Spatiotemporal Dependence Robust Inference for Linear Panel Models with Fixed Effects*

Min Seong Kim
Department of Economics
Ryerson University

Yixiao Sun
Department of Economics
UC San Diego

August 2011

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 that is flexible to nest existing estimators as special cases with certain choices of bandwidths. For distributional approximations, we embed the level of smoothing and the sample size in two different limiting sequences. In the first case where the level of smoothing increases with the sample size, the proposed covariance 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. In the second case where the level of smoothing is fixed, the covariance estimator has a random limit and we show by asymptotic expansion that the limiting distribution of the Wald 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 a convenient F-approximation to this distribution. For bandwidth selection, we employ and optimize a modified asymptotic mean square error criterion. The flexibility of our estimator and the proposed bandwidth selection procedure make our estimator adaptive to the dependence structure. This *adaptiveness* effectively automates the selection of covariance estimators. Simulation results show that our proposed testing procedure works reasonably well in finite samples.

Keywords: Adaptiveness, HAC estimator, F-approximation, Fixed-smoothing asymptotics, Increasing-smoothing asymptotics, Panel data, Optimal bandwidth, Robust inference, Spatiotemporal dependence

JEL Classification Number: C13, C14, C23

*Email: minseong.kim@ryerson.ca and yisun@ucsd.edu. We thank Brendan Beare, Alan Bester, Otilia Boldea, Pierre-Andre Chiappori, Gordon Dahl, Feico Drost, Graham Elliott, Patrik Guggenberger, Jinyong Hahn, James Hamilton, Christian Hansen, Hiroaki Kaido, Ivana Komunjer, Esfandiar Maasoumi, Jan Magnus, Bertrand Melenberg, Leo Michelis, Choon-Geol Moon, Philip Neary, Elena Pesavento, Ingmar Prucha, Andres Santos, Halbert White and the participants of Panel Data Conference at Amsterdam, Econometric Society World Congress at Shanghai and seminars at Tilburg University, UC San Diego, Emory University, UC Davis, Chicago Booth, Ryerson University, University of Maryland, KIPF and Hanyang University. Sun gratefully acknowledges partial research support from NSF under Grant No. SES-0752443.

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, robust inference in the panel setting is an 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 *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 an observable distance measure d_{ij} between them. 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), Bester, Conley, Hansen and Vogelsang (2008, BCHV hereafter) and Kim and Sun (2011, KS hereafter).

There are several robust covariance estimators with correlated panel data. Arellano (1987) proposes the clustered covariance estimator (CCE) by extending the White standard error (White, 1980) to account for serial correlation. Wooldridge (2003) provides a concise review on the CCE. Driscoll and Kraay (1998, DK hereafter) suggest a different approach that uses a time series HAC estimator (e.g. Newey and West, 1987) applied to cross-sectional averages of moment conditions. Gonçalves (2010) examines the properties of this estimator in linear panel models with fixed effects. Another approach considered in this paper is an extension of the spatial HAC estimator applied to time series averages of moment conditions, which we name the KS estimator. This is symmetric to the DK estimator. Conley (1999) is among the first to propose the spatial HAC estimator. Kelejian and Prucha (2007) argue that it can be extended to the panel setting with fixed T.

Our estimator includes these existing estimators as special cases, reducing to each of them with certain bandwidth choice. We refer to this as *flexibility*. If the sequence of the bandwidth in the spatial dimension, d_n, increases at a fast enough rate with the cross sectional sample size n, then our estimator with the rectangular kernel is asymptotically equivalent to the DK estimator. Similarly, if the sequence of the bandwidth in the time dimension, d_T, increases fast enough relative to the time series sample size T, then our estimator with the rectangular kernel is asymptotically equivalent to the KS estimator. On the other hand, if d_n is assumed to approach zero, our estimator reduces to a generalized CCE defined later in the paper.

For distributional approximations, we consider two types of asymptotics: the increasing-smoothing asymptotics and the fixed-smoothing asymptotics. The difference lies in whether the level of smoothing increases or stays fixed as the sample size increases. 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. We also define $\ell_{t,T}$ and ℓ_T in the same way along the time dimension. If $d_n, d_T \to \infty$ as $n, T \to \infty$ but slowly so that $nT/(\ell_n \ell_T) \to \infty$, then the level of smoothing increases with the sample size. Under this increasing-smoothing asymptotics, our covariance estimator is consistent and the limiting distribution of the associated Wald statistic is a χ^2 distribution.

The alternative estimators are also consistent under the increasing-smoothing asymptotics, but each estimator has an important limitation in practice. The performance of the CCE heavily depends on spatial correlation. While this estimator is quite efficient in the presence of spatial independence, even moderate spatial correlation may lead to substantial bias and hence size distortion in statistical testing. Though spatial independence is sometimes assumed for convenience, it may not hold due to, for example, spill-over effects, competition and so on.[1] Collapsing spatial dependence by the cross-sectional averaging, the DK estimator is robust to arbitrary forms 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. Similarly, the KS estimator is not efficient, as 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 KS 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 KS estimator if $n = o(T^4)$. For second-order kernels, the conditions become much weaker, i.e. $T = o(n^{5/2})$ and $n = o(T^6)$, respectively.

If we embed the bandwidth parameters d_n and d_T in a sequence such that $nT/(\ell_n \ell_T)$ holds fixed as n and T increase, then the level of smoothing is fixed with the sample size. Under this fixed-smoothing asymptotics, the covariance estimator converges in distribution to a random matrix and the limiting distribution of the Wald statistic is nonstandard but pivotal. The fixed-smoothing asymptotic approximation is first 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 the '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 than the χ^2 approximation. Jansson (2004), Sun, Phillips and Jin (2008), and Sun and Phillips (2009) provide theoretical explanations in different time series settings.

We adopt the fixed-smoothing asymptotics in the panel setting with our covariance estimator. Using asymptotic expansions we show that the deviation of this limiting distribution from the χ^2 distribution depends on the smoothing parameters, the kernel function, and the number of restrictions being tested. We can accommodate the estimation uncertainty of the parameter estimation and the randomness of the covariance estimator under the 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 the fixed-smoothing asymptotics, the covariance estimator converges in distribution to an infinite weighted sum of independent Wishart distributions. We approximate this by 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

[1] Recently, Bester, Conley and Hansen (2010) present consistency results for the CCE with spatially dependent data by constructing clusters to be asymptotically independent. In this paper, we consider a rather traditional panel CCE for which the cluster is defined based on each individual so that the asymptotic independence condition is not valid. Cameron, Gelbach, and Miller (2006) 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.

based on the generalized CCE and the DK estimator. Besides the kernel methods, Hansen (2007) and Bester, Conley and Hansen (2011) 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 the fixed-smoothing asymptotics for the Fama and MacBeth (1973) type procedure by fixing the number of groups. Sun and Kim (2010) consider a testing procedure using a series-type covariance estimator in the spatial setting. They show that, when the number of basis functions is held fixed, their series covariance estimator converges in distribution to a Wishart distribution, and that the scaled Wald statistic converges to an F distribution. Our F-approximation is motivated from the series method of Sun and Kim (2010). While for the other two 'non-kernel' methods critical values are readily available from the standard t or F distribution, critical values for the kernel methods by BCHV and Vogelsang (2008) have to be simulated. From this point of view, this paper fills the gap in the literature, providing an F-approximation for the kernel method in the panel setting.

In this paper, we select the bandwidth parameters to minimize an upper bound of the asymptotic mean square error (called AMSE*) of the covariance estimator. The AMSE* criterion has a minimax flavor. Though it is standard practice to use the asymptotic mean square error (AMSE) criterion in the HAC estimation literature (e.g. Andrews, 1991 and Newey and West, 1994), it is not tractable for our bivariate kernel estimator. Our AMSE* 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 opposite directions. We suggest a parametric plug-in procedure for practical implementation using the spatiotemporal models in Anselin (2001).

Our bandwidth selection procedure does not apply directly to the rectangular kernel estimator and, more broadly, flat-top kernel estimators. However, it is interesting to consider flat-top kernel estimators because they are higher-order accurate (Politis, 2011). This is particularly important in our setting because the rectangular-kernel-based covariance estimator is more flexible in that it can approach each of the existing estimators with appropriate bandwidth choice. We modify our bandwidth selection procedure to be applicable to the rectangular kernel. The rectangular kernel, combined with our modified bandwidth selection procedure, delivers a covariance estimator with better asymptotic properties than the covariance estimators based on second-order kernels.

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

The remainder of the paper is as follows. Section 2 introduces the panel model, the covariance estimator and hypothesis testing we consider. In Section 3, we examine the properties of our estimator and the associated test statistic under the 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 develop the limiting theory for our covariance estimator and the associated test statistic under the fixed-smoothing asymptotics. We also prove the validity of an F-approximation to the Wald statistic. Section 8 reports simulation evidence. The last section concludes. Proofs are given in the appendix or a supplementary appendix.

2 Panel model, covariance estimator and hypothesis testing

In this paper, we consider a static 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} can be consistently estimated with its sample analog, our central interest is on J_{nT}. Letting $V_{(i,t)} = \tilde{X}_{it}u_{it}$, J_{nT} can be rewritten as

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

We propose a bivariate kernel covariance estimator given by

$$\hat{J}_{nT} = \frac{1}{nT}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\hat{V}_{(i,t)}\hat{V}_{(j,s)}', \tag{3}$$

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, 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).

[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 that the fixed-effects estimators of nonlinear and dynamic panel models usually suffer from.

[3] For simplicity of our analysis, we employ a product kernel with the same kernel function in each dimension.

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

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

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:

$$V_{(i,t)} = \tilde{R}_{(i,t)}\tilde{\varepsilon}, \tag{4}$$

where

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

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 KS, we assume that

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

and

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

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 $\varepsilon := \left(\varepsilon_1, \ldots, \varepsilon_l, \ldots, \varepsilon_{nTp}\right)' = \left(\Sigma^{-1/2} \otimes I_{nT}\right)\tilde{\varepsilon}$. Then,

$$V_{(i,t)} = R_{(i,t)}\varepsilon \text{ and } \text{var}\left(\varepsilon\right) = I_{nTp}. \tag{5}$$

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

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

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 the 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

$$\Gamma_{(it,js)} := \left(\gamma^{(cd)}_{(it,js)}\right) = E\left[V_{(i,t)}V'_{(j,s)}\right] = R_{(i,t)}R'_{(j,s)}, \tag{6}$$

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

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

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

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

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

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

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

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 $|\sigma^{cd}| \leq c_\sigma$ for a constant $c_\sigma > 0$, then Assumption I2 holds if $\sum_{i=1}^{n}\sum_{t=1}^{T}|\tilde{r}^{(d)}_{(it,j,s)}| < c_R/c_\sigma$. Since $|\tilde{r}^{(d)}_{(it,j,s)}|$ can be regarded as the (absolute) change of $V^{(d)}_{(i,t)}$ in response to one unit change in $\tilde{\varepsilon}^{(d)}_{(j,s)}$, the summability condition requires that the aggregate response to an innovation be finite. Assumption

I3 implies that $\Gamma_{(it,js)}$ decays to zero fast enough as d_{ij} and d_{ts} increase so that the two summability conditions hold. These conditions hold if

$$\frac{1}{nT}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}\left|\sum_{a=1}^{n}\sum_{b=1}^{T}\tilde{r}_{(it,a,b)}^{(c)}\tilde{r}_{(js,a,b)}^{(d)}\right|d_{ij}^{q_1}<\infty, \tag{7}$$

$$\frac{1}{nT}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}\left|\sum_{a=1}^{n}\sum_{b=1}^{T}\tilde{r}_{(it,a,b)}^{(c)}\tilde{r}_{(js,a,b)}^{(d)}\right|d_{ts}^{q_2}<\infty \tag{8}$$

for all c and d. (7) and (8) 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 temporally 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 a distance measure 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 does not affect the asymptotic properties of our estimator from a perspective of the measurement error problem as explained below.

Distance measures observable 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 KS, we can show that our asymptotic results are still valid under the following conditions: (i) the measurement error is independent of ε_l for all l; (ii) it is of order $o(d_n)$ as d_n increases; and (iii) the first summability condition in Assumption I3 holds with the 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 areas. To be symmetric, we also define

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

[4] In Conley and Ligon (2000), the asymmetric transportation costs are replaced by the minimum cost between two countries.

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 that the size of a boundary set depends on choice of the bandwidth parameters. 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 $\ell_T/T = o(1)$.

3.2 Increasing-smoothing asymptotics

We present the consistency, the rate of convergence, and the AMSE of the covariance estimator $\hat{J}_{nT}$ and the limiting distribution of the Wald statistic W_{nT} under the increasing-smoothing asymptotics. We begin by introducing the assumption on the kernel used in the covariance estimator.

Assumption I5 *(i) The kernel function $K(\cdot)$ satisfies $K(0) = 1$, $|K(x)| \leq 1$, $K(x) = K(-x)$, $K(x) = 0$ for $|x| \geq 1$. (ii) For all $x_1, x_2 \in \mathbb{R}$ 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 generalize our results to include the QS kernel but this requires much longer proofs. 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

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

for $\bar{\mathcal{K}}_1$. Similarly 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, the limit value of J_{nT}:

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

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

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

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 (2011). They assume that the covariance matrix in each cluster 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 $q \leq q_i$ with $i = 1, 2$, we define

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

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

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,

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

Under the assumptions above, the effect of using $\hat{\beta}$ instead of β_0 on the asymptotic property is $o_p(1)$ as shown by Theorem 1(c) below. Therefore, we can 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 $\mathbb{K}_{pp}$ denote 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, $d_n,d_T\to\infty$, $n_2=o(n)$, $T_2=o(T)$, $\ell_n=o(n)$ and $\ell_T=o(T)$.*

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

(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,*

$$\begin{aligned}&\lim_{n,T\to\infty}MSE\left(\frac{nT}{\ell_n\ell_T},\hat{J}_{nT},S_{nT}\right)\\&=\lim_{n,T\to\infty}MSE\left(\frac{nT}{\ell_n\ell_T},\tilde{J}_{nT},S\right)\\&=\frac{1}{\tau}K_q^2vec\left(b_1^{(q)}+c_d^{-q}b_2^{(q)}\right)'Svec\left(b_1^{(q)}+c_d^{-q}b_2^{(q)}\right)+\bar{\mathcal{K}}_1\bar{\mathcal{K}}_2tr\left(S(I+\mathbb{K}_{pp})(J\otimes J)\right).\end{aligned}$$

[5]Instead of Assumption I8, we can consider asymptotic truncated MSE as Andrews (1991) and Kim and Sun (2011).

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 increase d_n and/or d_T, the asymptotic bias decreases while the asymptotic variance increases. 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 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, the rate of convergence of $\hat{J}_{nT}$ is $\sqrt{nT/\left(\ell_n \ell_T\right)}$. If we set $\ell_n = O(d_n^{\eta_n})$ and $\ell_T = O(d_T^{\eta_T})$ for some $\eta_n > 0$ and $\eta_T = 1$, then the rate of convergence under the rate condition $d_n^{2q}\ell_n\ell_T/\left(nT\right) \to \tau \in (0,\infty)$ is $(nT)^{-q/(2q+\eta_n+\eta_T)}$.

As $\hat{J}_{nT}$ is consistent, the limiting distribution of the Wald statistic is the 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 that minimizes an 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

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

where the inequality holds by the Cauchy inequality. AMSE* can be regarded as AMSE in the worst case:

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

where

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

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

$$(d_n^\star, d_T^\star) = \arg\min_{d_n, d_T} 2K_q^2\left(\frac{B_{11}}{d_n^{2q}} + \frac{B_{22}}{d_T^{2q}}\right) + \frac{\ell_n\ell_T}{nT}\bar{\mathcal{K}}_1\bar{\mathcal{K}}_2 C, \tag{9}$$

where $C = \text{tr}\left[S_{nT}(I_{pp} + \mathbb{K}_{pp})(J\otimes J)\right]$.

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, e.g. Andrews (1991) and Newey and

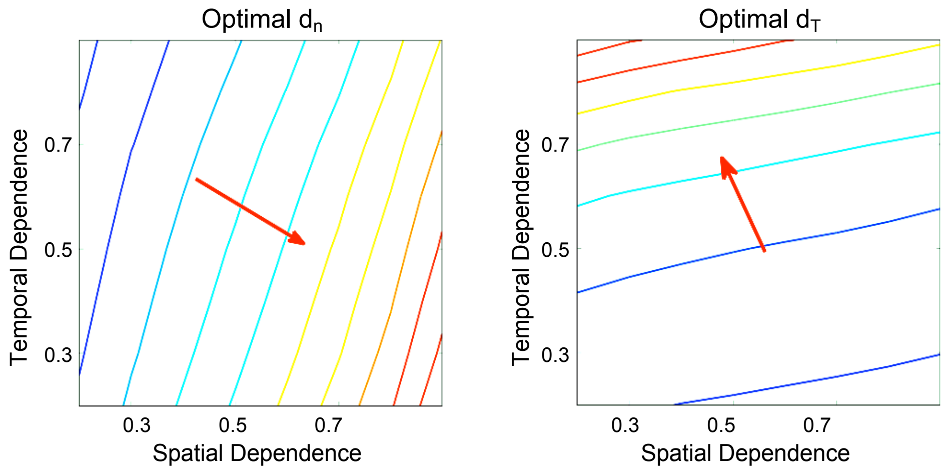

Figure 1 – Level curves of $d_n^\star$ and $d_T^\star$ as functions of spatial and temporal dependences

West (1994). In our setting, though, it is intractable. The source of the problem is that B_{12} can be negative. In theory, we may choose d_n and d_T to zero out the bias terms under some conditions. For example, consider the case $B_{12} = -\sqrt{B_{11}B_{22}}$. This may occur when we are interested in a single component of β. In this case, bandwidth parameters 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, we need to select d_n or d_T to tradeoff the second-order bias with the variance. However, this choice is infeasible in practice. As B_{11}/B_{22} is unknown, we have to estimate this ratio and the estimation error is of the same order as the first order bias. So the first order bias cannot be reduced by an order of magnitude in practice. Our minimax criterion avoids this problem. It is also simple to implement, as $d_n^\star$ and $d_T^\star$ depend only on two bias terms but not on their interaction B_{12}. It also effectively controls for the AMSE in terms of an upper bound, which is achievable under some data generating processes.

Under the boundary condition in the time dimension, we have $\ell_T/T \to 0$, $\ell_T = 2d_T + o(d_T)$. In some cases, it is also possible to approximate ℓ_n as a function of d_n. For example, if individuals are located 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:

$$d_n^\star = \left(\frac{4qK_q^2 B_{11}}{\eta_n \alpha_n \alpha_T \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C} nT \right)^{1/(2q+\eta_n+\eta_T)} \left(\frac{\eta_T B_{11}}{\eta_n B_{22}} \right)^{\eta_T/[2q(2q+\eta_n+\eta_T)]}, \tag{10}$$

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

The optimal bandwidth formulae in (10) and (11) show that the degree of persistence in one dimension affects both $d_n^\star$ and $d_T^\star$ but in opposite directions. 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 truncation in the spatial domain. But, the increase of $d_n^\star$, at the same time, magnifies the variance term. Therefore, in order to minimize the 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 temporal and spatial 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.

The corollary below gives a precise sense that $(d_n^\star, d_T^\star)$ is optimal.

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 $d_n^{2q}\ell_n\ell_T/(nT) \to \tau \in (0,\infty)$, $\{d_n^\star, d_T^\star\}$ is preferred in the sense that*

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

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 estimator, and more broadly, flat-top kernel estimators because their asymptotic bias is of smaller order than that in Theorem 1(b). However, it is interesting to consider flat-top kernel estimators because they are higher-order accurate. 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. In time series HAC estimation, Andrews (1991, footnote on p. 834) and Lin and Sakata (2009) suggest a practical bandwidth 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. We extend these methods to the present setting. For any finite-order kernel estimator set as the target, we can select the bandwidth parameters for the rectangular kernel $(d_{rec,n}^\star, d_{rec,T}^\star)$ such that the rectangular-kernel-based covariance estimator has a smaller AMSE*.

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

$$d_{rec,n}^\star = d_{tar,n}^\star\left(\frac{\bar{\mathcal{K}}_{tar,1}}{\bar{\mathcal{K}}_{rec,1}}\right)^{1/\eta_n} \text{ and } d_{rec,T}^\star = d_{tar,T}^\star\left(\frac{\bar{\mathcal{K}}_{tar,2}}{\bar{\mathcal{K}}_{rec,2}}\right)^{1/\eta_T}, \tag{12}$$

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 smoothness conditions, the asymptotic bias of the rectangular-kernel estimator is of smaller order. As a result, the rectangular kernel estimator has smaller AMSE* than that based on the target kernel.

The unknown values such as B_{11}, B_{22} and C in the optimal bandwidth formula (9) can be estimated in a parametric (e.g. Andrews, 1991; and Kim and Sun, 2011) 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):

$$V_{(i,t)}^{(c)} = \rho_c \left[W_n^{(c)} V_{t-1}^{(c)}\right]_i + \tilde{\varepsilon}_{(i,t)}^{(c)}, \tag{13}$$

$$V_{(i,t)}^{(c)} = \lambda_c V_{(i,t-1)}^{(c)} + \rho_c \left[W_n^{(c)} V_{t-1}^{(c)}\right]_i + \tilde{\varepsilon}_{(i,t)}^{(c)} \tag{14}$$

$$V_{(i,t)}^{(c)} = \lambda_c V_{(i,t-1)}^{(c)} + \phi_c \left[W_n^{(c)} V_{t}^{(c)}\right]_i + \tilde{\varepsilon}_{(i,t)}^{(c)} \tag{15}$$

$$V_{(i,t)}^{(c)} = \lambda_c V_{(i,t-1)}^{(c)} + \phi_c \left[W_n^{(c)} V_{t}^{(c)}\right]_i + \rho_c \left[W_n^{(c)} V_{t-1}^{(c)}\right]_i + \tilde{\varepsilon}_{(i,t)}^{(c)} \tag{16}$$

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, consider the model in (13). It can be rewritten recursively 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

$$\hat{J}(c,d) = \frac{\hat{\sigma}_{cd}}{nT} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} \left(\hat{\tilde{R}}_{(i,t)}^{(c)}\right) \left(\hat{\tilde{R}}_{(j,s)}^{(d)}\right)', \tag{17}$$

$$\hat{b}_1^{(q)}(c,d) = \frac{\hat{\sigma}_{cd}}{nT} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} \left(\hat{\tilde{R}}_{(i,t)}^{(c)}\right) \left(\hat{\tilde{R}}_{(j,s)}^{(d)}\right)' d_{ij}^q, \tag{18}$$

$$\hat{b}_2^{(q)}(c,d) = \frac{\hat{\sigma}_{cd}}{nT} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} \left(\hat{\tilde{R}}_{(i,t)}^{(c)}\right) \left(\hat{\tilde{R}}_{(j,s)}^{(d)}\right)' d_{ts}^q, \tag{19}$$

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 (9) for the true parameters, we obtain the data-driven bandwidth parameters, $(\hat{d}_n, \hat{d}_T)$ as follows:

$$\left(\hat{d}_n, \hat{d}_T\right) = \underset{d_n, d_T}{\arg\min} 2K_q^2 \left(\frac{\hat{B}_{11}}{d_n^{2q}} + \frac{\hat{B}_{22}}{d_T^{2q}}\right) + \frac{\ell_n \ell_T}{nT}\hat{C}, \tag{20}$$

where

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

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

$$\hat{d}_n = \left(\frac{4qK_q^2 \hat{B}_{11}}{\eta_n \alpha_n \alpha_T \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 \hat{C}} nT\right)^{1/(2q+\eta_n+\eta_T)} \left(\frac{\eta_T \hat{B}_{11}}{\eta_n \hat{B}_{22}}\right)^{\eta_T/[2q(2q+\eta_n+\eta_T)]}, \tag{21}$$

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

It also follows

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

Since the models in (14), (15) and (16) 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 (13). While the OLS estimator is consistent for (14), it is not for (15) and (16) due to the endogeneity of $[W_n^{(c)} V_t^{(c)}]_i$. For these models, we can obtain 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 details. However, we argue that the simple OLS can still be used for (15) and (16). 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 KS estimators

For comparison, we examine the asymptotic properties of the CCE, DK and KS estimators based on our data representation in (4) and (5) under the increasing-smoothing asymptotics. We also derive the optimal bandwidth parameters for DK and KS estimators using the AMSE criterion.

5.1 CCE

The CCE is defined as

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

Define $\tilde{J}_{nT}^A$ in the same way but with $\hat{V}_{(i,t)}$ replaced by $V_{(i,t)}$. The crucial condition for $\hat{J}_{nT}^A$ to be consistent is that covariates for two different individuals (or clusters) are 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 spatiotemporal representation accommodates spatial independence by imposing the following restriction.

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

Under Assumption I10,

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

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

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

Similarly, we define $\tilde{J}_{nT}^{DK}$ as above but with $\hat{V}_{(i,t)}$ replaced by $V_{(i,t)}$.

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

Assumption I12 *There exists $q_2 \geq q$ such that*

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

for all n, T.

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

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

Compared with Assumption I3, Assumption I12 is sufficient for $\hat{J}_{nT}^{DK}$ because it does not suffer from the bias due to kernel downweighing in the spatial dimension. Theorem 3 below gives 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,*

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

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, the rate of convergence of $\hat{J}_{nT}$ is faster than that of $\hat{J}_{nT}^{DK}$ if $T = o(n^{(2q+\eta_T)/\eta_n})$.

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

$$d_T^{DK} = \left(\frac{2qK_q^2 B_{22}}{\eta_T \alpha_T \bar{\mathcal{K}}_2 C} T \right)^{1/(2q+\eta_T)}, \tag{23}$$

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

5.3 KS estimator

Analogous to the DK estimator, we can also consider the usage of spatial HAC estimation applied to time series averages, especially when n is large. The KS estimator based on the time series averages is

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

Let $\tilde{J}_{nT}^{KS}$ denote the infeasible version of $\hat{J}_{nT}^{KS}$ with $\hat{V}_{(i,t)}$ replaced by $V_{(i,t)}$.

Assumption I14 *There exists $q_1 \geq q$ such that*

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

for all n, T.

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

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

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

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

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

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

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

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

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

$$d_n^{KS} = \left(\frac{2qK_q^2 B_{11}}{\eta_n \alpha_n \bar{\mathcal{K}}_1 C} n\right)^{1/(2q+\eta_n)}. \tag{24}$$

We can obtain the data-driven bandwidth parameter following KS.

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 in large samples with certain choice of the bandwidths and kernel function. In order to illustrate the flexibility, we first introduce the generalized CCE, $\hat{J}_{nT}^{GA}$:

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

where $K_{RE}(x) = 1\{|x| \leq 1\}$ is the rectangular kernel function.

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 *For $\hat{J}_{nT}$ 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 $\hat{J}_{nT} - \hat{J}_{nT}^{KS} = o_p(1)$.*

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

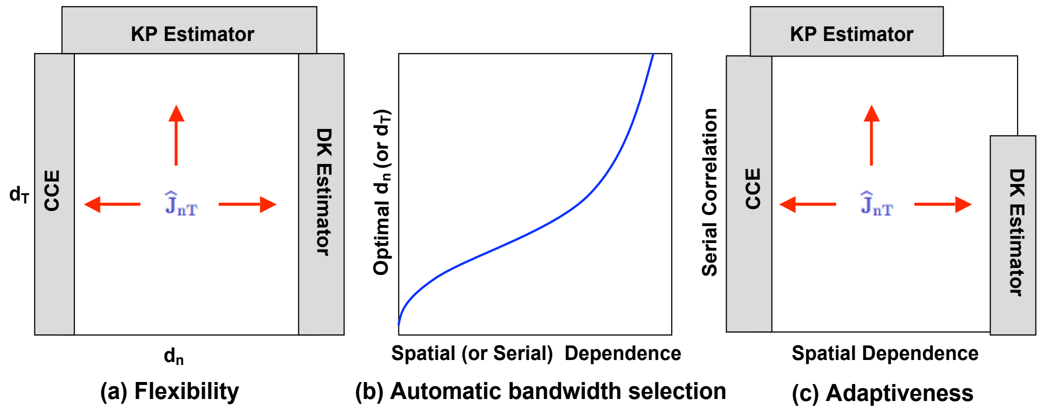

Figure 2 – Adaptiveness of $\hat{J}_{nT}$

6.2 Adaptiveness

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

The attractiveness 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 estimators. 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 $\hat{J}_{nT}^{KS}$ if a process is very persistent in the time dimension. In the absence of spatial dependence, $\hat{J}_{nT}$ becomes 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 the existing estimators. We can fix this by employing the rectangular kernel. In this case, with appropriate bandwidth choices, $\hat{J}_{nT}$ is asymptotically equivalent to the best estimator. The bandwidth selection rule in (12) meets the requirement, as the selected bandwidths from (12) are proportional to those from (9).[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

[6] Another issue with rectangular kernel estimators is that they are not positive semi-definite. Politis (2011) 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 (2011).

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 conditions that define $\hat{\beta}$ 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. 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*

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

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

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

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

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

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

for some positive definite matrix Q.

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

$$\sqrt{nT}\left(\hat{\beta} - \beta\right) = \left(\frac{1}{L_n M_n T} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^{T} \tilde{X}^*_{(i_1,i_2,t)} \tilde{X}^{*\prime}_{(i_1,i_2,t)}\right)^{-1} \frac{1}{\sqrt{L_n M_n T}} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^{T} V^*_{(i_1,i_2,t)}$$
$$\xrightarrow{d} Q\Lambda \mathcal{W}_p(1,1,1) := \Lambda^* \mathcal{W}_p(1,1,1). \tag{25}$$

Therefore,

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

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 the distance between the two units located 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.

Let $b_1 = d_n/L_n$, $b_2 = d_n/M_n$ and $b_3 = d_T/T$. Suppose that the level of smoothing is held fixed such that b_1, b_2 and b_3 are fixed constants. Under Assumption F3, we have

$$\hat{J}_{nT} := \frac{1}{L_n M_n T} \sum_{i_1,j_1=1}^{L_n} \sum_{i_2,j_2=1}^{M_n} \sum_{t,s=1}^{T} \mathbb{K}_b\left(\frac{i_1 - j_1}{L_n}, \frac{i_2 - j_2}{M_n}, \frac{t-s}{T}\right) \hat{V}^*_{(i_1,i_2,t)} \hat{V}^*_{(j_1,j_2,s)}$$

where

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

We also 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 *new* 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) $K(\cdot)$ is symmetric with $K(0) = 1$, $|K(z)| \leq 1$ (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 of $K(\cdot)$ is greater than or equal to 1.*

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

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

where $\varphi_{b,k}(x) = \exp\left(i\frac{x}{b}\pi(k-1)\right)$ and $\left\{\Phi_{b,k\ell m}\left(\frac{i_1}{L_n}, \frac{i_2}{M_n}, \frac{t}{T}\right) \Phi_{b,k\ell m}\left(-\frac{j_1}{L_n}, -\frac{j_2}{M_n}, -\frac{s}{T}\right)\right\}$ 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

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

Using the representation, we can obtain the following result:

Proposition 2 *Let Assumptions F1 - F3 hold. For $b_1, b_2, b_3 \in (0,1]$, we have*

$$\hat{J}_{nT} \xrightarrow{d} \Lambda \left[\int_0^1 \mathbb{K}_b \left(r_1 - v_1, r_2 - v_2, \tau - \kappa\right) dB_p \left(r_1, r_2, \tau\right) dB_p' \left(v_1, v_2, \kappa\right) \right] \Lambda'. \tag{26}$$

Here and hereafter, we use "$\int$" to indicate multivariate integration to simplify the notation. 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).

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

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

Using $\mathbb{K}_b^* (\cdot, \cdot)$, (26) is equivalent to

$$\Lambda \left[\int_0^1 \mathbb{K}_b^* \left((r_1, r_2, \tau), (v_1, v_2, \kappa) \right) d\mathcal{W}_p \left(r_1, r_2, \tau\right) d\mathcal{W}_p' \left(v_1, v_2, \kappa\right) \right] \Lambda'. \tag{27}$$

In (27), the integration is with respect to the standard Wiener process because the centered kernel function captures the estimation uncertainty in $\hat{\beta}$. With (25) and (27), we can show that under H_0,

$$\begin{aligned}
F_{nT} =& \sqrt{nT} \left[\mathcal{R} \left(\hat{\beta} - \beta_0 \right) \right]' \left(\mathcal{R} \hat{Q}_{nT} \hat{J}_{nT} \hat{Q}_{nT}' \mathcal{R}' \right)^{-1} \sqrt{nT} \left[\mathcal{R} \left(\hat{\beta} - \beta_0 \right) \right] / g \\
\xrightarrow{d}& \left(\mathcal{R} \Lambda^* \mathcal{W}_p (1,1,1) \right)' \left(\mathcal{R} \Lambda^* \left[\int_0^1 \mathbb{K}_b^* \left((r_1, r_2, \tau), (v_1, v_2, \kappa) \right) d\mathcal{W}_p \left(r_1, r_2, \tau\right) d\mathcal{W}_p' \left(v_1, v_2, \kappa\right) \right] \Lambda^{*\prime} \mathcal{R}' \right)^{-1} \\
& \times \left(\mathcal{R} \Lambda^* \mathcal{W}_p (1,1,1) \right) / g \\
\overset{d}{=}& \mathcal{W}_g' (1,1,1) \left[\int_0^1 \mathbb{K}_b^* \left((r_1, r_2, \tau), (v_1, v_2, \kappa) \right) d\mathcal{W}_g \left(r_1, r_2, \tau\right) d\mathcal{W}_g' \left(v_1, v_2, \kappa\right) \right]^{-1} \mathcal{W}_g (1,1,1) / g \\
:=& F_\infty (g, b),
\end{aligned} \tag{28}$$

where the equality in distribution holds because $\mathcal{R} \Lambda^* \mathcal{W}_p (x, y, z) \overset{d}{=} \mathcal{R}^* \mathcal{W}_g (x, y, z)$ for a Wiener process $\mathcal{W}_g (x, y, z)$ and some $g \times g$ matrix $\mathcal{R}^*$ such that $\mathcal{R}^* (\mathcal{R}^*)' = \mathcal{R} Q J Q' \mathcal{R}'$.

Because of the random limit of $\hat{J}_{nT}$ with fixed b_1, b_2 and b_3 as $n, T \to \infty$, the distribution of $F_\infty (g, b)$ is nonstandard. As b_1, b_2 and $b_3 \to 0$, however, the effect of this randomness diminishes and $g F_\infty (g, b)$ converges in distribution to the χ_g^2 distribution.

7.2 Expansion of the limiting distribution and F-approximation

We present the asymptotic expansion of the distribution of $F_\infty (g, b)$ in (28) and establish the validity of a standard F-approximation.

Let

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

where v_{11} is a scalar. Following Sun (2010), we can show that

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

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

$$P\left\{gF_{\infty}\left(g,b\right)\leq z\right\}=G_{g}\left(z\right)+A\left(z\right)b_{1}b_{2}b_{3}+o\left(b_{1}b_{2}b_{3}\right)$$

where

$$A\left(z\right)=G_{g}''\left(z\right)z^{2}c_{2}-G_{g}'\left(z\right)z\left[c_{1}+\left(g-1\right)c_{2}\right],$$

$$c_{1}=\int_{-\infty}^{\infty}\mathbb{K}\left(x,y,z\right)dxdydz\ \text{ and }\ c_{2}=\int_{-\infty}^{\infty}\mathbb{K}^{2}\left(x,y,z\right)dxdydz.$$

Theorem 5 characterizes the nonstandard distribution $gF_{\infty}\left(g,b\right)$ when b_1, b_2 and b_3 are small. It clearly shows that the difference between $gF_{\infty}\left(g,b\right)$ and χ_g^2 depends on the smoothing parameters, kernel function and the number of restrictions being tested.

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

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

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

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

Using this representation, we have

$$\begin{aligned}
&\int_0^1\mathbb{K}_b^*\left(\left(r_1,r_2,\tau\right),\left(v_1,v_2,\kappa\right)\right)d\mathcal{W}_g\left(r_1,r_2,\tau\right)d\mathcal{W}_g'\left(v_1,v_2,\kappa\right)\\
&=\sum_{k,\ell,m,k',\ell',m'=1}^{\infty}\lambda_{k\ell mk'\ell'm'}\xi_{b,k\ell m}\xi_{b,k'\ell'm'}',
\end{aligned}\tag{29}$$

where $\xi_{b,k\ell m}=\int_0^1\varrho_{b,k\ell m}(x,y,z)d\mathcal{W}_g\left(x,y,z\right)\overset{i.i.d.}{\sim}N(0,I_g)$.

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

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

With this result, we follow Sun and Kaplan (2010) to obtain

$$\sum_{k,\ell,m,k',\ell',m'=1}^{\infty} \lambda_{k\ell mk'\ell'm'}\xi_{b,k\ell m}\xi_{b,k'\ell'm'}' = \sum_{k=1}^{\infty}\lambda_k\zeta_k\zeta_k',$$

where $\zeta_k \overset{i.i.d.}{\sim} N(0,I_g)$. 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 $\phi = \mathcal{W}_g(1,1,1)$. Then, we have

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

where ζ_k is independent of ϕ for all k. It is interesting to see that this representation of $gF_\infty(g,b)$ is exactly the same as that obtained by Sun (2010) for the fixed-smoothing asymptotic distribution of the Wald statistic in a time series context.

Let

$$\mu_1 = \sum_{k=1}^{\infty}\lambda_k = \int_0^1 \mathbb{K}_b^*\left((r_1,r_2,\tau),(r_1,r_2,\tau)\right)dr_1dr_2d\tau,$$
$$\mu_2 = \sum_{k=1}^{\infty}\lambda_k^2 = \int_0^1 \left[\mathbb{K}_b^*\left((r_1,r_2,\tau),(v_1,v_2,\kappa)\right)\right]^2 dr_1dr_2d\tau dv_1dv_2d\kappa.$$

Define $D = \left[\mu_1^2/\mu_2\right]$. Then using the same argument as in Sun (2010), we have the following approximation:

$$\frac{\mu_1(D-g+1)}{D}F_\infty(g,b) \overset{d}{\approx} F_{g,D-g+1}. \tag{30}$$

It can be shown that

$$\frac{\mu_1(D-g+1)}{D} = \frac{1}{1+b_1b_2b_3\left[c_1+(g-1)c_2\right]} + o(b_1b_2b_3). \tag{31}$$

The following theorem gives a rigorous description of 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

$$\nu = 1+\left[c_1+(g-1)c_2\right]b_1b_2b_3.$$

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

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

where $D^* = \max(5,\lceil 1/(b_1b_2b_3c_2)\rceil)$ and $\lceil\cdot\rceil$ denotes the integer part.

In Theorem 6 we use D^* in place of $D-g+1$ 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,D^*}^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 D^*. Theorem 6 suggests that for the F-test version of Wald statistic, F_{nT}, we use

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

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

8 Monte Carlo simulation

In this section, we provide some simulation evidence on the finite sample performance of our covariance estimator and the associated testing procedure. 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 $\hat{J}_{nT}^{KS}$. We evaluate the covariance estimators using the RMSE criterion and the coverage error of the associated confidence intervals (CIs) or regions. The latter is equivalent to the error of rejection probability of the underlying tests under the null. We examine the robustness to the measurement errors in economic distance. It is also investigated how the number of restrictions being tested affects the performance of the Wald test under the two different limiting thought experiments.

We assume a lattice structure, in which each individual is located on a square grid of integers. We 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 $u_t = (u_{1t}, \ldots, u_{nt})'$. $\tilde{W}_n$ is a contiguity matrix and individuals i and j are neighbors if $d_{ij} = 1$. Following the convention, it is row-standardized and its diagonal elements are zero. The parameters λ and θ determine the strength of the temporal and spatial correlation. We consider the following values for λ and θ: 0, 0.3, 0.6 and 0.9.

DGP1 is used for the RMSE criterion and DGP2 is used 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 by the sample average.

For the estimators $\hat{J}_{nT}^{DK}$ and $\hat{J}_{nT}^{KS}$, we employ the respective data-driven bandwidth in (23) and (24), using the time series AR(1) or spatial AR(1) as the approximating plug-in model. For $\hat{J}_{nT}$ with the Parzen kernel, we employ the bandwidths given in (10) and (11), using the spatiotemporal parametric model in (15) as the approximating plug-in model. W_n is the contiguity matrix in which individuals i and j are neighbors if $d_{ij} = 1$. We set $\eta_n = 2$ and $\ell_n = \pi d_n^2$. Note that the approximating parametric models for $\hat{J}_{nT}^{KS}$ and $\hat{J}_{nT}$ are mis-specified whereas the AR(1) model for $\hat{J}_{nT}^{DK}$ is correctly specified. We employ the QMLE to estimate parameters in (15) and (24).

For $\hat{J}_{nT}$ with the rectangular kernel, we use the Parzen kernel as the target kernel to obtain the data-driven bandwidths.

To obtain a positive semi-definite covariance estimator with the rectangular kernel, we follow Politis (2011) and modify $\hat{J}_{nT}$. According to the spectral decomposition, $\hat{J}_{nT} = \hat{U}\hat{\Lambda}\hat{U}'$, 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, $\hat{J}_{nT}^+$ is positive semi-definite.

The number of simulation replications is 1000, and three different sample sizes are considered; (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 following values are used for each kernel.

	$\bar{\mathcal{K}}_1$	$\bar{\mathcal{K}}_2$	c_1	c_2	K_q
Parzen	0.2889	0.2697	0.4123	0.1558	-6
Rectangular	1	1	6.2926	6.2926	

We 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. If $d_{ij} \geq 2$, then we observe d_{ij}^* :

$$d_{ij}^* = d_{ij} + e_{ij},$$

where $e_{ij} = -1, 0, 1$ with equal probabilities. PHAC, CCE, DK and KS denote the test statistics based on $\hat{J}_{nT}$, $\hat{J}_{nT}^A$, $\hat{J}_{nT}^{DK}$, and $\hat{J}_{nT}^{KS}$, respectively. We use the F-approximation to obtain critical values under the fixed-smoothing asymptotics.

Table 1 presents the ratios of the RMSE to J_{nT} for $\hat{J}_{nT}$ and $\hat{J}_{nT}^{DK}$ evaluated at the data dependent bandwidth parameters $(\hat{d}_n, \hat{d}_T)$ and $\hat{d}_T^{DK}$ and at infeasible optimal bandwidth parameters $(d_n^\star, d_T^\star)$ and d_T^{DK}. The infeasible bandwidth parameters are obtained by plugging the true data generating process into the AMSE* formula. Several patterns emerge. First, $\hat{J}_{nT}$ outperforms $\hat{J}_{nT}^{DK}$ in almost all the cases. When spatial dependence is absent or weak, $\hat{J}_{nT}$ has a substantially smaller RMSE than $\hat{J}_{nT}^{DK}$. Even when $\theta = 0.9$, these two estimators are not much different. In particular, when the rectangular kernel is used, $\hat{J}_{nT}$ is as accurate as and sometimes more accurate than $\hat{J}_{nT}^{DK}$. This implies that adaptiveness works well in this setting. Second, increasing n reduces only the RMSE of $\hat{J}_{nT}$ while increasing T reduces the RMSEs of both estimators. This is expected, as the rate of convergence of $\hat{J}_{nT}^{DK}$ depends only on T while that of $\hat{J}_{nT}$ depends on both n and T. Finally, the results under both feasible and infeasible AMSE*-optimal bandwidths show that the AMSE* criterion is effective in controlling 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 $\hat{J}_{nT}^{KS}$. DGP2 is used with a univariate regressor $(p = 1)$. For the testing with $\hat{J}_{nT}$, we use both the fixed-smoothing asymptotics and the increasing-smoothing asymptotics. The simulation results verify our theoretical results. First, we compare $\hat{J}_{nT}$ with the other estimators under the increasing-smoothing asymptotics. When $\theta = 0$ with high temporal autocorrelation, CCE performs better than PHAC. However, as θ increases, the performance of PHAC becomes better than that of CCE. Compared with KS, the CIs associated with PHAC have more accurate coverage probability unless the process is temporally highly persistent. Even with $\lambda = 0.9$ PHAC is almost as accurate as KS especially when n is small. Both PHAC and KS become more accurate with large n, but only the performance of PHAC improves when T increases. Comparison with DK is very similar to the case based on the RMSE criterion, as given in Table 1. Second, Table 2 compares the performances of PHAC under two different asymptotics. The results indicate that the fixed-smoothing asymptotic approximation is substantially more accurate than the increasing-smoothing asymptotic approximation. 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 the fixed-smoothing asymptotics is 79.6% but it is only 64.4% under the increasing-smoothing asymptotics. Third, Table 2 provides strong evidence that the rectangular kernel performs better than the finite-order kernel under the fixed-smoothing asymptotics. The performance of PHAC with the rectangular kernel is very robust to spatial dependence so that the size distortion does not increase with spatial dependence. This size advantage of the rectangular kernel arises from its bias reducing property and the adaptiveness of the bandwidth choice rule. Finally, Table 2 shows that our testing procedure based on the fixed-smoothing asymptotics is reasonably robust to measurement errors. 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 performances of the two different asymptotics when more than one parameters or restrictions are considered. DGP2 is used with $p = 3$. The confidence regions are obtained by inverting the Wald test of $H_0 : \ \beta_1 = 0$ with $g = 1$ and $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ with $g = 3$, respectively. The table evidently indicates that under the increasing-smoothing asymptotics the error in coverage probability increases with the number of parameters being considered. The coverage error becomes 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 the increasing-smoothing asymptotics. The coverage error of PHAC also increases under the fixed-smoothing asymptotics with the number of parameters or restrictions being tested but much lesser. This is consistent with our asymptotic expansion in Theorem 5. The theorem shows that the fixed-smoothing asymptotics and F-approximation correct for the number of restrictions being jointly tested.

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 statistic under both the increasing-smoothing asymptotics and the fixed-smoothing asymptotics. We also derive the optimal bandwidth selection procedure based on an upper bound of the AMSE. For the fixed-smoothing asymptotic distribution, we establish the validity of an F-approximation. The adaptiveness of our estimator ensures that it can be safely used without the knowledge of the dependence structure.

Instead of using the upper bound of the AMSE as the criterion, we can study the optimal bandwidth selection based on a criterion that is most suitable for hypothesis testing and CI construction. It is interesting to extend the bandwidth selection methods in time series HAC estimation by Sun (2010) and Sun and Kaplan (2010) 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	$\hat{J}_{nT}$ $(\hat{d}_n, \hat{d}_T)_{PA}$	0.09	0.20	0.27	0.46	$\hat{J}_{nT}$ $(d_n^\star, d_T^\star)_{PA}$	0.42	0.21	0.25	0.43
0.3		0.16	0.34	0.45	0.65		0.13	0.33	0.43	0.60
0.6		0.22	0.41	0.55	0.72		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	$\hat{J}_{nT}$ $(\hat{d}_n, \hat{d}_T)_{RE}$	0.13	0.23	0.29	0.38	$\hat{J}_{nT}$ $(d_n^\star, d_T^\star)_{RE}$	1.00	0.36	0.38	0.36
0.3		0.21	0.36	0.51	0.62		0.19	0.31	0.41	0.50
0.6		0.29	0.47	0.62	0.68		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	$\hat{J}_{nT}^{DK}$ $(\hat{d}_T^{DK})$	0.48	0.46	0.48	0.47	$\hat{J}_{nT}^{DK}$ (d_T^{DK})	0.36	0.36	0.38	0.36
0.3		0.54	0.56	0.57	0.54		0.56	0.56	0.58	0.54
0.6		0.68	0.70	0.69	0.70		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	$\hat{J}_{nT}$ $(\hat{d}_n, \hat{d}_T)_{PA}$	0.05	0.13	0.18	0.40	$\hat{J}_{nT}$ $(d_n^\star, d_T^\star)_{PA}$	0.42	0.12	0.17	0.39
0.3		0.10	0.24	0.34	0.55		0.14	0.24	0.33	0.50
0.6		0.14	0.33	0.50	0.64		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	$\hat{J}_{nT}$ $(\hat{d}_n, \hat{d}_T)_{RE}$	0.08	0.14	0.18	0.21	$\hat{J}_{nT}$ $(d_n^\star, d_T^\star)_{RE}$	1.00	0.20	0.19	0.20
0.3		0.13	0.26	0.37	0.57		0.21	0.22	0.29	0.32
0.6		0.19	0.41	0.67	0.58		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	$\hat{J}_{nT}^{DK}$ $(\hat{d}_T^{DK})$	0.28	0.29	0.27	0.28	$\hat{J}_{nT}^{DK}$ (d_T^{DK})	0.21	0.20	0.19	0.20
0.3		0.40	0.41	0.40	0.40		0.38	0.38	0.38	0.37
0.6		0.53	0.54	0.55	0.56		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	$\hat{J}_{nT}$ $(\hat{d}_n, \hat{d}_T)_{PA}$	0.05	0.13	0.18	0.29	$\hat{J}_{nT}$ $(d_n^\star, d_T^\star)_{PA}$	0.43	0.20	0.21	0.27
0.3		0.09	0.24	0.33	0.54		0.07	0.24	0.32	0.47
0.6		0.13	0.30	0.42	0.57		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	$\hat{J}_{nT}$ $(\hat{d}_n, \hat{d}_T)_{RE}$	0.07	0.15	0.21	0.30	$\hat{J}_{nT}$ $(d_n^\star, d_T^\star)_{RE}$	1.00	0.34	0.36	0.35
0.3		0.11	0.27	0.37	0.62		0.09	0.23	0.28	0.41
0.6		0.15	0.36	0.51	0.62		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	$\hat{J}_{nT}^{DK}$ $(\hat{d}_T^{DK})$	0.47	0.43	0.48	0.47	$\hat{J}_{nT}^{DK}$ (d_T^{DK})	0.37	0.34	0.36	0.35
0.3		0.53	0.56	0.55	0.55		0.54	0.56	0.56	0.55
0.6		0.68	0.70	0.69	0.69		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. Left and right panels are based on data-driven bandwidths and infeasible bandwidths, respectively.

Table 2: Empirical Coverage Probabilities of Nominal 95% CIs Constructed
Using Alternative Covariance Estimators - 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		93.9	94.2	91.8	88.0		89.4	89.1	88.6	90.5		94.2	94.6	93.3	94.6
0.3	PHAC	91.4	90.3	90.9	83.7	DK	87.0	83.7	88.3	86.1	PHAC	91.5	90.4	92.6	94.3
0.6	(PA,F)	87.5	88.2	85.1	79.2		77.4	79.0	76.1	77.0	(RE,F)	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		93.7	94.2	91.3	87.9		94.9	93.6	86.6	56.6		93.6	94.3	92.8	92.0
0.3	PHAC$_e$	91.0	90.0	89.8	83.1	CCE	93.0	91.3	86.6	54.2	PHAC$_e$	91.0	89.9	91.4	88.6
0.6	(PA,F)	86.7	87.1	82.7	77.2		92.9	91.9	83.9	53.7	(RE,F)	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		93.7	94.0	91.5	87.5		91.3	90.2	86.9	67.1		93.7	94.0	91.6	90.3
0.3	PHAC	91.0	90.0	90.3	82.3	KS	89.5	88.0	86.2	64.2	PHAC	90.6	89.4	90.9	85.1
0.6	(PA,I)	86.8	87.6	82.6	71.8		88.1	88.9	84.4	62.1	(RE,I)	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		94.7	92.7	91.5	88.0		92.6	93.1	92.7	93.9		94.8	93.4	92.9	95.5
0.3	PHAC	92.9	93.3	89.6	83.9	DK	90.5	92.1	90.1	90.1	PHAC	92.9	93.8	91.1	93.7
0.6	(PA,F)	93.1	91.5	90.2	84.4		87.6	87.4	88.5	87.3	(RE,F)	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		94.8	92.8	91.0	87.8		93.9	91.6	83.7	55.0		95.0	93.4	92.2	93.6
0.3	PHAC$_e$	92.5	92.9	88.6	83.8	CCE	93.5	92.3	81.9	54.1	PHAC$_e$	92.4	93.2	89.4	89.0
0.6	(PA,F)	93.0	91.0	88.8	83.8		94.3	91.9	85.9	55.5	(RE,F)	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		94.6	92.7	91.3	88.7		90.2	88.1	83.3	66.7		94.7	93.0	92.2	94.4
0.3	PHAC	92.4	93.1	88.8	84.8	KS	88.5	89.6	82.7	65.3	PHAC	92.7	93.2	89.8	89.6
0.6	(PA,I)	93.1	91.2	88.6	82.5		90.5	88.5	85.1	65.5	(RE,I)	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		93.6	92.4	93.2	90.8		86.1	87.7	88.8	89.4		93.6	93.0	94.2	92.9
0.3	PHAC	92.1	92.6	92.0	89.4	DK	85.0	86.1	85.0	87.3	PHAC	92.2	91.2	91.1	92.3
0.6	(PA,F)	91.0	89.9	88.2	88.2		80.1	82.7	80.1	75.1	(RE,F)	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		93.5	92.3	92.7	89.0		94.5	92.7	84.9	50.3		93.4	92.8	94.0	91.9
0.3	PHAC$_e$	92.1	91.6	89.8	88.0	CCE	94.7	92.2	85.0	50.3	PHAC$_e$	91.4	90.4	90.9	90.1
0.6	(PA,F)	90.4	88.7	86.9	83.7		94.4	94.2	86.5	47.7	(RE,F)	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		93.6	92.3	92.7	88.9		93.3	92.9	90.0	79.4		93.4	92.8	93.3	90.7
0.3	PHAC	92.3	90.8	89.8	86.5	KS	93.9	92.1	90.0	78.8	PHAC	92.1	90.9	90.3	87.0
0.6	(PA,I)	89.9	91.0	87.9	76.5		93.6	94.2	89.6	74.6	(RE,I)	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% Confidence Regions 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 notes 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

We have

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

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}, under Assumptions I1 and I2

$$|C_{1nT}| \leq \frac{c_R^4}{\ell_n\ell_T} \frac{1}{nT} \sum_{l=1}^{nTp} \left|E\varepsilon_l^4 - 3\right| \leq \frac{c_R^4 c_E p}{\ell_n \ell_T} = o(1) \tag{A.1}$$

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

$$C_{2nT} := D_{1nT} + D_{2nT} + D_{3nT} + D_{4nT} + D_{5nT}$$

where

$$D_{1nT} = \frac{1}{nT\ell_n\ell_T} \sum_{i,a\in E_n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t,u\in E_T} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ij_{(i)}}}{d_n}\right) K\left(\frac{d_{ab_{(a)}}}{d_n}\right) K\left(\frac{d_{ts_{(t)}}}{d_T}\right) K\left(\frac{d_{uv_{(u)}}}{d_T}\right) \times \gamma^{(c_1c_2)}_{(it,au)} \gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}$$

$$D_{2nT} = \frac{1}{nT\ell_n\ell_T} \sum_{i,a=1}^{n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t\notin E_T} \sum_{u=1}^{T} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ij_{(i)}}}{d_n}\right) K\left(\frac{d_{ab_{(a)}}}{d_n}\right) K\left(\frac{d_{ts_{(t)}}}{d_T}\right) K\left(\frac{d_{uv_{(u)}}}{d_T}\right) \times \gamma^{(c_1c_2)}_{(it,au)} \gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}$$

$$D_{3nT} = \frac{1}{nT\ell_n\ell_T} \sum_{i,a=1}^{n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t\in E_T} \sum_{u\notin E_T} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ij_{(i)}}}{d_n}\right) K\left(\frac{d_{ab_{(a)}}}{d_n}\right) K\left(\frac{d_{ts_{(t)}}}{d_T}\right) K\left(\frac{d_{uv_{(u)}}}{d_T}\right) \times \gamma^{(c_1c_2)}_{(it,au)} \gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}$$

$$D_{4nT} = \frac{1}{nT\ell_n\ell_T} \sum_{i\notin E_n} \sum_{a=1}^{n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t,u\in E_T} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ij_{(i)}}}{d_n}\right) K\left(\frac{d_{ab_{(a)}}}{d_n}\right) K\left(\frac{d_{ts_{(t)}}}{d_T}\right) K\left(\frac{d_{uv_{(u)}}}{d_T}\right) \times \gamma^{(c_1c_2)}_{(it,au)} \gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}$$

$$D_{5nT} = \frac{1}{nT\ell_n\ell_T} \sum_{i\in E_n} \sum_{a\notin E_n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t,u\in E_T} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} K\left(\frac{d_{ij_{(i)}}}{d_n}\right) K\left(\frac{d_{ab_{(a)}}}{d_n}\right) K\left(\frac{d_{ts_{(t)}}}{d_T}\right) K\left(\frac{d_{uv_{(u)}}}{d_T}\right) \times \gamma^{(c_1c_2)}_{(it,au)} \gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}$$

D_{1nT} is based on nonboundary units whereas D_{2nT}, D_{3nT}, D_{4nT} and D_{5nT} are based on boundary ones.

First, applying the proof of Theorem 1 in Kim and Sun (2011), we can show that

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

and

$$\lim_{n,T\to\infty} D_{1nT} = \lim_{n,T\to\infty} \frac{1}{nT\ell_n\ell_T} \sum_{i,a\in E_n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t,u\in E_T} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right) K^2\left(\frac{d_{ts_{(t)}}}{d_T}\right) \gamma^{(c_1c_2)}_{(it,au)} \gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}. \tag{A.3}$$

It is straightforward to show that (A.2) and (A.3) imply

$$\lim_{n,T\to\infty} D_{1nT} = \bar{\mathcal{K}}_1\bar{\mathcal{K}}_2 J(c_1,c_2)\, J(d_1,d_2).$$

For D_{2nT}, we have

$$D_{2nT} \le \frac{1}{nT} \sum_{i,a=1}^{n} \sum_{t\notin E_T} \sum_{u=1}^{T} \left|\gamma^{(c_1c_2)}_{(it,au)}\right| \left(\frac{1}{\ell_n\ell_T} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} \left|\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right|\right) = o(1), \tag{A.4}$$

as $T_2/T \to 0$. Using the similar procedure to (A.4), we can show $D_{3nT} = o(1)$, $D_{4nT} = o(1)$ and $D_{5nT} = o(1)$ given $T_2/T \to 0$ and $n_2/n \to 0$.

Thus,

$$\lim_{n,T\to\infty} C_{2nT} = \bar{\mathcal{K}}_1\bar{\mathcal{K}}_2 J(c_1,c_2)\, J(d_1,d_2).$$

By symmetry,

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

Therefore,

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

In terms of matrix form,

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

(b) Asymptotic Bias

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

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

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

(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), 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 $b \in \mathbb{R}^p$. In consequence, we can consider the case that J_{nT} is a scalar without loss of generality.

$$\begin{aligned} \sqrt{\frac{nT}{\ell_n\ell_T}}\left(\hat{J}_{nT} - \tilde{J}_{nT}\right) &= \left(\sqrt{nT}\left(\hat{\beta}-\beta_0\right)\right)^2\sqrt{\frac{\ell_n\ell_T}{nT}}\frac{1}{\ell_n\ell_T nT}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{it}^2\tilde{X}_{js}^2 \\ &\quad - 2\sqrt{nT}\left(\hat{\beta}-\beta_0\right)\sqrt{\frac{\ell_n\ell_T}{nT}}\frac{1}{\ell_n\ell_T\sqrt{nT}}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{js}^2\tilde{X}_{it}\tilde{u}_{it} \\ &\quad - \frac{2}{\sqrt{\ell_n\ell_T nT}}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{it}\tilde{X}_{js}u_{it}\left(\bar{u}_j + \bar{u}_s - \bar{u}\right) \\ &\quad + \frac{1}{\sqrt{\ell_n\ell_T nT}}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{it}\tilde{X}_{js}\left(\bar{u}_i + \bar{u}_t - \bar{u}\right)\left(\bar{u}_j + \bar{u}_s - \bar{u}\right) \\ &:= H_{1nT} + H_{2nT} + H_{3nT} + H_{4nT}. \end{aligned}$$

It is easy to show that $H_{1nT} = o_p(1)$ and $H_{2nT} = o_p(1)$ under Assumptions I4 and I7. For $H_{3nT,}$ we need to show that for all i and t

$$\frac{1}{\sqrt{\ell_n\ell_T}}\sum_{j=1}^{n}\sum_{s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{js}\left(\bar{u}_j + \bar{u}_s - \bar{u}\right) = o_p(1). \tag{A.5}$$

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 second 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

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

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

Proof of Corollary 1

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

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

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

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

since S is pd. Therefore,

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

and the 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

The proofs of (a) and (b) are analogous to the proofs of Theorem 1(a) and (c) respectively.

Proof of Theorem 3

The proofs of (a), (b), (c) and (d) are analogous to the proofs of Theorem1(a), (b), (c) and (d) respectively.

Proof of Theorem 4

The proofs of (a), (b), (c) and (d) are analogous to the proofs of Theorem 1(a), (b), (c) and (d) respectively.

Proof of Proposition 1

(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

$$\tilde{J}_{nT}(c,d) - \tilde{J}_{nT}^{GA}(c,d) = o_p(1), \tag{A.6}$$

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 $\Delta > 0$,

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

where

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

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

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

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

Following (A.1), we can show $\tilde{C}_{1nT} = o(1)$. For $\tilde{C}_{2nT}$,

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

as $d_n \to 0$ because $K_{RE}(d_{ij}/d_n) = 0$ 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.6) holds.

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

$$\tilde{J}_{nT}(c,d) - \tilde{J}_{nT}^{DK}(c,d) = o_p(1), \tag{A.7}$$

if $\ell_n/n \to 1$ as $n \to \infty$.

By Chebyshev's inequality, we have

$$P\left(\left|\tilde{J}_{nT}(c,d)-\tilde{J}_{nT}^{DK}(c,d)\right|>\Delta\right)\leq\frac{1}{\Delta^2}E\left(\tilde{J}_{nT}(c,d)-\tilde{J}_{nT}^{DK}(c,d)\right)^2$$
$$:=\check{C}_{1nT}+\check{C}_{2nT}+\check{C}_{3nT}+\check{C}_{4nT},$$

for any Δ, where

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

We can show that $\check{C}_{1nT}=o(1)$ using the procedure in (A.1). For $\check{C}_{2nT}$,

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

as $d_n\to\infty$. For $\check{C}_{3nT}$,

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

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

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

which implies $\check{C}_{3nT} \to 0$ as $n, T \to \infty$. With the same procedure, we can show that $\check{C}_{4nT} = o(1)$. Therefore, (A.7) holds.

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

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

The proof of Proposition 2 uses the lemma below whose proof is given in the supplementary appendix.

Lemma 1 *Let*

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

Then, under F1 - F2

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

Proof of Proposition 2

Let

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

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

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

by the Fourier series representation. This implies

$$\hat{J}_{nT} - \check{J}_{nT} = o_p(1). \tag{A.8}$$

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

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

where superscript 'H' denotes the conjugate transpose.

From Lemma 1 and (A.8), we have

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

where the equality in distribution holds because

$$\sum_{k=1}^{\mathcal{K}}\sum_{\ell=1}^{\mathcal{L}}\sum_{m=1}^{\mathcal{M}}\lambda_{k,\ell,m}\Phi_{b,k\ell m}\left(r_1,r_2,\tau\right)\Phi_{b,k\ell m}\left(-v_1,-v_2,-\kappa\right)\rightarrow\mathbb{K}_b\left(r_1-v_1,r_2-v_2,\tau-\kappa\right)$$

as $\mathcal{K}\rightarrow\infty$, $\mathcal{L}\rightarrow\infty$ and $\mathcal{M}\rightarrow\infty$ in L^2.

Lemma 2 *As b_1, b_2 and $b_3\rightarrow 0$, we have*

$$(a)\ \ \mu_1 = 1-b_1b_2b_3c_1+o\left(b_1b_2b_3\right); \quad (b)\ \ \mu_2 = b_1b_2b_3c_2+o\left(b_1b_2b_3\right).$$

Proofs are given in the supplementary appendix.

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

(a) $E\left(v_{11}-v_{12}v_{22}^{-1}v_{21}\right)=1-b_1b_2b_3c_1-(g-1)\,b_1b_2b_3c_2+o\left(b_1b_2b_3\right)$,
(b) $E\left(v_{11}-v_{12}v_{22}^{-1}v_{21}\right)^2=1-2b_1b_2b_3\left(c_1+(g-2)\,c_2\right)+o\left(b_1b_2b_3\right)$,
(c) $E\left[\left(v_{11}-v_{12}v_{22}^{-1}v_{21}\right)-1\right]^2=2b_1b_2b_3c_2+o\left(b_1b_2b_3\right)$.

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

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

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

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

Proof of Theorem 6

It follows from Theorem 5 that

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

By definition,

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

where we have used Lemma 2. Hence

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

References

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

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

[3] Arellano, M. (1987). Computing robust standard errors for within group estimators. *Oxford Bulletin of Economics and Statistics*, 49:431–434.

[4] Bertrand, M., Duo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275.

[5] 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.

[6] Bester, C., Conley, T., and Hansen, C. (2011). Inference with dependent data using cluster covariance estimators. *Journal of Econometrics*, forthcoming.

[7] Cameron, A., Gelbach, J., and Miller, D. (2009). Robust inference with multi-way clustering. Working Papers, UC Davis.

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

[9] Conley, T. (1999). GMM Estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45.

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

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

[12] Driscoll, J. and Kraay, A. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics,* 80(4):549–560.

[13] Fama, E. and MacBeth, J. (1973). Risk, return, and equilibrium: empirical tests. *Journal of Political Economy*, 81(3): 607–636

[14] Gonçlaves, S. (2010). The moving blocks bootstrap for panel linear regression models with individual fixed effects. Working paper, Universite de Montreal.

[15] 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.

[16] Ibragimov, R. and Müller, U. (2010). t-Statistic based correlation and heterogeneity robust inference. *Journal of Business and Economic Statistics*, 28, 453–468.

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

[18] Kelejian, H. and Prucha, I. (2007). HAC Estimation in a spatial framework. *Journal of Econometrics*, 140(1):131–154.

[19] Kiefer, N. and Vogelsang, T. (2002a). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory*, 18(06):1350–1366.

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

[21] Kiefer, N. and Vogelsang, T. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21(06):1130–1164.

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

[23] Kim, M. S. and Sun, Y. (2011). Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Journal of Econometrics*, 160(2):349–371.

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

[25] Lin, C. and Sakata, S. (2009). On long-run covariance matrix estimation with the truncated flat kernel. Working paper, University of British Columbia.

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

[27] Newey, W. and West, K. (1987). A Simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

[28] Newey, W. and West, K. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61:631–653.

[29] Petersen, M. (2009). Estimating standard errors in finance panel data sets: comparing approaches. *Review of Financial Studies*, 22(1):435–480

[30] Politis, D. (2011). Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econometric Theory*. Forthcoming.

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

[32] Sun, Y. and Kaplan, D. (2010). A new asymptotic theory for vector autoregressive long-run variance estimation and autocorrelation robust testing. Working paper, UCSD.

[33] Sun, Y. and Kim, M. S. (2010). Asymptotic F test in the presence of nonparametric spatial dependence. Working paper, UCSD.

[34] Sun, Y. and Phillips, P. (2009). Bandwidth choice for interval estimation in GMM regression. Working paper, UCSD.

[35] Sun, Y., Phillips, P., and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica*, 76(1):175–194.

[36] Vogelsang, T. (2008). Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. Working paper, Michigan State University.

[37] White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838.

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

[39] 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.