Title
Improved inference for spatial and panel models

Author
Kim, Min Seong

Publication Date
2011

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Improved Inference for spatial and panel models

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Min Seong Kim

Committee in charge:

Professor Yixiao Sun, Chair
Professor Graham Elliott
Professor Anthony Gamst
Professor Dimitris Politis
Professor Halbert White

2011

The dissertation of Min Seong Kim is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2011

DEDICATION

To Sora and Teao.

TABLE OF CONTENTS

Signature Page iii

Dedication iv

Table of Contents v

List of Figures vii

List of Tables viii

Acknowledgements ix

Vita x

Abstract of the Dissertation xi

Chapter 1 Heteroskedasticity and Spatiotemporal Dependence Robust Inference for Linear Panel Models with Fixed Effects 1
1.1 Introduction 1
1.2 Panel model, covariance estimator and hypothesis testing 7
1.3 Increasing smoothing asymptotics 8
1.3.1 Basic setting 8
1.3.2 Properties of $\hat{J}_{nT}$ and limiting distribution of Wald statistic under increasing smoothing asymptotics . 13
1.4 Optimal bandwidth selection procedure 17
1.5 Comparison with CCE, DK and KS estimators 24
1.5.1 CCE 24
1.5.2 DK estimator 25
1.5.3 KS estimator 26
1.6 Adaptiveness of $\hat{J}_{nT}$ 28
1.6.1 Flexibility 28
1.6.2 Adaptiveness 28
1.7 Fixed smoothing asymptotics 30
1.7.1 Limiting theory for $\hat{J}_{nT}$ under fixed smoothing asymptotics 30
1.7.2 Expansion of the distribution and F-approximation 34
1.8 Monte Carlo simulation 39
1.9 Conclusion 44
1.10 Acknowldegements 44
1.11 Appendix 51
1.12 Bibliography 71

Chapter 2 Spatial Heteroskedasticity and Autocorrelation Consistent Estimation of Covariance Matrix 75
2.1 Spatial Processes and HAC Estimators 78
2.2 Asymptotic Properties of Spatial HAC Estimators 85
2.3 Optimal Bandwidth Parameter and Data Dependent Bandwidth Selection 89
2.4 Properties of Data Dependent Bandwidth Parameter Estimators 94
2.5 Monte Carlo Simulation 95
2.5.1 Location Model 97
2.5.2 Univariate Model 103
2.6 Conclusion 104
2.7 Acknowldegements 105
2.8 Appendix 113
2.9 Bibliography 140

Chapter 3 *k*-step Bootstrap Bias Correction for Fixed Effects Estimators in Nonlinear Panel Models 144
3.1 Incidental Parameters Problem 147
3.2 Bootstrap Bias Correction 149
3.3 *k*-step Bootstrap Bias Correction 152
3.4 Asymptotic Properties 154
3.5 Bias Correction for Average Marginal Effects 156
3.6 Monte Carlo Study 157
3.7 Conclusion 161
3.8 Acknowledgments 162
3.9 Appendix 167
3.10 Bibliography 183

LIST OF FIGURES

Figure 1.1: Level curves of $d_n^\star$ and $d_T^\star$ with respect to spatio-temporal dependences 19
Figure 1.2: Adaptiveness of $\hat{J}_{nT}$ 30
Figure 2.1: Spatial AR(1) Process : $V_n = \rho W_n V_n + \varepsilon_n$, $n = 400$ 91

LIST OF TABLES

Table 1.1: RMSE/Estimand with $\hat{J}_{nT}$ and $\hat{J}_{nT}^{DK}$ 45
Table 1.2: RMSE/Estimand with $\hat{J}_{nT}$ and $\hat{J}_{nT}^{DK}$ - Cont. 46
Table 1.3: ECPs of Nominal 95% CIs with Alternative Covariance Estiamtors 47
Table 1.4: ECPs of Nominal 95% CIs with Alternative Covariance Estiamtors - Cont. 48
Table 1.5: ECPs of Nominal 95% CIs with Alternative Covariance Estimators - Cont. 49
Table 1.6: ECPs of Nominal 95% CIs with Different Number of Restrictions 50

Table 2.1: Ratio of the MSE of Spatial HAC Estimators with Different Bandwidths to the MSE of the Spatial HAC Estimator with Optimal Bandwidth, $\tilde{d}_n$ 106
Table 2.2: Bias, Variance, and MSE in Location Model with Spatial AR(1) Error 107
Table 2.3: ECPs in Location Model with Spatial AR(1) Error 108
Table 2.4: Performance under Misspecified Approximating Parametric Model 109
Table 2.5: ECPs with the Bootstrap and Standard Normal Approximations 110
Table 2.6: Performance in the Presence of Irregulairty and Sparsity 111
Table 2.7: Bias and MSE of $\hat{J}_{22,n}$ and ECPs with Different Weighting Matrices 112

Table 3.1: k-step Bootstrap Estimators 163
Table 3.2: Bias Corrected Estimators of θ 164
Table 3.3: Bias Corrected Estimators of the Average Marginal Effect μ . . 165
Table 3.4: Bias Corrected Estimators of the Overall Average Marginal Effect ν 166

ACKNOWLEDGEMENTS

I sincerely thank God for making me study econometrics under my advisor Yixiao Sun. I cannot imagine my study at UCSD without his immeasurable support. From the beginning of my research, he has been the "light" that I could always rely on. His warm encouragement, patient guidance and insightful advice has enabled me to finally complete the Ph.D. program. I have also learned the attitude I will keep in mind as a researcher from his enthusiasm and passion toward his research.

I thank Brendan Beare, Graham Elliott, Anthony Gamst, Patrik Guggenberger, James Hamilton, Ivana Komunjer, Dimitris Politis, Andres Santos and Halbert White for their considerate support.

Chapter 1 is coauthored with Yixiao Sun. The dissertation author was the primary author of this paper.

Chapter 2, which is coauthored with Yixiao Sun, is published in Journal of Econometrics, 2011. Vol 160 (2011), pp. 349-371. The dissertation author was the primary author of this paper.

Chapter 3, which is coauthored with Yixiao Sun, has been submitted for publication. The dissertation author was the primary author of this paper.

VITA

2000	B. A. in Economics, Seoul National University
2005	M. S. in Statistics, University of California, San Diego
2011	Ph. D. in Economics, University of California, San Diego

PUBLICATIONS

Min Seong Kim and Yixiao Sun "Spatial Heteroskedasticity and Autocorrelation Consistent Estimation of Covariance Matrix," *Journal of Econometrics*, Vol 160, 2007, pp 349-371.

ABSTRACT OF THE DISSERTATION

Improved Inference for spatial and panel models

by

Min Seong Kim

Doctor of Philosophy in Economics

University of California, San Diego, 2011

Professor Yixiao Sun, Chair

Chapter 1 is "Heteroskedasticity and Spatiotemporal Dependence Robust Inference for Linear Panel Models with Fixed Effects." This chapter studies robust inference for linear panel models with fixed effects in the presence of heteroskedasticity and spatiotemporal dependence of unknown forms. We propose a bivariate kernel covariance estimator, which is flexible to nest existing estimators as special cases with certain choices of bandwidths. For distributional approximations, we consider two different types of asymptotics. When the level of smoothing is assumed to increase with the sample size, the proposed estimator is consistent and the associated Wald statistic converges to a χ^2 distribution. We show that our covariance estimator improves upon existing estimators in terms of robustness and efficiency. When we assume the level of smoothing to be held fixed, the covariance

estimator has a random limit and we show by asymptotic expansion that the limiting distribution of the test statistic depends on the bandwidth parameters, the kernel function, and the number of restrictions being tested. As this distribution is nonstandard, we establish the validity of an F-approximation to this distribution, which greatly facilitates the test. For optimal bandwidth selection, we propose a procedure based on the upper bound of asymptotic mean square error criterion. The flexibility of our estimator and proposed bandwidth selection procedure make our estimator adaptive to the dependence structure in data. This *adaptiveness* automates the selection of covariance estimator. That is, our estimator reduces to the existing estimators which are designed to cope with the particular dependence structures. Simulation results show that the F-approximation and the adaptiveness work reasonably well.

Chapter 2 is "Spatial Heteroskedasticity and Autocorrelation Consistent Estimation of Covariance Matrix.". This chpater considers spatial heteroskedasticity and autocorrelation consistent (spatial HAC) estimation of covariance matrices of parameter estimators. We generalize the spatial HAC estimators introduced by Kelejian and Prucha (2007) to apply to linear and nonlinear spatial models with moment conditions. We establish its consistency, rate of convergence and asymptotic truncated mean squared error (MSE). Based on the asymptotic truncated MSE criterion, we derive the optimal bandwidth parameter and suggest its data dependent estimation procedure using a parametric plug-in method. The finite sample performances of the spatial HAC estimator are evaluated via Monte Carlo simulation.

Chapter 3 is "k-step Bootstrap Bias Correction for Fixed Effects Estimator in Nonlinear Panel Models." Fixed effects estimators in nonlinear panel models with fixed T usually suffer from inconsistency because of the incidental parameters problem first noted by Neyman and Scott (1948). Moreover, even though T grows at the same rate as n, they are asymptotically biased and therefore the associated confidence interval has a large coverage error. This chapter proposes a k-step parametric bootstrap bias corrected estimator. We prove that our estimator is asymptotically normal and is centered at the true parameter if T grows faster

than $\sqrt[3]{n}$. In addition to bias correction, we construct a confidence interval with a double bootstrap procedure, and Monte Carlo experiments confirm that the error in coverage probability of our CI's is smaller than those of the alternatives. We also propose bias correction for average marginal effects.

Chapter 1

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

1.1 Introduction

This paper studies robust inference for linear panel models with fixed effects in the presence of heteroskedasticity and spatiotemporal dependence of unknown forms. As economic data is potentially heterogeneous and correlated in unknown ways across individuals and time, the robust inference in the panel setting is 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 some *a priori* knowledge about the dependence structure. It is often assumed that the covariance of two random variables at locations i and j is a decreasing function of an observable distance measure d_{ij} between them. An example of d_{ij} is the economic distance. The idea of using a distance measure to characterize spatial dependence is common in the spatial econometrics literature. See, for example, Conley (1999), Kelejian and Prucha (2007), Bester, Conley, Hansen and Vogelsang (2009, BCHV hereafter) and Kim and Sun (2010, KS hearafter).

There are several robust covariance estimators with correlated panel data. Liang and Zeger (1986) and Arellano (1987) propose the clustered covariance estimator (CCE) by extending White standard error (White, 1980) to account for serial correlation. Wooldridge (2003) provides a concise review on the CCE and Kèzdi (2003) explores its properties in panel models with fixed effects. Driscoll and Kraay (1998, DK hereafter) suggest a different approach that uses a time series HAC estimator (Domowitz and White, 1982; and Newey and West, 1987) with the cross-sectional averages of moment conditions. Gonçlaves (2008) examines the properties of this estimator in linear panel models with fixed effects. Another approach considered in this paper is the extension of the spatial HAC estimator with serial averages of moment conditions, which we name the KS estimator. This is symmetric to the DK estimator. The spatial HAC estimator is firstly proposed by Conley (1999), and Kelejian and Prucha (2007) argue that it can be extended to the panel setting with fixed T.

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

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

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

The proposed estimator improves upon the above estimators by employing a bivariate kernel. It does not require zero spatial correlation for consistency in

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

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 $n = o(T^4)$. The conditions are more generous with the second order kernels, such as the Parzen kernel, i.e. $T = o(n^{5/2})$ and $n = o(T^6)$.

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

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

Several testing methods using the fixed smoothing asymptotics are recently proposed in the spatial or panel setting. BCHV extend the fixed-b asymptotics to the spatial context where dependence is indexed in more than one dimension, and propose an *i.i.d.* bootstrap method to obtain the critical values. Vogelsang (2008) develops a fixed-b asymptotic theory for statistics based on the generalized CCE and the DK estimator. Besides the kernel methods, Hansen (2007) and Bester, Conley and Hansen (2009) apply the fixed smoothing asymptotics to the testing procedure with the CCE. They assume the number of clusters to be fixed and the number of observations per cluster to increase with the sample size. Ibragimov and Müller (2010) consider fixed smoothing asymptotics for the Fama and MacBeth (1973) type procedure by fixing the number of groups. Sun and Kim (2010) considers a testing procedure using the series covariance estimator in the spatial setting. They show that, when the number of basis functions is assumed to be held fixed, the series estimator converges in distribution to a Wishart distribution, and that the scaled Wald statistic converges to F distribution. The motivation of our F-approximation arises from the series method. All the kernel covariance estimators with positive semi-definite kernels can be written as series covariance estimators (Percival and Walden, 1993, p. 353). Actually, the other two 'non-kernel' methods also use the t or F distribution as the reference distribution while the kernel methods by BCHV and Vogelsang (2008) have to simulate the critical values. From this point of view, this paper fills the gap in the literature, providing F-approximation for the kernel method in the panel setting.

We propose an optimal bandwidth selection procedure based on the upper bound of the asymptotic mean square error (AMSE*) criterion. Though it is standard practice to use the asymptotic mean square error (AMSE) criterion in the HAC estimation literature (e.g. Andrews, 1991 and Newey and West, 1994), it is not tractable for the proposed estimator. Our minimax criterion is simple to implement and makes the bias and variance tradeoff transparent. It is interesting to note that the level of persistence in each dimension affects both $d_T^\star$ and $d_n^\star$, the optimal bandwidth parameters in the time and spatial dimensions respectively, but in the opposite direction. 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 and, more broadly, flat-top kernel estimators. However, it is interesting to consider flat-top kernel estimators because they are higher order accurate (Politis, 2010). This is particularly more important in our setting because the flexibility is the main ingredient of the adaptiveness of our estimator which is explained below. We modify our bandwidth selection procedure to be applicable to the rectangular kernel. This modified procedure leads the rectangular kernel based estimator to better asymptotic properties than the one with any finite order kernel we target.

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

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

approximation. Section 8 reports simulation evidence. The last section concludes.

1.2 Panel model, covariance estimator and hypothesis testing

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

$$Y_{it} = X_{it}'\beta_0 + \alpha_i + f_t + u_{it}, \tag{1.1}$$

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

$$\hat{\beta} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{Y}_{it}. \tag{1.2}$$

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

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

where

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

To make inference on β_0, we have to estimate unknown quantities in the asymptotic variance of $\hat{\beta}$. Since Q_{nT} is consistently estimated with its sample analog, our central interest is on J_{nT}. Letting $V_{(i,t)} = \tilde{X}_{it}u_{it}$, J_{nT} can be rewritten as

$$J_{nT} = \frac{1}{nT}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}E\left[V_{(i,t)}V_{(j,s)}'\right] := \frac{1}{nT}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}\Gamma_{(it,js)}. \tag{1.3}$$

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

We propose a bivariate kernel covariance estimator which is given as

$$\hat{J}_{nT} = \frac{1}{nT} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} K\left(\frac{d_{ij}}{d_n}\right) K\left(\frac{d_{ts}}{d_T}\right) \hat{V}_{(i,t)} \hat{V}'_{(j,s)}, \tag{1.4}$$

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

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

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

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

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

1.3 Increasing smoothing asymptotics

1.3.1 Basic setting

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

$$V_{(i,t)} = \tilde{R}_{(i,t)}\tilde{\varepsilon}, \tag{1.5}$$

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

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

$$V_{(i,t)} = R_{(i,t)} \varepsilon \text{ and } \text{var}\left(\varepsilon \right) = I_{nTp}. \tag{1.6}$$

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

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

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

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

For simplicity, we assume that ε_l is independent of ε_k for $l \neq k$. We can relax independence assumption to zero correlation but with more tedious calculations. Under Assumption F5, the covariance matrix of $V_{(i,t)}$ and $V_{(j,s)}$ is given by

$$\Gamma_{(it,js)} := \left(\gamma_{(it,js)}^{(cd)}\right) = E\left[V_{(i,t)} V_{(j,s)}'\right] = R_{(i,t)} R_{(j,s)}', \tag{1.7}$$

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

$$J_{nT} = \frac{1}{nT} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} R_{(i,t)} R_{(j,s)}', \tag{1.8}$$

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

$$J_{nT}(c, d) = \frac{1}{nT} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} \left(\sum_{l=1}^{nTp} r_{(i,t),l}^{(c)} r_{(j,s),l}^{(d)} \right). \tag{1.9}$$

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

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

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

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

Assumptions F6 and F7 impose the conditions on the persistence of the process. If $|\sigma^{cd}| \leq C$ for a constant $C > 0$, then Assumption F6 holds if $\sum_{i=1}^{n} \sum_{t=1}^{T} |\tilde{r}_{(it,j,s)}^{(d)}| < c_R / C$. Since $|\tilde{r}_{(it,j,s)}^{(d)}|$ can be regarded as the (absolute) change of $V_{(i,t)}^{(d)}$ in response to one unit change in $\tilde{\varepsilon}_{(j,s)}^{(d)}$, the summability condition requires that the aggregate response to an innovation be finite. Assumption F7 implies that $\Gamma_{(it,js)}$ decays to zero fast as d_{ij} and d_{ts} increase so that the two summability conditions holds. These conditions hold if

$$\frac{1}{nT}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}\left|\sum_{a=1}^{n}\sum_{b=1}^{T}\tilde{r}_{(it,a,b)}^{(c)}\tilde{r}_{(js,a,b)}^{(d)}\right| d_{ij}^{q_1} < \infty, \tag{1.10}$$

$$\frac{1}{nT}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}\left|\sum_{a=1}^{n}\sum_{b=1}^{T}\tilde{r}_{(it,a,b)}^{(c)}\tilde{r}_{(js,a,b)}^{(d)}\right| d_{ts}^{q_2} < \infty \tag{1.11}$$

for all c and d. (1.10) and (1.11) imply that as d_{ij} or d_{ts} increases, the corresponding two row vectors in $\tilde{R}_{(i,t)}$ and $\tilde{R}_{(j,s)}$, $(\tilde{r}_{(it,1,1)}^{(c)}, \ldots, \tilde{r}_{(it,n,T)}^{(c)})$ and $(\tilde{r}_{(js,1,1)}^{(d)}, \ldots, \tilde{r}_{(js,n,T)}^{(d)})$ become nearly orthogonal. As the row vector represents the aggregate response of a unit to all the innovations, this assumption implies the responses of two units become independent as they become spatially or serially distant. Assumption F7 enables us to truncate the sum of $\Gamma_{(it,js)}$ and downweigh the summand without incurring much bias.

As Assumption F7 implies, the key property of d_{ij} is to characterize the decaying pattern of the spatial dependence. In addition, we assume that d_{ij} satisfies the properties of distance in a metric space: (i) $d_{ij} \geq 0$, (ii) $d_{ii} = 0$, (iii) $d_{ij} = d_{ji}$, and (iv) $d_{ij} \leq d_{ik} + d_{kj}$. In practice, nonetheless, the symmetry condition (iii) may not hold for some candidates of economic distance. Conley and Ligon (2000), for example, notice that transportation costs among countries violate this condition if tariff barriers are asymmetric. In such a case adjustment should be made.[4] This adjustment does not affect the asymptotic properties of our estimator from a perspective of the measurement error problem as explained below.

Observed distance data available to empirical researchers usually contain measurement errors, and the results in this paper can be generalized to the case when d_{ij} is error contaminated. Following KS, we can show that our asymptotic results are still valid under the following conditions: (i) the measurement error

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

is independent of ε_l; (ii) it is of order $o(d_n)$ as d_n increases; and (iii) the first summability condition in Assumption F7 holds with an error contaminated distance measure. In this paper, however, we do not consider measurement errors for simplicity.

Let

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

$\ell_{i,n}$ is the number of pseudo-neighbors that unit i has and ℓ_n is the average number of pseudo-neighbors. Here we use the terminology "pseudo-neighbor" in order to differentiate it from the common usage of "neighbor" in spatial modeling. We maintain the following assumption on the number of pseudo-neighbors.

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

Assumption F9 allows the units to be irregularly located but rules out the case that they are concentrated only in some limited area while other area is scarce. To be symmetric, we also define

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

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

In order to obtain the properties of the estimator in Theorem 7 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 relies on choice of the corresponding bandwidth. We can mitigate the boundary effects by raising d_n and d_T slowly as n and T increase to make the interior large enough. Provided that n_2/n and T_2/T are $o(1)$, the boundary effects are asymptotically negligible. When units are regularly spaced on a lattice in $\mathbb{R}^2$, $n_2/n = o(1)$ if $\ell_n/n = o(1)$. $T_2/T = o(1)$ holds if $\ell_T/T = o(1)$.

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

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

Assumption I5. *(i) The kernel $K : \mathbb{R} \to [0,1]$ satisfies $K(0) = 1, K(x) = K(-x), K(x) = 0$ for $|x| \geq 1$. (ii) For all $x_1, x_2 \in \mathbb{R}$ 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 F11(i) and (ii) are the Bartlett, Tukey-Hanning and Parzen kernels. The quadratic spectral (QS) kernel does not satisfy Assumption F11(i) because it does not truncate. We may generalized our results to include the QS kernel but this requires considerable amount of work. Assumption F11(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

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

In finite samples, we may use

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

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

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

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

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

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

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

Assumption F10 states that the covariance matrix defined locally for each nonboundary unit converges to the same limiting value of J_{nT}. This assumption is related to covariance stationarity but weaker. It is implied by covariance stationarity but it can hold even though covariance stationarity is violated. Stationarity seems to be a very strong assumption especially in the spatial dimension because a spatial process is nonstationary simply if each unit has different numbers of neighbors. This assumption is similar to the homogeneity assumption in Bester, Hansen and Conley (2009). They assume that the covariance matrix in each 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 F12 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 $\text{vec}(\cdot)$ 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 Theorem 7(c) states below. Therefore, we use $\tilde{J}_{nT}$ to analyze the asymptotic properties of $\hat{J}_{nT}$.

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

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

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

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

Theorem 1. *Suppose that Assumptions F5-F10 hold, $d_n, d_T \to \infty$, $n_2(d_n) = o(n)$, $T_2(d_T) = o(T)$, $\ell_n = o(n)$ 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) *Let $k_{nT} = d_T/d_n$ and $k_{nT} \to k > 0$ as $n, T \to \infty$. Then,* $\lim_{n,T\to\infty} d_n^q (E\tilde{J}_{nT} - J_{nT}) = -K_q \left(b_1^{(q)} + \frac{1}{k^q} b_2^{(q)}\right)$

(c) *If Assumption F12 holds and $d_n^{2q} \ell_n \ell_T / nT \to \tau \in (0, \infty)$, then*

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

(d) *Under the conditions of part (c), Assumptions I8 and A.17,*

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

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

Proofs are given in the appendix. For each element of $\tilde{J}_{nT}$, the asymptotic variance in Theorem 7(a) is rewritten as

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

Theorem 7(a) and (b) show that the asymptotic variance and bias of $\tilde{J}_{nT}$ depend on the choice of d_n and d_T. When we enlarge d_n and/or d_T, the bias decreases while the variance increases and vice versa. The second part of Theorem 7(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, $\ell_n \ell_T = o(nT)$ is the condition for the consistency and its rate of convergence is $\sqrt{nT/\ell_n \ell_T}$. It is also required that $T_2(d_T) = o(T)$ and $n_2(d_n) = o(n)$ to control the boundary effects. Let $\eta_T = 1$. If we set $\ell_n = O(d_n^{\eta_n})$ and $\ell_T = O(d_n^{\eta_T})$ for some $\eta_n > 0$, then the rate of convergence under the rate condition $d_n^{2q} \ell_n \ell_T / nT \to \tau \in (0, \infty)$ is $(nT)^{q/(2q+\eta_n+\eta_T)}$.

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

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

1.4 Optimal bandwidth selection procedure

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

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

Then, up to smaller order terms

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

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

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

where

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

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

$$(d_n^\star, d_T^\star) = \underset{d_n, d_T}{\arg\min}\ 2K_q^2\left(\frac{B_{11}}{d_n^{2q}} + \frac{B_{22}}{d_T^{2q}}\right) + \frac{\ell_n \ell_T}{nT}\bar{\mathcal{K}}_1\bar{\mathcal{K}}_2 C, \tag{1.12}$$

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

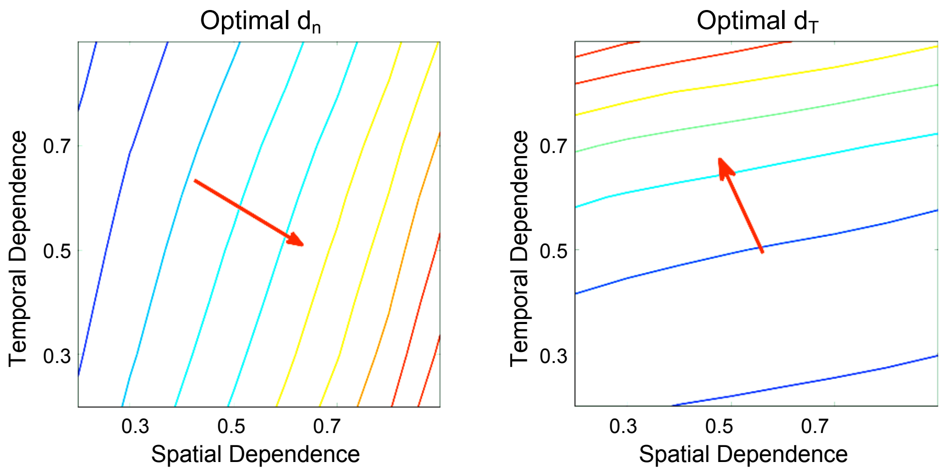

Figure 1.1: Level curves of $d_n^\star$ and $d_T^\star$ with respect to spatio-temporal dependences

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

$$d_n^\star = \left(\frac{4qK_q^2 B_{11}}{\eta_n \alpha_n \alpha_T \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C} nT \right)^{1/(2q+\eta_n+\eta_T)} \left(\frac{\eta_T B_{11}}{\eta_n B_{22}} \right)^{\eta_T/[2q(2q+\eta_n+\eta_T)]}, \tag{1.13}$$

$$d_T^\star = \left(\frac{4qK_q^2 B_{22}}{\eta_T \alpha_n \alpha_T \bar{\mathcal{K}}_1 \bar{\mathcal{K}}_2 C} nT \right)^{1/(2q+\eta_n+\eta_T)} \left(\frac{\eta_n B_{22}}{\eta_T B_{11}} \right)^{\eta_n/[2q(2q+\eta_n+\eta_T)]}. \tag{1.14}$$

(1.13) and (1.14) show that the degree of persistence in one dimension affects both $d_n^\star$ and $d_T^\star$ but in the opposite direction. For example, if a process becomes spatially persistent, $d_n^\star$ is increased to address the increasing bias, which comes from the usage of kernel in the spatial domain. But, the increase of $d_n^\star$, at the same time, magnifies the variance term. Therefore, in order to minimize AMSE*, $d_T^\star$ is decreased to moderate the inflation of the asymptotic variance. Figure 1 illustrates this relation of $d_n^\star$ and $d_T^\star$ with different dependence structure. The two graphs are the level curves of $d_n^\star$ and $d_T^\star$ as functions of λ and ρ, which determine the spatial and temporal persistence respectively in the following DGP:

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

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

Corollary 1. *Suppose Assumptions F5-A.17 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*

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

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 the one 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 Shinichi (2009) suggest a practical rule for the rectangular kernel estimator based on the AMSE criterion. Sun and Kaplan (2010) explore this problem rigorously and provide a bandwidth selection procedure that is testing optimal. By extending these preceding methods, we obtain the optimal bandwidth parameters for the rectangular kernel, $(d_{rec,n}^\star, d_{rec,T}^\star)$

which lead the rectangular kernel estimator to smaller $AMSE^*$ than any finite order kernel estimator we target.

Let $K_{tar}(\cdot)$ be the target kernel and $(d^{\star}_{tar,n}, d^{\star}_{tar,T})$ 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

$$d^{\star}_{rec,n} = d^{\star}_{tar,n} \left(\frac{\bar{\mathcal{K}}_{tar,1}}{\bar{\mathcal{K}}_{rec,1}} \right)^{1/\eta_n} \text{ and } d^{\star}_{rec,T} = d^{\star}_{tar,T} \left(\frac{\bar{\mathcal{K}}_{tar,2}}{\bar{\mathcal{K}}_{rec,2}} \right)^{1/\eta_T}, \tag{1.15}$$

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

As (A.9) is the function of unknown values such as B_{11}, B_{22} and C, they need to be estimated for implementation with given data in a parametric (e.g. Andrews, 1991; and Kim and Sun, 2010) or nonparametric way (e.g. Newey and West, 1994). In this paper, we suggest a parametric plug-in method. We consider the following four different spatiotemporal parametric models, which are introduced in Anselin (2001).

$$V^{(c)}_{(i,t)} = \rho_c \left[W^{(c)}_n V^{(c)}_{t-1} \right]_i + \tilde{\varepsilon}^{(c)}_{(i,t)}, \tag{1.16}$$

$$V^{(c)}_{(i,t)} = \lambda_c V^{(c)}_{(i,t-1)} + \rho_c \left[W^{(c)}_n V^{(c)}_{t-1} \right]_i + \tilde{\varepsilon}^{(c)}_{(i,t)} \tag{1.17}$$

$$V^{(c)}_{(i,t)} = \lambda_c V^{(c)}_{(i,t-1)} + \phi_c \left[W^{(c)}_n V^{(c)}_{t} \right]_i + \tilde{\varepsilon}^{(c)}_{(i,t)} \tag{1.18}$$

$$V^{(c)}_{(i,t)} = \lambda_c V^{(c)}_{(i,t-1)} + \phi_c \left[W^{(c)}_n V^{(c)}_{t} \right]_i + \rho_c \left[W^{(c)}_n V^{(c)}_{t-1} \right]_i + \tilde{\varepsilon}^{(c)}_{(i,t)} \tag{1.19}$$

where $\tilde{\varepsilon}^{(c)}_{(i,t)} \overset{i.i.d}{\sim} (0, \sigma_{cc})$ and $[W^{(c)}_n V^{(c)}_t]_i$ is the i^{th} element of vector $W^{(c)}_n V^{(c)}_t$. The spatial weight matrix $W^{(c)}_n$ is determined a priori and by convention it is row-standardized and its diagonal elements are zeros. For an illustrative purpose, let's consider the model in (1.16). It can be rewritten in a recursive way as follows:

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

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

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

and

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

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

$$
\hat{J}\left(c,d\right) = \frac{\hat{\sigma}_{cd}}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\hat{\tilde{R}}_{(i,t)}^{(c)}\right) \left(\hat{\tilde{R}}_{(j,s)}^{(d)}\right)', \tag{1.20}
$$

$$
\hat{b}_1^{(q)}\left(c,d\right) = \frac{\hat{\sigma}_{cd}}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\hat{\tilde{R}}_{(i,t)}^{(c)}\right) \left(\hat{\tilde{R}}_{(j,s)}^{(d)}\right)' d_{ij}^q, \tag{1.21}
$$

$$
\hat{b}_2^{(q)}\left(c,d\right) = \frac{\hat{\sigma}_{cd}}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\hat{\tilde{R}}_{(i,t)}^{(c)}\right) \left(\hat{\tilde{R}}_{(j,s)}^{(d)}\right)' d_{ts}^q, \tag{1.22}
$$

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

$$
\left(\hat{d}_n, \hat{d}_T\right) = \underset{d_n, d_T}{\arg\min}\ 2K_q^2 \left(\frac{\hat{B}_{11}}{d_n^{2q}} + \frac{\hat{B}_{22}}{d_T^{2q}}\right) + \frac{\ell_n \ell_T}{nT} \hat{C}. \tag{1.23}
$$

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

$$\hat{d}_n = \left(\frac{4qK_q^2\hat{B}_{11}}{\eta_n\alpha_n\alpha_T\bar{\mathcal{K}}_1\bar{\mathcal{K}}_2\hat{C}}nT\right)^{1/(2q+\eta_n+\eta_T)}\left(\frac{\eta_T\hat{B}_{11}}{\eta_n\hat{B}_{22}}\right)^{\eta_T/[2q(2q+\eta_n+\eta_T)]}, \tag{1.24}$$

$$\hat{d}_T = \left(\frac{4qK_q^2\hat{B}_{22}}{\eta_T\alpha_n\alpha_T\bar{\mathcal{K}}_1\bar{\mathcal{K}}_2\hat{C}}nT\right)^{1/(2q+\eta_n+\eta_T)}\left(\frac{\eta_n\hat{B}_{22}}{\eta_T\hat{B}_{11}}\right)^{\eta_n/[2q(2q+\eta_n+\eta_T)]}. \tag{1.25}$$

It also follows

$$\hat{d}_{rec,n} = \hat{d}_{tar,n}\left(\frac{\bar{\mathcal{K}}_{tar,1}}{\bar{\mathcal{K}}_{rec,1}}\right)^{1/\eta_n} \text{ and } \hat{d}_{rec,T} = \hat{d}_{tar,T}\left(\frac{\bar{\mathcal{K}}_{tar,2}}{\bar{\mathcal{K}}_{rec,2}}\right)^{1/\eta_T}. \tag{1.26}$$

Since the models in (1.17), (1.18) and (1.19) 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 (1.16). While the OLS estimator is consistent for (1.17), it is not for (1.18) and (1.19) due to endogeneity of $[W_n^{(c)}V_t^{(c)}]_i$. For these models, we can have consistent estimators using QMLE as follows:

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

See Yu, de Jong and Lee (2008) for detail. In fact, however, the simple OLS can still be used for (1.18) and (1.19). 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.

1.5 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 (1.5) and (1.6) under the increasing smoothing asymptotics. We also derive the optimal bandwidth parameters for DK and KS estimators using the AMSE criterion.

1.5.1 CCE

The CCE is defined as

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

The critical condition for this estimator to be consistent is that each variable from two different individuals (or clusters) is uncorrelated, i.e. $EV_{(i,t)} V'_{(j,s)} = 0$ if $i \neq j$. Under this condition, $\hat{J}_{nT}^{A}$ is robust to heteroskedasticity and arbitrary forms of serial correlation. Our 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 F5, F6, I10 and I11 hold.*

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

(b) *If Assumption F12 holds, then* $\sqrt{n}\left(\hat{J}^A_{nT} - J^A_{nT}\right) = O_p(1)$ *and* $\sqrt{n}\left(\hat{J}^A_{nT} - \tilde{J}^A_{nT}\right) = o_p(1)$.

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

1.5.2 DK estimator

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

$$\hat{J}^{DK}_{nT} = \frac{1}{nT} \sum_{i,j=1}^{n} \sum_{t,s=1}^{T} K\left(\frac{d_{ts}}{d_T}\right) \hat{V}_{(i,t)} \hat{V}'_{(j,s)}, \tag{1.27}$$

For the asymptotic properties, we introduce the following assumptions in place of Assumptions F7 and F10.

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

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

for all n, T.

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

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

Compared with Assumption F7, Assumption I12 is sufficient for $\hat{J}^{DK}_{nT}$ because $\hat{J}^{DK}_{nT}$ is not involved with the bias that arises from a kernel in the spatial dimension. Theorem 3 below states the asymptotic properties of $\hat{J}^{DK}_{T}$.

Theorem 3. *Suppose that Assumptions F5, F6, F11(i) and (ii), I12 and I13 hold, and* $d_T \to \infty$, $\ell_T = o(T)$.

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

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

(c) *If Assumption F12 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 A.17,*

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

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

$$d_T^{DK} = \left(\frac{2qK_q^2 B_{22}}{\eta_T\alpha_T\bar{\mathcal{K}}_2 C}T\right)^{1/(2q+\eta_T)}, \tag{1.28}$$

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

1.5.3 KS estimator

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

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

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

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

for all n, T.

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

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

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

Theorem 4. *Suppose that Assumptions F5, F6, F9, F11, 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 F12 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 A.17,*

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

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

$$d_n^{\star KS} = \left(\frac{2qK_q^2 B_{11}}{\eta_n \alpha_n \bar{\mathcal{K}}_1 C} n\right)^{1/(2q+\eta_n)}. \tag{1.29}$$

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

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

1.6.1 Flexibility

$\hat{J}_{nT}$ is flexible in the sense that it includes the estimators in the previous section as special cases, reducing to each of them with certain choice of the bandwidths and kernel function. In order to illustrate the flexibility, we 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 $\hat{J}_{nT} - \hat{J}_{nT}^{KS} = o_p(1)$.*

Proofs are given in the appendix. The flexibility of our estimator relies on the property that the rectangular kernel does not downweigh the covariances between spatially or serially remote units. 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.

1.6.2 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 form of spatial correlation. For the same reason, $\hat{J}_{nT}^{KS}$ tends to perform better than the others, if a process is serially 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 estimator. As illustrated in Figure 2, adaptiveness arises from the flexibility and automatic bandwidth selection procedure. In case that a process is spatially highly persistent, the automatic bandwidth selection procedure yields large $\hat{d}_n$ so that $\hat{J}_{nT}$ gets close to $\hat{J}_{nT}^{DK}$. Analogously, $\hat{J}_{nT}$ becomes close to $\hat{J}_{nT}^{KS}$ 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 (1.15) meets the requirement, as the selected bandwidths from (1.15) are the proportional to those from (A.9).[6]

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

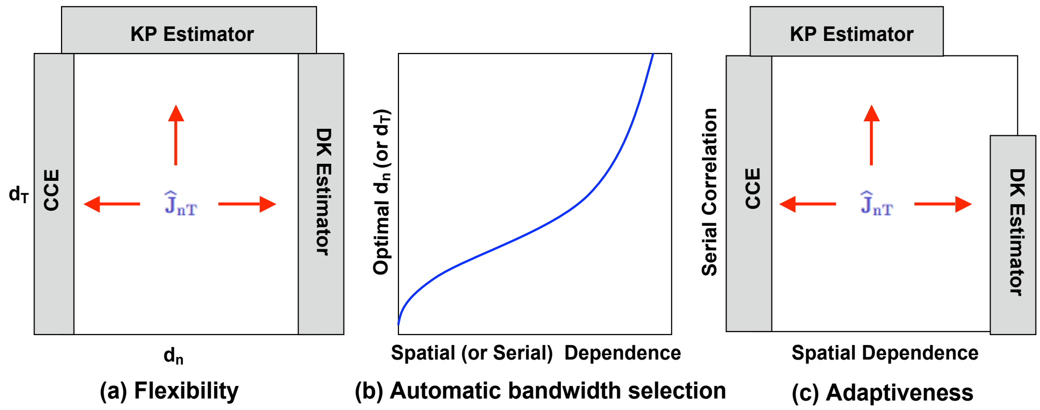

Figure 1.2: Adaptiveness of $\hat{J}_{nT}$

1.7 Fixed smoothing asymptotics

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

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

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

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

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

We maintain the following high level assumptions.

Assumption F1. *The functional central limit theorem*
Let $\Lambda\Lambda' = J$ *and* $W_p(r_1, r_2, \tau) = \left(W^{(1)}(r_1, r_2, \tau), \ldots, W^{(p)}(r_1, r_2, \tau)\right)'$ *be a p-dimensional independent Wiener process with covariance given by*

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

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

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

holds for all $(r_1, r_2, \tau) \in [0,1]^3$.

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

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

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

$$\sqrt{nT}\left(\hat{\beta} - \beta\right) = \left(\frac{1}{L_n M_n T} \sum_{i_1=1}^{L_n} \sum_{i_2=1}^{M_n} \sum_{t=1}^{T} \tilde{X}^{*}_{(i_1,i_2,t)} \tilde{X}^{*\prime}_{(i_1,i_2,t)}\right)^{-1} \quad (1.30)$$

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

$$\xrightarrow{d} Q\Lambda W_p(1,1,1) := \Lambda^{*} W_p(1,1,1). \quad (1.31)$$

Therefore,

$$\begin{aligned}
&\frac{1}{\sqrt{L_nM_nT}}\sum_{i_1=1}^{[r_1L_n]}\sum_{i_2=1}^{[r_2M_n]}\sum_{t=1}^{[\tau T]}\hat{V}^*_{(i_1,i_2,t)} \\
&=\frac{1}{\sqrt{L_nM_nT}}\sum_{i_1=1}^{[r_1L_n]}\sum_{i_2=1}^{[r_2M_n]}\sum_{t=1}^{[\tau T]}V^*_{(i_1,i_2,t)}-\frac{1}{L_nM_nT}\sum_{i_1=1}^{[r_1L_n]}\sum_{i_2=1}^{[r_2M_n]}\sum_{t=1}^{[\tau T]}\tilde{X}^*_{(i_1,i_2,t)}\tilde{X}^{*\prime}_{(i_1,i_2,t)} \\
&\times\sqrt{L_nM_nT}\left(\hat{\beta}-\beta\right) \\
&\xrightarrow{d}\Lambda\left[W_p\left(r_1,r_2,\tau\right)-r_1r_2\tau W_p\left(1,1,1\right)\right] \\
&:=\Lambda B_p\left(r_1,r_2,\tau\right),
\end{aligned}$$

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

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

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

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

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

$$\hat{J}_{nT}:=\frac{1}{L_nM_nT}\sum_{i_1,j_1=1}^{L_n}\sum_{i_2,j_2=1}^{M_n}\sum_{t,s=1}^{T}\mathbb{K}_b\left(\frac{i_1-j_1}{L_n},\frac{i_2-j_2}{M_n},\frac{t-s}{T}\right)\hat{V}^*_{(i_1,i_2,t)}\hat{V}^*_{(j_1,j_2,s)} \tag{1.32}$$

where

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

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 function of two variables, from K, a function of a single variable. Note that K_n does not depend on the sample size n.

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

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

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

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

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

It follows from Assumption F4(i) that

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

Using the representation, we can obtain the following results.

Proposition 2. *Under Assumptions F1 - F3,* [7]

$$\hat{J}_{nT}\quad\xrightarrow{d}\Lambda\left[\int_0^1\mathbb{K}_b\left(r_1-v_1,r_2-v_2,\tau-\kappa\right)dB_p\left(r_1,r_2,\tau\right)dB_p'\left(v_1,v_2,\kappa\right)\right]\Lambda'. \tag{1.33}$$

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

[7] Let $\int=\int\cdots\int$ for simplification.

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

$$
\begin{aligned}
&\mathbb{K}_b^*((r_1,r_2,\tau),(v_1,v_2,\kappa)) \\
&= \mathbb{K}_b(r_1-v_1,r_2-v_2,\tau-\kappa) - \int_0^1 \mathbb{K}_b(x_1-v_1,y_1-v_2,z_1-\kappa)\,dx_1dy_1dz_1 \\
&- \int_0^1 \mathbb{K}_b(r_1-x_2,r_2-y_2,\tau-z_2)\,dx_2dy_2dz_2 \\
&+ \int_0^1 \mathbb{K}_b(x_1-x_2,y_1-y_2,z_1-z_2)\,dx_1dy_1dz_1dx_2dy_2dz_2.
\end{aligned}
$$

Using $\mathbb{K}_b^*(\cdot,\cdot)$, (1.33) is equivalent to

$$
\Lambda\left[\int_0^1 \mathbb{K}_b^*((r_1,r_2,\tau),(v_1,v_2,\kappa))\,dW_g(r_1,r_2,\tau)\,dW_p'(v_1,v_2,\kappa)\right]\Lambda'. \tag{1.34}
$$

In (1.34), the integration is with respect to the standard Wiener process because the centered kernel function captures the estimation uncertainty in $\hat{\beta}$. With (1.31) and (1.34), we can show that under H_0,

$$
\begin{aligned}
F_{nT} &= \sqrt{nT}\left[R\left(\hat{\beta}-\beta_0\right)\right]'\left(R\hat{Q}_{nT}\hat{J}_{nT}\hat{Q}_{nT}'R'\right)^{-1}\sqrt{nT}\left[R\left(\hat{\beta}-\beta_0\right)\right]/g \\
&\xrightarrow{d} (R\Lambda^*W_p(1,1,1))' \\
&\times\left(R\Lambda^*\left[\int_0^1 \mathbb{K}_b^*((r_1,r_2,\tau),(v_1,v_2,\kappa))\,dW_p(r_1,r_2,\tau)\,dW_p'(v_1,v_2,\kappa)\right]\Lambda^{*\prime}R'\right)^{-1} \\
&\times(R\Lambda^*W_p(1,1,1))/g \\
&\overset{d}{=} W_g'(1,1,1)\left[\int_0^1 \mathbb{K}_b^*((r_1,r_2,\tau),(v_1,v_2,\kappa))\,dW_g(r_1,r_2,\tau)\,dW_g'(v_1,v_2,\kappa)\right]^{-1} \\
&\times W_g(1,1,1)/g \\
&:= F_\infty(g,b),
\end{aligned} \tag{1.35}
$$

where the equality in distribution holds because $R\Lambda^*W_p(x,y,z)\overset{d}{=}DW_g(x,y,z)$ for a Wiener process $W_g(x,y,z)$ and some $g\times g$ matrix D such that $DD'=RQJQ'R'$.

1.7.2 Expansion of the distribution and F-approximation

We present the asymptotic expansion of the distribution of $F_\infty(g,b)$ in (1.35) and establish the validity of a standard F-approximation. The distribution of

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

Let

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

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

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

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 $G_g(zv_{11.2})$ around $G_g(z)$ and computing the moments of $v_{11.2}$, we can prove the following theorem.

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

$$P\left\{ gF_\infty(g,b) \le z \right\} = G_g(z) + A(z)\, b_1 b_2 b_3 + o(b_1 b_2 b_3)$$

where

$$A(z) = G_g''(z)\, z^2 c_2 - G_g'(z)\, z \left[c_1 + (g-1) c_2 \right],$$
$$c_1 = \int_{-\infty}^{\infty} \mathbb{K}(x,y,z)\, dxdydz \text{ and } c_2 = \int_{-\infty}^{\infty} \mathbb{K}^2(x,y,z)\, dxdydz.$$

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

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

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

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

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

Using this representation, we have

$$\begin{aligned}
&\int_0^1 \mathbb{K}_b^*\left((r_1, r_2, \tau), (v_1, v_2, \kappa)\right) dW_g\left(r_1, r_2, \tau\right) dW_g'\left(v_1, v_2, \kappa\right) \\
&= \sum_{k,\ell,m,k',\ell',m'=1}^{\infty} \lambda_{k\ell mk'\ell'm'} \xi_{b,k\ell m} \xi'_{b,k'\ell'm'}, \qquad (1.36)
\end{aligned}$$

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

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

$$\lambda_{h_1 h_2 h_3 h_1' h_2' h_3'} = \lambda_{hh'} \text{ and } \xi_{b,h_1 h_2 h_3} = \xi_{b,h}.$$

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

$$gF_\infty(g, b) = \phi' \left[\sum_{k=1}^{\infty} \lambda_k \zeta_k \zeta_k' \right]^{-1} \phi, \tag{1.37}$$

where ζ_k is independent of ϕ for all k.

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

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

where

$$\mu_1 = \sum_{k=1}^{\infty} \lambda_k = \int_0^1 \mathbb{K}_b^* \left((r_1, r_2, \tau), (r_1, r_2, \tau) \right) dr_1 dr_2 d\tau.$$

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

$$E\Psi = I_g \tag{1.38}$$

$$E\Psi D\Psi = \frac{1}{K} tr(D) I_g + \left(1 + \frac{1}{K}\right) D \tag{1.39}$$

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

$$E\Phi D\Phi = \frac{\mu_2}{\mu_1^2} tr(D) I_g + \left(1 + \frac{\mu_2}{\mu_1^2}\right) D \tag{1.40}$$

where

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

From (1.39) and (1.40), it is reasonable to approximate the distribution of Φ by $\mathbb{W}_g(I_g, K)/K$ with

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

K is called equivalent degree of freedom (EDF) of $\hat{J}_{nT}$. Combining this results with (1.37), we have

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

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

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

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

$$\frac{\mu_1 (K - g + 1)}{K} F_\infty(g, b) \overset{d}{\approx} F_{g, K-g+1}. \tag{1.41}$$

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

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

Proofs are given in the supplementary appendix.[8] Lemma 1 implies

$$\frac{\mu_1 (K - g + 1)}{K} = \frac{1}{1 + b_1 b_2 b_3 (c_1 + (g-1) c_2)} + o(b_1 b_2 b_3). \tag{1.42}$$

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

The following theorem summarizes the F-approximation.

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

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

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

where

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

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

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

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

Proof is given in the appendix. Some comments on Theorem 6 are in order. We use K^* in place of $K - g + 1$ for the second degree of freedom in the F-approximation. This modification ensures that the variance of the F distribution exists. Let $F_\infty^\alpha(g,b)$ and F_{g,K^*}^α denote the $1-\alpha$ quantiles of the distribution of $F_\infty(g,b)$ and the F distribution with the degrees of freedom g and K^*. Theorem 6 suggests that for the F-test version of Wald statistic, F_{nT}, we use

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

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

As the EDF is proportional to $1/(b_1 b_2 b_3)$, we may want to choose the large EDF by increasing the degree of smoothing for more powerful inference. However, for a given sample size, small $b_1 b_2 b_3$ introduces a large bias of $\hat{J}_{nT}$, which will cause F_{nT}/ν to deviate substantially from F_{g,K^*} distribution. At the same time, there can be a significant loss in power if we choose too large bandwidths.

1.8 Monte Carlo simulation

In this section, we provide some simulation evidence on the finite sample performance of our estimator and the associated testing procedures. 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 estimators using the RMSE criterion and the coverage accuracy of the confidence intervals (CIs). We examine the robustness to the measurement errors in economic distance. It is also investigated how the number of restrictions being tested affects the performance of the Wald tests under the two different smoothing asymptotics.

We assume a lattice structure, in which each individual is located on a square grid of integers over time. 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 $u_t = (u_{1t}, \ldots, u_{nt})'$. The parameter λ determines the strength of the temporal correlation of data. X_t and u_t are also spatially dependent as $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{nt})'$ and $\nu_t = (\nu_{1t}, \ldots, \nu_{nt})'$ follow spatial AR(1) processes with parameter θ. The values considered for the model parameters λ and θ are 0, 0.3, 0.6 and 0.9. $\tilde{W}_n$ is a contiguity matrix and individuals i and j are neighbors if $d_{ij} = 1$. Following convention, it is row-standardized and its diagonal elements are zero.

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

For bandwidth selection, the formulas in (1.28) and (1.29) are considered with the AR(1) and spatial AR(1) models for $\hat{J}_{nT}^{DK}$ and $\hat{J}_{nT}^{KS}$ respectively. For $\hat{J}_{nT}$

with the Parzen kernel, we choose the bandwidths based on (1.13) and (1.14) using the spatiotemporal parametric model in (1.18). We use the Parzen kernel as the target kernel to obtain the bandwidths for $\hat{J}_{nT}$ with the rectangular kernel. One concern of the rectangular kernel estimator is that it is not positive semi-definite. However, we can attain positive semi-definiteness with a simple modification suggested by Politis (2010). As $\hat{J}_{nT}$ is symmetry, $\hat{J}_{nT} = \hat{U}\hat{\Lambda}\hat{U}'$, where $\hat{U}$ is an orthogonal matrix and $\hat{\Lambda} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\hat{J}_{nT}$. Let $\hat{\Lambda}^+ = \mathrm{diag}(\hat{\lambda}_1^+, \ldots, \hat{\lambda}_p^+)$ where $\hat{\lambda}_s^+ = \max(\hat{\lambda}_s, 0)$. Then, we define our modified estimator as $\hat{J}_{nT}^+ = \hat{U}\hat{\Lambda}^+\hat{U}'$. As each eigenvalue of $\hat{J}_{nT}$ is nonnegative, it is positive semi-definite.

W_n is the contiguity matrix in which individuals i and j are neighbors if $d_{ij} = 1$. We set $\eta = 2$ and $\ell_n = \pi d_n^2$. Note that the approximating parametric models for $\hat{J}_{nT}^{KS}$ and $\hat{J}_{nT}$ are mis-specified whereas the AR(1) model for $\hat{J}_{nT}^{DK}$ is correctly specified. We estimate parameters in (1.18) and (1.29) with the QMLE.

We consider three different sample sizes; (i) small T and n; $T = 15, n = 49$ (7×7), (ii) large T and small n; $T = 50, n = 49$, and (iii) small T and large n; $T = 15, n = 196$ (14×14).

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

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

PHAC, CCE, DK and 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 for 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 and at infeasible optimal bandwidth parameters. First, we find that $\hat{J}_{nT}$ performs better than $\hat{J}_{nT}^{DK}$ in almost all the cases. When spatial dependence is absent or weak, $\hat{J}_{nT}$ has substantially smaller ratio than $\hat{J}_{nT}^{DK}$. This is because the DK estimator loses information in the spatial dimension with the cross-sectional averaging. When $\theta = 0.9$, both the estimators are not much different. Especially $\hat{J}_{nT}$ with the rectangular kernel is as accurate as and sometimes better than $\hat{J}_{nT}^{DK}$. This implies that adaptiveness works well in this setting. Second, the increase of n reduces only the ratio of $\hat{J}_{nT}$ while increasing T improves the performance of both the estimators. Finally, the AMSE* criterion tends to effectively controls the RMSE of $\hat{J}_{nT}$.

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

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

The DGP2 is used with a univariate regressor ($p = 1$). For the testing with $\hat{J}_{nT}$, we use both the fixed smoothing asymptotics and increasing smoothing asymptotics. The simulation results verify our theoretical results. First, we compare $\hat{J}_{nT}$ with the other estimators under increasing smoothing asymptotics. PHAC performs as well as CCE when $\theta = 0$ and is significantly more accurate when there exists spatial dependence. Compared with the KS, the CIs associated with PHAC is more precise unless the process is temporally highly persistent. Even under strong temporal persistence ($\lambda = 0.9$) PHAC is almost as good as KS especially if n is small. Both the PHAC and KS become more accurate with large n, but only the performance of PHAC improves when T increases. Comparison with DK is quite similar to the RMSE criterion case. The PHAC tends to perform better than DK except in some cases with $\theta = 0.9$. Even when $\theta = 0.9$, the ECP of PHAC especially with the rectangular kernel is very close to that of DK. Second, Table 2 compares the performance of PHAC using increasing smoothing asymptotics and fixed smoothing asymptotics. The results indicate that fixed smoothing asymptotic approximation is substantially more accurate than increasing smoothing asymptotics. The difference increases as the process becomes more persistent. When $\theta = 0.9$, $\lambda = 0.9$ and $T = 15, n = 49$, the ECP of the PHAC with the Parzen kernel under fixed smoothing asymptotics is 79.6% but it is 64.4% under increasing smoothing asymptotics. This result shows that fixed smoothing asymptotics captures the demeaning bias arising from the estimation errors of $\hat{\beta}$ while it is ignored under increasing smoothing asymptotics. Third, Table 2 provides a strong evidence that the rectangular kernel performs better than the finite order kernel under fixed smoothing asymptotics. Especially in our simulation, PHAC with the rectangular kernel is very robust to spatial dependence so that size distortion does not increase with the spatial dependence. This size advantage of the rectangular kernel may arise from its bias reducing property and adaptiveness. Finally, Table 2 shows that our testing procedure based on fixed smoothing asymptotics is reasonably robust to measurement errors. As economic distance data is likely to be error contaminated in practice, the robustness to measurement errors is a highly desirable property. Comparing PHAC with PHAC$_e$, we see that the performance of PHAC$_e$ is quite close to that of PHAC in most cases.

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

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

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

1.9 Conclusion

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

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

1.10 Acknowldegements

Chapter 1 is coauthored with Yixiao Sun.

Table 1.1: RMSE/Estimand with $\hat{J}_{nT}$ and $\hat{J}_{nT}^{DK}$

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

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

Table 1.2: RMSE/Estimand with $\hat{J}_{nT}$ and $\hat{J}_{nT}^{DK}$ - Cont.

λ		θ								
		0.0	0.3	0.6	0.9		0.0	0.3	0.6	0.9
T=15, n=196										
0.0		0.05	0.13	0.18	0.29		0.43	0.20	0.21	0.27
0.3	$\hat{J}_{nT}$	0.09	0.24	0.33	0.54	$\hat{J}_{nT}$	0.07	0.24	0.32	0.47
0.6	$(\hat{d}_n, \hat{d}_T)$	0.13	0.30	0.42	0.57	$(d_n^\star, d_T^\star)$	0.12	0.29	0.39	0.56
0.9	(PA)	0.29	0.43	0.52	0.72	(PA)	0.28	0.43	0.51	0.69
0.0		0.07	0.15	0.21	0.30		1.00	0.34	0.36	0.35
0.3	$\hat{J}_{nT}$	0.11	0.27	0.37	0.62	$\hat{J}_{nT}$	0.09	0.23	0.28	0.41
0.6	$(\hat{d}_n, \hat{d}_T)$	0.15	0.36	0.51	0.62	$(d_n^\star, d_T^\star)$	0.10	0.26	0.37	0.50
0.9	(RE)	0.22	0.43	0.55	0.74	(RE)	0.10	0.35	0.45	0.66
0.0		0.47	0.43	0.48	0.47		0.37	0.34	0.36	0.35
0.3	$\hat{J}_{nT}^{DK}$	0.53	0.56	0.55	0.55	$\hat{J}_{nT}^{DK}$	0.54	0.56	0.56	0.55
0.6	$(\hat{d}_T^{DK})$	0.68	0.70	0.69	0.69	(d_T^{DK})	0.70	0.72	0.71	0.70
0.9		0.88	0.87	0.88	0.88		0.89	0.88	0.89	0.89

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

Table 1.3: ECPs of Nominal 95% CIs with Alternative Covariance Estiamtors

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	KP	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

'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 1.4: ECPs of Nominal 95% CIs with Alternative Covariance Estiamtors - Cont.

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=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	KP	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

'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 1.5: ECPs of Nominal 95% CIs with Alternative Covariance Estimators - Cont.

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=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	KP	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 1.6: ECPs of Nominal 95% CIs with Different Number of Restrictions

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

See note to Table 2.

1.11 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

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

where

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

For C_{1nT}, under Assumptions F5 and F6

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

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

$$C_{2nT} := D_{1nT} + D_{2nT} + D_{3nT} + D_{4nT} + D_{5,nT} \tag{A.2}$$

where

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

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

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

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

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

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

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

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

and the next step is to prove that

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

For (A.2), let $\gamma^{(d_1d_2)}_{(\bar{\imath}\bar{t},b_{(a)}v_{(u)})} = (\ell_n\ell_T)^{-1}\sum_{h_{(i)}=1}^{\ell_{i,n}}\sum_{w_{(t)}=1}^{\ell_{t,T}}\gamma^{(d_1d_2)}_{(h_{(i)}w_{(t)},b_{(a)}v_{(u)})}$. Then,

$$\frac{1}{nT\ell_n\ell_T} \sum_{i,a\in E_n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t,u\in E_T} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right) K^2\left(\frac{d_{ts_{(t)}}}{d_T}\right)$$
$$\times \gamma^{(c_1c_2)}_{(it,au)} \gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})} \tag{A.5}$$
$$= \frac{1}{nT\ell_n\ell_T} \sum_{i,a\in E_n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t,u\in E_T} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right) K^2\left(\frac{d_{ts_{(t)}}}{d_T}\right)$$
$$\times \gamma^{(c_1c_2)}_{(it,au)} \gamma^{(d_1d_2)}_{\left(\bar{\imath}\bar{t},b_{(a)}v_{(u)}\right)}$$
$$+ \frac{1}{nT\ell_n\ell_T} \sum_{i,a\in E_n} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{t,u\in E_T} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right) K^2\left(\frac{d_{ts_{(t)}}}{d_T}\right) \gamma^{(c_1c_2)}_{(it,au)}$$
$$\times \left(\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})} - \gamma^{(d_1d_2)}_{\left(\bar{\imath}\bar{t},b_{(a)}v_{(u)}\right)}\right)$$
$$:= L_{1nT} + L_{2nT}. \tag{A.6}$$

L_{1nT} can be decomposed as

$$L_{1nT} := G_{1nT} + G_{2nT}$$

where

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

$$G_{2nT} = \frac{1}{nT} \sum_{i,a \in E_n} \sum_{t,u \in E_T} \gamma_{(it,au)}^{(c_1 c_2)} \big[1\{d_{ia} > c_n, d_{ut} > c_T\} + 1\{d_{ia} > c_n, d_{ut} \le c_T\}$$

$$+ 1\{d_{ia} \le c_n, d_{ut} > c_T\} \big] \frac{1}{\ell_n \ell_n} cov \left(\sum_{j:d_{ij} \le d_n} \sum_{s:d_{ts} \le d_T} V_{(j,s)}^{(d_1)}, \sum_{b:d_{ab} \le d_n} \sum_{v:d_{uv} \le d_T} V_{(b,v)}^{(d_2)} \right)$$

$$\times \left(\frac{1}{\ell_n} \sum_{j=1}^{n} K^2 \left(\frac{d_{ij}}{d_n} \right) \right) \left(\frac{1}{\ell_T} \sum_{s=1}^{T} K^2 \left(\frac{d_{ts}}{d_T} \right) \right)$$

$$= o(1)$$

as c_n,$c_T \to \infty$. Thus, it suffices to consider G_{1nT}. When $d_{ia} \le c_n$ and $d_{tu} \le c_T$, we have

$$cov \left(\sum_{j:d_{ij} \le d_n} \sum_{s:d_{ts} \le d_T} V_{(j,s)}^{(d_1)}, \sum_{b:d_{ab} \le d_n} \sum_{v:d_{uv} \le d_T} V_{(b,v)}^{(d_2)} \right)$$

$$= cov \left(\sum_{j:d_{ij} \le d_n} \sum_{s:d_{ts} \le d_T} V_{(j,s)}^{(d_1)}, \sum_{j:d_{ij} \le d_n} \sum_{s:d_{ts} \le d_T} V_{(j,s)}^{(d_2)} \right)$$

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

but

$$\sum_{b:d_{ab} \le d_n} \sum_{v:d_{uv} \le d_T} V_{(b,v)}^{(d_2)} - \sum_{b:d_{ib} \le d_n} \sum_{v:d_{tv} \le d_T} V_{(b,v)}^{(d_2)}$$

$$= \sum_{b:d_{ab} \le d_n} \sum_{v:d_{uv} \le d_T, d_{tv} > d_T} V_{(b,v)}^{(d_2)} + \sum_{b:d_{ab} \le d_n, d_{ib} > d_n} \sum_{v:d_{tv} \le d_T} V_{(b,v)}^{(d_2)}.$$

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

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

$$\le \frac{1}{\ell_n \ell_T} \sum_{j:d_{ij} \le d_n} \sum_{b:d_n < d_{ib} \le d_n + c_n} \sum_{s:d_{ts} \le d_T} \sum_{v:d_{tv} \le d_T} \left| E V_{(j,s)}^{(d_1)} V_{(b,v)}^{(d_2)} \right| = o(1),$$

by choosing c_n such that $\sum_{b=1}^{n} 1\{d_n < d_{ib} \le d_n + c_n\} = o(\ell_n)$ for all i. Similarly,

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

Hence

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

$$= \frac{1}{\ell_n \ell_T} cov\left(\sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T} V_{(j,s)}^{(d_1)}, \sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T} V_{(j,s)}^{(d_2)}\right) + o\left(1\right) \qquad \text{(A.7)}$$

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

Now under Assumption F10, we have

$$G_{1nT} = \frac{1}{nT}\sum_{i,a\in E_n}\sum_{t,u\in E_T}\gamma_{(it,au)}^{(c_1c_2)}\frac{1}{\ell_n\ell_T} cov\left(\sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T} V_{(j,s)}^{(d_1)}, \sum_{j:d_{ij}\leq d_n}\sum_{s:d_{ts}\leq d_T} V_{(j,s)}^{(d_2)}\right)$$

$$\times\left(\frac{1}{\ell_n}\sum_{j=1}^{n}K^2\left(\frac{d_{ij}}{d_n}\right)\right)\left(\frac{1}{\ell_T}\sum_{s_{(t)}=1}^{T}K^2\left(\frac{d_{ts}}{d_T}\right)\right)(1+o(1))$$

$$\to \bar{\mathcal{K}}_1\bar{\mathcal{K}}_2 J\left(c_1,c_2\right)J(d_1,d_2).$$

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

For L_{2nT} in (A.4), the first step is to show

$$\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left(K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right)K^2\left(\frac{d_{ts_{(t)}}}{d_T}\right)\right.$$

$$\left.-K^2\left(\frac{d_{ia}}{d_n}\right)K^2\left(\frac{d_{tu}}{d_T}\right)\right)\gamma_{(it,au)}^{(c_1c_2)}\left(\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}-\gamma^{(d_1d_2)}_{\left(\bar{i}\bar{t},b_{(a)}v_{(u)}\right)}\right) = o(1), \qquad \text{(A.8)}$$

and the second step is to prove

$$\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}K^2\left(\frac{d_{ia}}{d_n}\right)K^2\left(\frac{d_{tu}}{d_T}\right)$$

$$\times\gamma_{(it,au)}^{(c_1c_2)}\left(\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}-\gamma^{(d_1d_2)}_{\left(\bar{i}\bar{t},b_{(a)}v_{(u)}\right)}\right) = o(1). \qquad \text{(A.9)}$$

For (A.8),

$$\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left(K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right)K^2\left(\frac{d_{ts_{(t)}}}{d_T}\right)\right.$$

$$\left.-K^2\left(\frac{d_{ia}}{d_n}\right)K^2\left(\frac{d_{tu}}{d_T}\right)\right)\gamma_{(it,au)}^{(c_1c_2)}\left(\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}-\gamma^{(d_1d_2)}_{\left(\bar{i}\bar{t},b_{(a)}v_{(u)}\right)}\right)$$

$$\leq M_{1nT}+M_{2nT}+M_{3nT}+M_{4nT}$$

where

$$M_{1nT} := \frac{1}{nT} \sum_{(i,a)\in\mathcal{F}_1} \sum_{(t,u)\in\mathcal{G}_1} \left|\gamma_{(it,au)}^{(c_1c_2)}\right| \left(\frac{1}{\ell_n \ell_T} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} \left|\gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)} - \gamma_{(\bar{\imath}\bar{t},b_{(a)}v_{(u)})}^{(d_1d_2)}\right| \right),$$

$$M_{2nT} := \frac{1}{nT} \sum_{(i,a)\in\mathcal{F}_2} \sum_{(t,u)\in\mathcal{G}_1} \left|\gamma_{(it,au)}^{(c_1c_2)}\right| \left(\frac{1}{\ell_n \ell_T} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} \left|\gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)} - \gamma_{(\bar{\imath}\bar{t},b_{(a)}v_{(u)})}^{(d_1d_2)}\right| \right),$$

$$M_{3nT} := \frac{1}{nT} \sum_{(i,a)\in\mathcal{F}_1} \sum_{(t,u)\in\mathcal{G}_2} \left|\gamma_{(it,au)}^{(c_1c_2)}\right| \left(\frac{1}{\ell_n \ell_T} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} \left|\gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)} - \gamma_{(\bar{\imath}\bar{t},b_{(a)}v_{(u)})}^{(d_1d_2)}\right| \right),$$

$$M_{4nT} := \frac{1}{nT} \sum_{(i,a)\in\mathcal{F}_2} \sum_{(t,u)\in\mathcal{G}_2} \left|\gamma_{(it,au)}^{(c_1c_2)}\right| \left(\frac{1}{\ell_n \ell_T} \sum_{j_{(i)}=1}^{\ell_{i,n}} \sum_{b_{(a)}=1}^{\ell_{a,n}} \sum_{s_{(t)}=1}^{\ell_{t,T}} \sum_{v_{(u)}=1}^{\ell_{u,T}} \left|\gamma_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}^{(d_1d_2)} - \gamma_{(\bar{\imath}\bar{t},b_{(a)}v_{(u)})}^{(d_1d_2)}\right| \right),$$

and

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

in which $f_n/d_n = O(1)$ and $g_T/d_T = O(1)$. It is straightforward to show

$$M_{1nT} = O\left(\frac{\ell_n \ell_T}{nT}\right), \ M_{2nT} = O\left(\frac{\ell_T}{T}\right) \text{ and } M_{3nT} = O\left(\frac{\ell_n}{n}\right).$$

For M_{4nT},

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

Therefore, (A.8) holds.

We also obtain (A.9) because

$$\frac{1}{\ell_n\ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left(\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}-\gamma^{(d_1d_2)}_{\left(\bar{\imath}\bar{t},b_{(a)}v_{(u)}\right)}\right)$$
$$=\left[\frac{1}{\ell_n\ell_T}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right.$$
$$\left.-\left(\frac{1}{\ell_n\ell_T}\right)^2\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{s_{(t)}=1}^{\ell_{t,T}}\left(\sum_{h_{(i)}=1}^{\ell_{i,n}}\sum_{w_{(t)}=1}^{\ell_{t,T}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\gamma^{(d_1d_2)}_{(h_{(i)}w_{(t)},b_{(a)}v_{(u)})}\right)\right]$$
$$=\left[\frac{1}{\ell_n\ell_T}\sum_{j:d_{ij}\le d_n}\sum_{b:d_{ib}\le d_n}\sum_{s:d_{ts}\le d_T}\sum_{v:d_{tv}\le d_T}\gamma^{(d_1d_2)}_{(js,bv)}\right.$$
$$\left.--\left(\frac{1}{\ell_n\ell_T}+o\left(1\right)\right)\sum_{h:d_{ih}\le d_n}\sum_{b:d_{ib}\le d_n}\sum_{w:d_{tw}\le d_T}\sum_{v:d_{tv}\le d_T}\gamma^{(d_1d_2)}_{(hw,bv)}+o\left(1\right)\right]$$
$$=o(1).$$

Therefore, $L_{1nT}=o\left(1\right)$ and $L_{2nT}=(1)$, which completes the proof of (A.2). The next step is to prove (A.3). In view of previous derivations, it suffices to show that

$$\lim_{n,T\to\infty}\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)\right.$$
$$\left.-K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right]^2\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}=0.$$

$$\frac{1}{nT\ell_n\ell_T}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{t,u\in E_T}\sum_{s_{(t)}=1}^{\ell_{t,T}}\sum_{v_{(u)}=1}^{\ell_{u,T}}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)\right.$$
$$\left.-K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right]^2\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}$$
$$=\frac{1}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_1}\sum_{(t,s,u,v)\in\mathcal{J}_1}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)-K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right]^2$$
$$\times\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{\left(j_{(i)}s_{(t)},b_{(a)}v_{(u)}\right)}$$
$$+\frac{1}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_2}\sum_{(t,s,u,v)\in\mathcal{J}_1}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)-K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right]^2$$
$$\times\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{\left(j_{(i)}s_{(t)},b_{(a)}v_{(u)}\right)}$$
$$+\frac{1}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_1}\sum_{(t,s,u,v)\in\mathcal{J}_2}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)-K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right]^2$$
$$\times\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{\left(j_{(i)}s_{(t)},b_{(a)}v_{(u)}\right)}$$
$$+\frac{1}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_2}\sum_{(t,s,u,v)\in\mathcal{J}_2}\left[K\left(\frac{d_{ij_{(i)}}}{d_n}\right)K\left(\frac{d_{ts_{(t)}}}{d_T}\right)-K\left(\frac{d_{ab_{(a)}}}{d_n}\right)K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right]^2$$
$$\times\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{\left(j_{(i)}s_{(t)},b_{(a)}v_{(u)}\right)}$$
$$:=F_{1nT}+F_{2nT}+F_{3nT}+F_{4nT},$$

where

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

For F_{1nT}, we have

$$F_{1nT} \leq \left| \frac{2}{nT\ell_n\ell_T} \sum_{(i,j,a,b)\in\mathcal{I}_1} \sum_{(t,s,u,v)\in\mathcal{J}_1} \left[K^2\left(\frac{d_{ij_{(i)}}}{d_n}\right)\left(K\left(\frac{d_{ts_{(t)}}}{d_T}\right) - K\left(\frac{d_{uv_{(u)}}}{d_T}\right)\right)^2 \right.\right.$$
$$\left.\left. +K^2\left(\frac{d_{uv_{(u)}}}{d_T}\right)\left(K\left(\frac{d_{ij_{(i)}}}{d_n}\right) - K\left(\frac{d_{ab_{(a)}}}{d_n}\right)\right)^2\right] \gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})} \right|$$

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

For F_{2nT}, we note that if $\left|d_{ij_{(i)}} - d_{ab_{(a)}}\right| > 2c_n$, then either $d_{ia} > c_n$ or $d_{j_{(i)}b_{(a)}} > c_n$. Without the loss of generality, we assume that $d_{ia} > c_n$ for $\left(i, j_{(i)}, a, b_{(a)}\right) \in \mathcal{I}_2$. In this case

$$F_{2nT} \leq \frac{8c_L^2}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_2}\sum_{(t,s,u,v)\in\mathcal{J}_1}\left(\frac{c_T}{d_T}\right)^2\left|\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right|$$
$$+ \frac{8}{nT\ell_n\ell_T}\sum_{(i,j,a,b)\in\mathcal{I}_2}\sum_{(t,s,u,v)\in\mathcal{J}_1}(d_{ia})^q\left|\gamma^{(c_1c_2)}_{(it,au)}\gamma^{(d_1d_2)}_{(j_{(i)}s_{(t)},b_{(a)}v_{(u)})}\right|(d_{ia})^{-q}$$
$$= O\left(\frac{1}{c_n^q}\right) + O\left(\frac{c_T^2}{d_T^2}\right).$$

With the similar procedure, we can show that $F_{3nT} = O\left(c_T^{-q}\right) + O\left(c_n^2/d_n^2\right)$ and $F_{4nT} = O\left(1/c_n^q\right)$.

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

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

and (A.3) is proved.

It is easy to show $D_{2nT}=o(1)$, $D_{3nT}=o(1)$, $D_{4nT}=o(1)$ and $D_{5nT}=o(1)$ given $T_2/T\to 0$ and $n_2/n\to 0$.

By symmetry,

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

Therefore,

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

In terms of matrix form,

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

(b) Asymptotic Bias

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

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

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

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

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

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

It is easy to show that $H_{1nT}=o_p(1)$ and $H_{2nT}=o_p(1)$ under Assumptions F9 and F12. For $H_{3nT,}$ we need to show that for all i and t

$$\frac{1}{\sqrt{\ell_n\ell_T}}\sum_{j=1}^{n}\sum_{s=1}^{T}K\left(\frac{d_{ij}}{d_n}\right)K\left(\frac{d_{ts}}{d_T}\right)\tilde{X}_{js}\left(\bar{u}_j+\bar{u}_s-\bar{u}\right)=o_p\left(1\right). \tag{A.11}$$

First, $\bar{u}_j+\bar{u}_s-\bar{u}=o_p\left(1\right)$ uniformly. Second, by Assumption F12(iv)

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

as $\Delta\rightarrow\infty$. 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

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

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

Proof of Corollary 1

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

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

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

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

since S is pd. Therefore,

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

and the sequence $\{(d_n, d_T)\}$ satisfies $d_n^{2q}\ell_n\ell_T/nT \to \tau^\star$ if and only if

$$d_n = d_n^\star + o\left((nT)^{1/(2q+\eta+1)}\right) \text{ and } d_T = d_T^\star + o\left((nT)^{1/(2q+\eta+1)}\right).$$

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

(a) Asymptotic Variance

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

$$\tilde{J}_{nT}(c,d) - \tilde{J}_{nT}^{GA}(c,d) = o_p(1), \tag{A.12}$$

if $d_n \to 0$ as $n \to \infty$. By Chebyshev's inequality, for any $\Delta > 0$,

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

where

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

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

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

as $d_n \to 0$ because $K(d_{ij}/d_n) = 0$ for all $i \neq j$ provided $d_n < \min_{i,j} d_{ij}$. With the similar procedures, we can show that $\tilde{C}_{3nT} \to 0$ and $\tilde{C}_{4nT} \to 0$. Therefore, (A.12) holds.

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

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

$$\tilde{J}_{nT}(c,d) - \tilde{J}_{nT}^{DK}(c,d) = o_p(1), \tag{A.13}$$

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

By Chebyshev's inequality, we have

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

for any Δ, where

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

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

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

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

as $d_n \to \infty$.

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

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

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

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

which implies $\check{C}_{3nT} \to 0$ as $n, T \to \infty$. With the same procedure, we can show that $\check{C}_{4nT} = o(1)$. Therefore, (A.13) holds.

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

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

Lemma 2. *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

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

Proof of Lemma 2

Proofs are in the supplementary appendix.

Proof of Proposition 2

To simplify the notation, we use $\int_{-\infty}^{\infty}$ to stand for $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}$ and $\int_0^1$ to stand for $\int_0^1\int_0^1\cdots\int_0^1$ when there is no confusion. Let

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

Then, for any given $\Delta > 0$

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

because by Assumption I8

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

and

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

by the Fourier series representation. This implies

$$\hat{J}_{nT} - \check{J}_{nT} = o_p(1). \tag{A.14}$$

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

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

where superscript 'H' denotes the conjugate transpose.

From Lemma 2 and (A.14), we have

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

where the equality in distribution holds because

$$P\Bigg(\Bigg\|\int_0^1\Bigg(\sum_{k,\ell,m=1}^{\infty}\lambda_{k,\ell,m}\Phi_{b,k\ell m}\left(r_1,r_2,\tau\right)\Phi_{b,k\ell m}\left(-v_1,-v_2,-\kappa\right)$$

$$-\mathbb{K}_b\left(r_1-v_1,r_2-v_2,\tau-\kappa\right)\Bigg)dB_p\left(r_1,r_2,\tau\right)dB_p'\left(v_1,v_2,\kappa\right)\Bigg\|\geq\Delta\Bigg)$$

$$\leq\int_0^1\Bigg(\sum_{k,\ell,m=1}^{\infty}\lambda_{k,\ell,m}\Phi_{b,k\ell m}\left(r_1,r_2,\tau\right)\Phi_{b,k\ell m}\left(r_1,r_2,\tau\right)-1\Bigg)I_p dr_1 dr_2 d\tau$$

$$=0.$$

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

(a) $E\left(v_{11}-v_{12}v_{22}^{-1}v_{21}\right)=1-b_1b_2b_3c_1-(g-1)\,b_1b_2b_3c_2+o\left(b_1b_2b_3\right),$

(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

$$P\left\{gF_\infty\left(g,b\right)\leq z\right\}$$

$$=EG_g\left(z\left(v_{11}-v_{12}v_{22}^{-1}v_{21}\right)\right)$$

$$=G_g\left(z\right)+G_g'\left(z\right)zE\left[\left(v_{11}-v_{12}v_{22}^{-1}v_{21}\right)-1\right]$$

$$+\frac{1}{2}G_g''\left(z\right)z^2E\left[\left(v_{11}-v_{12}v_{22}^{-1}v_{21}\right)-1\right]^2$$

$$+\frac{1}{2}E\left[G_g''\left(\tilde{z}\right)-G_g''\left(z\right)\right]z^2\left[\left(v_{11}-v_{12}v_{22}^{-1}v_{21}\right)-1\right]^2$$

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

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

Proof of Theorem 6

It follows from Theorem 5 that

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

By definition,

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

Hence

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

1.12 Bibliography

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Chapter 2

Spatial Heteroskedasticity and Autocorrelation Consistent Estimation of Covariance Matrix

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

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

Kelejian and Prucha (2007, hereafter KP) also develop a spatial HAC estimator. As in many empirical studies, they model spatial dependence in terms of a spatial weighting matrix. The difference is that the weighting matrix is not assumed to be known and is not parametrized. Typical examples of this type of processes include the spatial autoregressive processes and spatial moving average processes. Local nonstationarity and heteroskedasticity are built-in features of these type of processes. This is in sharp contrast with Conley (1996, 1999) and Robinson (2005) in which the process is assumed to be stationary.

In spatial HAC estimation literature, an *economic distance* is commonly employed to characterize the decaying pattern of the spatial dependence. The covariance of random variables at locations i and j is a function of $d_{ij,n}$, the economic distance between them. As the economic distance increases, the covariance decreases in absolute value and vice versa. A variety of distance measures can be considered depending on applications. For example, Pinkse, Slade and Brett (2002) use geographic distance and Conley (1999) uses transportation cost. The existence of such an economic distance enables us to use the kernel method for the standard error estimation. The estimator is a weighted sum of sample covariances with weights depending on the relative distances, that is, $d_{ij,n}/d_n$ for some bandwidth parameter d_n.

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

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

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

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

The remainder of the paper is as follows. Section 2 describes the estimation problem and the underlying spatial process we consider and introduces our spatial

HAC estimator. Section 3 establishes the consistency, the rate of convergence, and the asymptotic truncated MSE of the spatial HAC estimator. Section 4 derives asymptotically optimal sequences of fixed bandwidth parameters and proposes a data-dependent implementation. Section 5 studies the consistency, the rate of convergence, and the asymptotic truncated MSE of the spatial HAC estimator with the estimated optimal bandwidth parameter. Section 6 presents Monte Carlo simulation results. Section 7 concludes.

2.1 Spatial Processes and HAC Estimators

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where

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

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

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

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

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

where $\otimes$ denotes the Kronecker product. The process $V_{i,n}(\theta_0)$ may be nonlinear in parameter θ_0 but we assume that $V_{i,n}(\theta_0)$ follows a linear array process as in

(A.3). This type of processes allows for nonstationarity and unconditional heteroskedasticity. It also includes typical spatial parametric models such as spatial autoregressive processes and spatial moving average processes.

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

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

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

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

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

We make the following assumption on ε_n.

Assumption F5. *For each* $n \geq 1$, $\{\varepsilon_{\ell,n}\}$ *are i.i.d.*$(0, 1)$ *with* $E\varepsilon_{\ell,n}^4 \leq c_E$ *for a constant* $c_E < \infty$.

For simplicity, we assume that $\varepsilon_{i,n}$ is independent of $\varepsilon_{j,n}$ for $i \neq j$. Our results can be generalized but with more tedious calculations. Under Assumption F5, the covariance matrix between $V_{i,n}$ and $V_{j,n}$ is given by

$$\Gamma_{ij,n} \equiv \left(\gamma_{ij,n}^{(cd)}\right) = E[V_{i,n}V_{j,n}'] = R_{in}R_{jn}' \tag{A.4}$$

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

$$J_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma_{ij,n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}R_{in}R_{jn}' \tag{A.5}$$

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

$$J_n(c,d) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left[\sum_{m=1}^{np}\sum_{\ell=1}^{np} r_{im,n}^{(c)} r_{j\ell,n}^{(d)} \varepsilon_{m,n}\varepsilon_{\ell,n}\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{m=1}^{np} r_{im,n}^{(c)} r_{jm,n}^{(d)}. \quad \text{(A.6)}$$

Assumption F6. *For all* $j = 1, 2, ..., np$, *and* $s = 1, 2, ..., p$, $\sum_{k=1}^{n} \left| r_{kj,n}^{(s)} \right| < c_R$ *for some constant* c_R, $0 < c_R < \infty$.

Assumption F7. *There exists* $q_d > 0$ *such that* $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\Gamma_{ij,n}\| d_{ij,n}^{q_d} < \infty$ *for all* n, *where* $\|A\|$ *denotes the Euclidean norm of matrix* A.

Assumptions F6 and F7 impose conditions on the persistence of the spatial process. If $|\sigma^{ij}| \leq C$ for some constant $C > 0$, then Assumption F6 holds if $\sum_{k=1}^{n} \left| \tilde{r}_{kj,n}^{(s)} \right| < c_R/C$. Since $\left| \tilde{r}_{kj,n}^{(s)} \right|$ can be regarded as the (absolute) change of $V_{k,n}^{(s)}$ in response to one unit change in $\tilde{\varepsilon}_{jn}^{(s)}$, the summability condition requires that the aggregate response be finite. The condition holds trivially if the set $\{\tilde{r}_{kj,n}^{(s)}, k = 1, 2, ..., n\}$ has only a finite number of nonzero elements. In this case, the dependence induced by the innovation $\tilde{\varepsilon}_{jn}^{(s)}$ are limited to a finite number of units. Assumption F7 states that $\Gamma_{ij,n}$ decays to zero fast enough such that $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\Gamma_{ij,n}\| d_{ij,n}^{q_d}$ is finite for all n. This excludes the case in which the sample size increases because of more intensive sampling within a given distance. This condition enables us to truncate the sum $\sum_{j=1}^{n} \|\Gamma_{ij,n}\|$ and downweigh the summand without incurring a large error. As in the time series literature, this assumption helps us control the asymptotic bias of the spatial HAC estimator. By (A.6), Assumption 3 holds if $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} \left| \sum_{m=1}^{np} r_{im,n}^{(c)} r_{jm,n}^{(d)} \right| d_{ij,n}^{q_d} < \infty$ for all c, d, and n. This implies that if units i and j are far away from each other, then one of $r_{im,n}^{(c)}$ and $r_{jm,n}^{(c)}$ must be small for any given m. In other words, given an innovation at any location m, the responses at locations i and j cannot be both large if i and j are far away from each other.

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

$$\hat{J}_n = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{V}_{i,n} \hat{V}_{j,n}' K\left(\frac{d_{ij,n}^*}{d_n}\right), \tag{A.7}$$

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

Assumption F8. *(i)* $\{\nu_{ij,n}\}$ *are independent of* $\{\varepsilon_{\ell,n}\}$. *(ii)* $\nu_{ij,n} = o(d_n)$ *as* $d_n \to \infty$.
(iii) $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\Gamma_{ij,n}\| E|\nu_{ij,n}|^{q_d} < \infty$ *for all* n.

We allow a measurement error to increase as the distance of two units becomes farther as long as Assumptions F8 (ii) and (iii) hold. Under this assumption, it is straightforward that $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\Gamma_{ij,n}\| E(d_{ij,n}^*)^{q_d} < \infty$ for all n. Essentially, measurement errors in location can not be so large as to change the summability of $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\Gamma_{ij,n}\| E(d_{ij,n}^*)^{q}$.

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

In order to obtain the properties of the estimator in Theorem 7 below, it is important to control the boundary effects. That is, the effects of the units on the boundary should become negligible as the sample size increases so that the asymptotic properties of the estimator depend only on the behavior of the interior units. We define the boundary in terms of the number of pseudo-neighbors. If $E\left|\ell_{i,n}-E\ell_n\right|=o(E\ell_n)$, then we say that i is not on the boundary, otherwise it is on the boundary. Let

$$E_n \equiv \left\{i : E\left|\ell_{i,n}-E\ell_n\right|=o(E\ell_n)\right\},\ n_1=\sum_{i=1}^{n}1\left\{i\in E_n\right\} \text{ and } n_2=n-n_1.$$

Then E_n represents the set of nonboundary locations and n_1 and n_2 are the sizes of nonboundary set and boundary set respectively. A unit i is in the nonboundary set E_n as long as the difference between $\ell_{i,n}$ and $E\ell_n$ does not grow too fast as n increases. The size of boundary depends on the choice of the bandwidth. We can mitigate the boundary effects by raising d_n slowly as n increases. If n_2/n is $o\left(1\right)$, the boundary effect is asymptotically negligible. When the units are regularly spaced on a lattice in $\mathbb{R}^2$, this condition is satisfied if $E\ell_n/n=o\left(1\right)$.

We maintain the following assumption on the number of pseudo-neighbors.

Assumption F9. *For $i\in E_n$, $\ell_{i,n}\leq CE\ell_n$ for some constant C.*

Assumption F9 is very weak as C can be a large constant. This assumption rules out the case that the units are concentrated only in some limited area while other area is scarce.

Assumption F10. *For $i\in E_n$,*

$$\lim_{n\to\infty} var\left(\frac{1}{\sqrt{E\ell_n}}\sum_{j:d^*_{ij,n}\leq d_n}V_{j,n}\right)=g,$$

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

In this assumption, $1/\sqrt{E\ell_n}\sum_{j:d^*_{ij,n}\leq d_n}V_{j,n}$ can be regarded as a local version of the (scaled) global average: $1/\sqrt{n}\sum_{j=1}^{n}V_{j,n}$. Assumption F10 states that the asymptotic variance of each local average (around an interior point) is the same as that of the global average. There are many copies of local averages. Assumption F10 amounts to stating that there are multiple (noisy) observations of the quantity that we want to estimate. From this perspective, Assumption F10 is likely to be the weakest possible assumption we have to maintain. It is similar to the homogeneity assumption in Bester, Conley, Hansen and Vogelsang (2009), which assumes that the covariance matrix of each group converges to the same limit.

Assumption F10 is related to but weaker than covariance stationarity. It is implied by covariance stationarity but it can hold even though covariance stationarity is violated. As an example, consider a nonstationary spatial process on a regular lattice in $\mathbb{R}^2$ represented by

$$S(x_1, x_2) = \int_0^{2\pi} \int_0^{2\pi} \exp(i\omega_1 x_1 + i\omega_2 x_2) f_{(x_1,x_2)}(\omega_1, \omega_2) dZ(\omega_1, \omega_2)$$

where $f_{(x_1,x_2)}(\omega_1, \omega_2)$ is the local spectral density function at location (x_1, x_2), and $Z(\omega_1, \omega_2)$ is a stochastic process with independent increments. The dependence of $f_{(x_1,x_2)}(\omega_1, \omega_2)$ on (x_1, x_2) induces nonstationarity. For this type of nonstationary processes, Assumption F10 amounts to assuming that $f_{(x_1,x_2)}(0,0)$ does not depend on (x_1, x_2), which is much weaker than assuming that $f_{(x_1,x_2)}(\omega_1, \omega_2)$ does not depend on (x_1, x_2) for all (ω_1, ω_2).

In Assumptions F7-F10, we consider the case of a single distance measure. Our framework can be easily extended to allow for multiple distance measures. As in KP, suppose we have M distance measures $d^*_{ij,m,n}$ for $m = 1, 2, ..., M$, each of which is possibly error-ridden. If one of the distance measures, say $d_{ij,1,n}$ and its associated measurement error satisfy Assumptions F7-F10, then

$$n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\Gamma_{ij,n}\| E\left(d^*_{ij,1,n}\right)^{q_d} < \infty.$$

Let

$$d^*_{ij,\min} = \min_m d^*_{ij,m,n}$$

then

$$n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\Gamma_{ij,n}\| E\left(d^*_{ij,\min}\right)^{q_d} < \infty.$$

Hence we can use the $d^*_{ij,\min}$ as the "aggregate distance measure" and construct the HAC estimator based on $d^*_{ij,\min}$. More specifically, our HAC estimator is

$$\hat{J}_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{V}_{i,n}\hat{V}'_{j,n}K\left(\frac{d^*_{ij,\min}}{d_n}\right). \tag{A.8}$$

All our results remain valid as $d^*_{ij,\min}$ is a single distance measure and it is not hard to show that Assumptions F7-F10 hold with $d^*_{ij,\min}$.

2.2 Asymptotic Properties of Spatial HAC Estimators

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

Assumption F11. *(i) The kernel* $K : \mathbb{R} \to [-1, 1]$ *satisfies* $K(0) = 1, K(x) = K(-x)$, $K(x) = 0$ *for* $|x| \geq 1$. *(ii) For all* $x_1, x_2 \in R$ *there is a constant,* $c_L < 0$, *such that*

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

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

Examples of kernels which satisfy Assumptions F11 (i) and (ii) are the Bartlett, Tukey-Hanning and Parzen kernels. The quadratic spectral (QS) kernel does not satisfy Assumption F11(i) because it does not truncate. We may generalize our results to include the QS kernel but this requires a considerable amount of work. Assumption F11 (iii) is more of an assumption on the distribution of the units. In the case of a 2-dimensional lattice structure and $d^*_{ij,n}$ is the Euclidean distance, we have

$$\bar{K} = \frac{1}{\pi}\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} K^2(\sqrt{x^2+y^2})dydx = \int_0^1 rK^2(r)dr.$$

In finite samples, we may use

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

for $\bar{K}$.

As positive semi-definiteness is a desirable property of $\hat{J}_n$, it is important to examine the class of kernel functions that yields a psd covariance estimator. If we map a set of observations into a lattice and assign zero at a grid where an observation is missing, then $\hat{J}_n$ is numerically equivalent to the weighted average of periodogram. Therefore, even though the spatial process of $V_{i,n}$ is not covariance stationary, a kernel with a positive Fourier transformation guarantees the positive semi-definiteness of $\hat{J}_n$ as the case of spectral density estimation.

The asymptotic variance of $\hat{J}_n$ depends on g, the limit value of J_n and the asymptotic bias of $\hat{J}_n$ is determined by the smoothness of the kernel at zero and the rate of decaying of the spatial dependence as a function of the distance. Define

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

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

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

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

Assumption F12. *(i)* $\sqrt{n} \left(\hat{\theta} - \theta_0 \right) = O_p(1)$. *(ii)* $\sup_i E \sup_{\theta \in \Theta_n} \| V_{i,n}(\theta) \|^2 < \infty$ *where* Θ_n *is a small neighborhood around* θ_0. *(iii)* $\sup_i E \sup_{\theta \in \Theta_n} \| \frac{\partial}{\partial \theta'} V_{i,n}(\theta) \|^2 < \infty$. *(iv) For* $r = 1, ..., p$, $\sup_i E \sup_{\theta \in \Theta_n} \| \frac{\partial^2}{\partial \theta \partial \theta'} V_{i,n}^{(r)}(\theta) \|^2 < \infty$. *(v) For* $r, s = 1, \ldots, p$ $\sup_{i,a,b} \sum_{j=1}^{\infty} \left\| E V_{j,n}^{(r)} V_{b,n}^{(r)} \frac{\partial}{\partial \theta'} V_{i,n}^{(s)} \frac{\partial}{\partial \theta} V_{a,n}^{(s)} \right\| < \infty$.

Assumption F12(i) usually holds by the asymptotic normality of parameter estimators. Assumption F12 (ii) is implied by Assumptions F5 and F6. Assumptions F12 (iii), (iv) and (v) are trivial in a linear regression case.

We define the MSE criterion as

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

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

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

Under the assumptions above, the effect of using $\hat{\theta}$ instead of θ_0 on the asymptotic property is $o_p(1)$ as Theorem 7(c) states below. Therefore, we use $\tilde{J}_n$ to analyze the asymptotic properties of $\hat{J}_n$. Notwithstanding, if $\hat{\theta}$ has an infinite second moment, the underlying estimation error can dominate the MSE criterion. To circumvent the undue influence of $\hat{\theta}$ on the criterion of performance, we follow Andrews (1991) and replace the MSE criterion with a truncated MSE criterion. We define

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

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

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

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

Assumption F13. *(i)* $E\varepsilon_{l,n}^8 < \infty$. *(ii)* $S_n \xrightarrow{p} S$ *for a positive definite matrix* S.

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

Theorem 7. *Suppose that Assumptions F5-F11 hold,* $n_2/n \to 0$*,* $E\ell_n$ *and* $d_n \to \infty$ *and* $E\ell_n/n \to 0$*.*

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

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

(c) *If Assumption F12 holds and* $\frac{d_n^{2q} E\ell_n}{n} \to \tau \in (0, \infty)$*, then* $\sqrt{\frac{n}{E\ell_n}}\left(\hat{J}_n - J_n\right) = O_p(1)$ *and* $\sqrt{\frac{n}{E\ell_n}}\left(\hat{J}_n - \tilde{J}_n\right) = o_p(1)$*.*

(d) *Under the conditions of part (c) and Assumption A.17,*

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

Proofs are given in the appendix. For each element, $\lim_{n\to\infty} \frac{n}{E\ell_n} \text{cov}\left(\tilde{J}_{rs,n}, \tilde{J}_{cd,n}\right) = \bar{K}\left(g_{rc}g_{sd} + g_{rd}g_{sc}\right)$ and $\lim_{n\to\infty} d_n^q (E\tilde{J}_{rs,n} - J_{rs,n}) = -K_q g_{rs}^{(q)}$. Theorem 7(a) and (b) show that the asymptotic variance and bias of $\tilde{J}_n$ depend on the choice of the bandwidth. When we increase the bandwidth, the bias decreases and the variance increases because $E\ell_n$ increases with d_n.

The second part of Theorem 7(c) shows that, compared with the variance term in part (a), the effect of using $V_{i,n}(\hat{\theta})$ instead of $V_{i,n}(\theta_0)$ in the construction of the spatial HAC estimator is of a smaller order. Therefore, the rate of convergence is obtained by balancing the variance and the squared bias. Accordingly, $E\ell_n = o(n)$ is the condition for the consistency of $\hat{J}_n$ and its rate of convergence is $\sqrt{E\ell_n/n}$ $(= O(d_n^{-q}))$. If we assume that $E\ell_n = O(d_n^{\eta})$ for some $\eta > 0$, then the rate of convergence can be rewritten as $n^{q/(\eta+2q)}$. The results here are different from those provided by KP. In their paper, the condition for consistency is $E\ell_n = o(n^{\tau})$

where $\tau \le \frac{1}{2}$ and the rate of convergence is $n^{q/(\eta+4q)}$. They obtain this slower rate of convergence by balancing the terms from the estimation error in $\hat{\theta}$ and the asymptotic bias. Their rate is not the best obtainable because their bound for the estimation error term is too loose.

It is also interesting that the asymptotic properties of the spatial HAC estimator are very similar to those of the time series HAC estimator even though their DGPs and dependence structures are different from each other. Instead of using d_n as the bandwidth parameter, we can also use $E\ell_n$ as the bandwidth parameter. In the time series case, $d_n = E\ell_n$. Substituting this relationship into Theorem 7, we obtain the same results as given in Parzen (1957), Hannan (1970) and Andrews (1991).

2.3 Optimal Bandwidth Parameter and Data Dependent Bandwidth Selection

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

We obtain the optimal bandwidth parameter directly as a corollary to Theorem 7 (d). Let $d_n^\star$ be the optimal bandwidth parameter. Then

$$d_n^\star = \arg\min_{d_n} \frac{1}{d_n^{2q}} K_q^2 \left(\text{vec } g^{(q)}\right)' S_n \left(\text{vec } g^{(q)}\right) + \frac{E\ell_n}{n} \bar{K} \text{tr}\left(S_n(I + K_{pp})(g \otimes g)\right) \quad \text{(A.9)}$$

If the relation between $E\ell_n$ and d_n is specified, (A.9) can be restated in an explicit form. For example, we may assume that $E\ell_n = \alpha_n d_n^\eta$ and $\alpha_n = O(1)$ for some $\eta > 0$. Then (A.9) is reduced to:

$$\begin{aligned} d_n^\star &= \arg\min_{d_n} \frac{1}{d_n^{2q}} K_q^2 \left(\text{vec } g^{(q)}\right)' S_n \left(\text{vec } g^{(q)}\right) + \frac{\alpha_n d_n^\eta}{n} \bar{K} \text{tr}\left(S_n(I + K_{pp})(g \otimes g)\right) \\ &= \left(\frac{q K_q^2 \kappa(q) n}{\alpha_n \eta \bar{K}}\right)^{\frac{1}{2q+\eta}} \end{aligned} \quad \text{(A.10)}$$

where

$$\kappa(q) = \frac{2\left(\mathrm{vec}\ g^{(q)}\right)' S_n \left(\mathrm{vec}\ g^{(q)}\right)}{\mathrm{tr}\left(S_n(I+K_{pp})(g\otimes g)\right)}.$$

Corollary 2. *Suppose Assumptions F5-A.17 hold. Assume that* $E\ell_n = \alpha_n d_n^{\eta}$ *for some* $\eta > 0$, $\alpha_n = \alpha + o(1)$. *Then, for any sequence of bandwidth parameters* $\{d_n\}$ *such that* $\frac{d_n^{2q} E\ell_n}{n} \to \tau \in (0,\infty)$ *as* $n \to \infty$, $\{d_n^{\star}\}$ *is preferred in the sense that*

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

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

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

$d_n^{\star}$ is a function of g and $g^{(q)}$ which are unknown in finite samples. Therefore, the optimal bandwidth $d_n^{\star}$ is not feasible in practice. For this reason, a data dependent estimation procedure is needed for implementation. Among several data dependent bandwidth selection methods, plug-in methods are appropriate in this case because we consider the estimation of J_n at given data. In the plug-in methods, unknown parameters are estimated using a parametric or nonparametric method (e.g. Andrews (1991). Newey and West (1987, 1994)). The former yields a less variable bandwidth parameter but may introduce an asymptotic bias due to the mis-specification of the parametric model. In contrast, the latter does not require the knowledge of the DGP, but it converges more slowly than the former, which causes bandwidth selection to be less reliable. Since the optimal bandwidth involves $g^{(q)}$, a quantity that is very hard to estimate, we focus on the parametric plug-in method in this paper. In fact, the rate of convergence for a nonparametric estimator of $g^{(q)}$ is generally slower than that for g itself. Figure 1 presents the

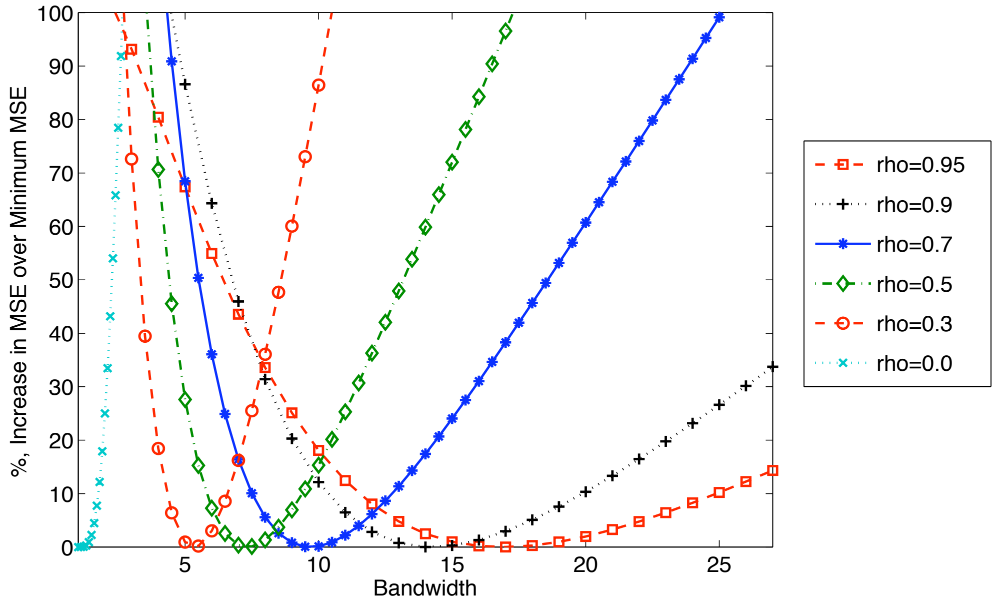

Figure 2.1: Spatial AR(1) Process : $V_n = \rho W_n V_n + \varepsilon_n$, $n = 400$

percentage increase in MSE relative to the minimum MSE as a function of the bandwidth. The graph is based on the spatial AR(1) process $V_n = \rho W_n V_n + \varepsilon_n$ on a square grid of integers, where W_n is a contiguity matrix whose threshold is $\sqrt{2}$ and $\varepsilon_{i,n} \overset{i.i.d}{\sim} N(0,1)$. The sample size is $n = 400$. As a standard practice, W_n is row-standardized and its diagonal elements are zero. The curve is U-shaped for each ρ and therefore our goal is to choose the bandwidth which is reasonably close to $d_n^\star$. As argued by Andrews (1991), good performance of a HAC estimator only requires the automatic bandwidth parameter to be near the optimal bandwidth value and not precisely equal to it. The simplest and most popular approximating parametric model is the spatial AR(1) model for $V_n^{(c)}$, $c = 1, \ldots, p$. Depending on the correlation structure, spatial MA(q) or spatial ARMA(p,q) models can also be used. As an example, consider the case that $V_n^{(c)}$ follows a spatial AR(1) process of the form:

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

where $\tilde{\varepsilon}_{i,n}^{(c)} \overset{i.i.d}{\sim} (0, \sigma_\varepsilon^2)$ and $W_n^{(c)}$ is a spatial weight matrix. $W_n^{(c)}$ is determined a priori and by convention it is row-standardized and its diagonal elements are zero. See Anselin (1988). We can estimate ρ_c by quasi-maximum likelihood (QML) or spatial two stage least squares (2SLS) estimators (e.g. Kelejian and Prucha (1998)). In fact, a simple OLS estimator can be used. If the spatial AR(1) model is the true data generating process, then the OLS estimator is inconsistent while the QML and 2SLS estimators are consistent. Since the spatial AR(1) model is likely to be mis-specified, the QML and 2SLS estimators are not necessarily preferred.

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

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

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

$$\hat{g}_{cd} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{a}_{ij}^{(cd)}, \hat{g}_{cd}^{(q)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{a}_{ij}^{(cd)}\left(d_{ij,n}^*\right)^q. \tag{A.11}$$

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

$$\hat{d}_n = \arg\min_{d_n}\frac{1}{d_n^{2q}}K_q^2\left(\text{vec } \hat{g}^{(q)}\right)' S_n\left(\text{vec } \hat{g}^{(q)}\right) + \bar{K}\frac{\ell_n}{n}\text{tr}\left(S_n(I + K_{pp})(\hat{g}\otimes\hat{g})\right). \tag{A.12}$$

For spatial MA(1) and spatial ARMA(1,1) models, (A.11) is restated as

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

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

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

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

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

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

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

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

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

2.4 Properties of Data Dependent Bandwidth Parameter Estimators

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

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

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

$$\ddot{d}_n = \arg\min_{d_n}\frac{1}{d_n^{2q}}K_q^2\left(\text{vec }\ddot{g}^{(q)}\right)' S_n\left(\text{vec }\ddot{g}^{(q)}\right) + \bar{K}\frac{E\ell_n}{n}\text{tr}\left(S_n(I+K_{pp})\left(\ddot{g}\otimes\ddot{g}\right)\right). \tag{A.13}$$

We study the properties of $\hat{J}_n(\hat{d}_n)$ by investigating $\hat{J}_n(\ddot{d}_n)$ because the asymptotic properties of $\hat{J}_n(\hat{d}_n)$ are equivalent to those of $\hat{J}_n(\ddot{d}_n)$ as stated in Theorem 8 below. For Theorem 8, we introduce the following assumption.

Assumption F14. $\sqrt{n}\left(\frac{\ddot{d}_n}{\hat{d}_n}-1\right) = O_p(1)$.

Since $\ddot{d}_n$ is defined by replacing ℓ_n, $\hat{g}_{cd}$ and $\hat{g}_{cd}^{(q)}$ with $E\ell_n$, the probability limits of $\hat{g}_{cd}$ and $\hat{g}_{cd}^{(q)}$ in the definition of $\hat{d}_n$, the assumption holds if ℓ_n, $\hat{g}_{cd}$ and $\hat{g}_{cd}^{(q)}$ converge to $E\ell_n$, $\ddot{g}_{cd}$ and $\ddot{g}_{cd}^{(q)}$ respectively at the parametric rate. This is a rather weak assumption.

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

Theorem 8. *Suppose Assumptions 1-F14 hold.*

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

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

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

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

Corollary 3. *Suppose Assumptions 1-F14 hold. Assume that* $E\ell_n = \alpha_n d_n^\eta$ *for some* $\eta > 0$ *and* $\alpha_n = \alpha + o(1)$. *Then for any sequence of data dependent bandwidth estimators* $\{\dot{d}_n\}$ *such that for some fixed sequence,* $\{d_n\}$, *which satisfies* $\lim_{n\to\infty} \frac{d_n^{2q} E\ell_n}{n} \to \tau \in (0,\infty)$ *we have*

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

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

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

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

2.5 Monte Carlo Simulation

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

The data generating process we consider here is

$$y_n = X_n\theta_0 + u_n \tag{A.14}$$

$$u_n = \rho_0 W_{0n} u_n + \varepsilon_n, \ |\rho_0| < 1, \tag{A.15}$$

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

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

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

2.5.1 Location Model

For the location model, model (A.14) reduces to

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

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

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

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

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

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

Depending on the choice of W_n, we obtain a different $\hat{\rho}$ and hence a different bandwidth parameter $\hat{d}_n(W_n)$ from equation (A.12). To find $\hat{d}_n$, we search the minimizer numerically instead of using the plug-in version of (A.10). In our simulation experiment, we take W_n to be the contiguity matrix in which units i and j are neighbors if $d_{ij,n} \leq \mathcal{D}$, a threshold parameter. We consider three values for the threshold: $\mathcal{D} = 1, \sqrt{2}, 2$, leading to three bandwidth choices $\hat{d}_n^{(\ell)}, \hat{d}_n$ and $\hat{d}_n^{(h)}$. Note that when $\mathcal{D} = \sqrt{2}$, the spatial weighting matrix is equal to the true spatial weighting matrix W_{0n}. We also consider the case with measure errors in distance. When $d_{ij,n} > 1$, we take $P(\nu_{ij,n} = -1) = P(\nu_{ij,n} = 0) = P(\nu_{ij,n} = 1) = 1/3$. We use the contiguity matrix as the weighting matrix and take the threshold parameter to be $\sqrt{2}$. This gives us the data driven bandwidth estimator $\hat{d}^{(e)}$.

KP suggest taking $d_n = c[n^{\frac{1}{4}}]$, where c is a constant and $[z]$ denotes the largest integer that is less than or equal to z and they use $[n^{\frac{1}{4}}]$ in their simulation. We compare the performances of $\hat{d}_n^{(\ell)}, \hat{d}_n, \hat{d}_n^{(h)}$ and $\hat{d}^{(e)}$ with that of $d_n^F = [n^{\frac{1}{4}}]$.[1] We

[1] F denotes "fixed."

also include the heteroskedasticity consistent estimator of White (1980, 1984) in our comparison. The estimator is defined to be

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

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

Table 1 presents the ratio of the MSE of the spatial HAC estimator with different bandwidth choices to the spatial HAC estimator with the infeasible finite sample optimal bandwidth $\tilde{d}_n$. It also reports the average bandwidth choice in each scenario. When ρ is high, the ratio is usually less than 1.20 even with incorrect W_n and measurement errors. When ρ is negative, the ratio with $\hat{d}_n^{(\ell)}$ is larger than 2.0. Table 1 also illustrates how mis-specification in the spatial weighting matrix affects our choice of bandwidth. If W_n includes fewer units as neighbors, the bandwidth estimator tends to be smaller than the one with correct W_n. In contrast, if W_n includes more units as neighbors, the bandwidth estimator tends to be larger. This coincides with our intuition. If we think we have a larger neighborhood, we need to choose a larger bandwidth to reflect the dependence structure.

Table 1 also presents the performance of the spatial HAC estimator with measurement errors. The effects of measurement errors are related to the mis-specification of W_n. For a given bandwidth parameter, positive measurement errors lead to a smaller number of neighbors and vice versa. Whereas, in contrast to the mis-specification of W_n, measurement errors are different across different individuals. Table 1 shows that the estimator contaminated by measurement errors performs very poorly compared to other estimators when $\rho = -0.5$, while it performs reasonably well when ρ is positive.

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

Table 2 provides the bias, variance and MSE of the spatial HAC estimators with different bandwidth selection and those of INID. We use SHAC_0, SHAC_l, SHAC$_h$, SHAC$_e$ and SHAC$_F$ to denote the spatial HAC estimators with $\hat{d}_n$, $\hat{d}_n^{(l)}$, $\hat{d}_n^{(h)}$, $\hat{d}_n^{(e)}$ and d_n^F respectively. We can see that SHAC$_0$ is reasonably accurate in general but that it suffers from severe underestimation when ρ is extremely high. Spatial HAC estimators do not capture high dependence well even if we choose a large bandwidth since spatial HAC estimators are constructed with the estimated residuals not the true disturbances. Our asymptotic theory does not capture the effect of demeaning on the SHAC estimator. This is analogous to the time series case, see for example, Sun, Phillips and Jin (2008) and Sun and Phillips (2008).

When there is no spatial dependence ($\rho = 0$), SHAC$_0$ is quite reliable in that the RMSE is only 15% of the true value even though INID is slightly more accurate. When there exists some spatial dependence, SHAC$_0$ is much more accurate than INID. Furthermore, INID is rarely improved with an increasing sample size, which is in sharp contrast to SHAC$_0$. For example, when $\rho = 0.3$ and $n = 400$, the MSE of SHAC$_0$ is less than a third of that of INID. When $n = 1024$, the difference increases with the former less than a fifth of the latter. Therefore, when there is no spatial dependence, the loss of efficiency from using a spatial HAC estimator with data dependent bandwidth is small. Whereas, there is a remarkable reduction in RMSE by using a spatial HAC estimator when there exists spatial dependence.

Table 2 also shows how mis-specification in W_n and measurement errors affect the performance of the spatial HAC estimator using the bandwidth choice we suggest. Comparing SHAC$_e$ with SHAC$_0$, we find that measurement errors lead to higher MSE. However, the difference in MSE is not very large, reflecting the robustness of the SHAC to the presence of measurement errors. Similarly, mis-specification in W_n is not critical in our simulation design. Among the three bandwidth choices $\hat{d}_n^{(l)}$, $\hat{d}_n^{(h)}$, $\hat{d}_n^{(e)}$, none of them performs consistently better than others and the difference gets smaller when $n = 1024$. Compared to SHAC$_F$, all of them tend to yield smaller MSEs especially when $n = 400$ and ρ is high.

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

Table 3 shows that, when $\rho = 0.9$ or 0.95, the error in coverage probability (ECP) is substantial. For example, when $\rho = 0.95$, the ECP for the 95% CI with SHAC_0 is 14.6% even when $n = 1024$. As seen in Table 2, the downward bias of spatial HAC estimators becomes very large when spatial dependence is very high. For this reason, the CIs tend to be very tight. The ECP comes from two sources. First, the spatial HAC estimator is biased downward. Second, the CIs are based on the asymptotic normal approximation. In order to alleviate this problem, we investigate the performance of some bootstrap procedures in Table 5.

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

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

The thresholds for W_{1n}, W_{2n}, W_{3n} and W_{4n} are $d_{ij,n} \leq \sqrt{2}$, $\sqrt{2} < d_{ij,n} \leq 2$, $2 < d_{ij,n} < \sqrt{5}$ and $\sqrt{5} < d_{ij,n} \leq 2\sqrt{2}$ respectively. Regardless of the number of lags the true process has, we use spatial AR(1) as the approximating parametric model. Table 4 illustrates that as the number of lags increases, the accuracy of the spatial HAC estimator using the spatial AR(1) model becomes lower. However, comparison with d_n^F clearly shows that the plug-in method using spatial AR(1) model performs reasonably well. For example, when $\rho = 0.4$ and the DGP is spatial AR(4) the empirical coverage probability of the 99% CI with SHAC_0 is 93.5% and that with SHAC_F is 86.5%.

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

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

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

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

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

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

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

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

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

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

See Davidson and Flachaire (2001) for more details.

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

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

Table 5 shows that the bootstrap methods implemented here improve the accuracy of the CIs compared to the standard normal approximation, especially when the dependence is extremely high. As we have seen in previous tables, the standard normal approximation yields a large size distortion when spatial dependence is very high. However, we don't find this problem from the bootstrap procedures. Between the i.i.d. *naive* bootstrap and the wild bootstrap, there is no significant difference. For example, when $\rho = 0.95$, the empirical coverage probabilities of the 95% CI by the i.i.d. *naive* bootstrap and the wild bootstrap are 85.8% and 83.2% respectively, while that of CLT is 69.2%.

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

2.5.2 Univariate Model

In the second part, the regression model we consider is

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

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

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

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

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

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

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

Table 7 reports the bias and MSE of the SHAC estimator $\hat{J}_{22,n}$ for the above two weighting matrices. The coverage probability of the associated 95% CI is also reported. When $\psi = 0.3$, $\hat{S}_n$ always yields more accurate $\hat{J}_{22,n}$ if $\rho > 0$. In the case that ρ is very high, the reduction in MSE and improvement in coverage accuracy by using $\hat{S}_n$ over $\check{S}_n$ are remarkable. For example, when $\rho = 0.95$ and $n = 400$, the MSE of $\hat{J}_{22,n}$ with weighting matrix $\hat{S}_n$ is 27.43 while that with $\check{S}_n$ is 42.93. The empirical coverage probability of the CI with $\hat{S}_n$ is 91.4% and that with $\check{S}_n$ is 82.6%. When $n = 1024$, the difference is still very large but become less dramatic. When $\psi = 0.9$, $\hat{S}_n$ performs better than $\check{S}_n$ in most cases although the margin of improvement is small.

2.6 Conclusion

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

In this paper, we focus on the asymptotic truncated MSE criterion, which may not be most suitable for hypothesis testing or CI construction. It is interesting to extend the methods by Sun, Phillips and Jin (2008) and Sun and Phillips (2008) on time series HAC estimation to the spatial setting.

2.7 Acknowldegements

Chapter 2, which is coauthored with Yixiao Sun, is published in Journal of Econometrics, 2011. Vol 160 (2011), pp. 349-371.

Table 2.1: Ratio of the MSE of Spatial HAC Estimators with Different Bandwidths to the MSE of the Spatial HAC Estimator with Optimal Bandwidth, $\tilde{d}_n$

	ρ							
	-0.5	-0.3	0	0.3	0.5	0.7	0.9	0.95
$\hat{d}_n$	1.08	1.16	5.02	1.18	1.14	1.12	1.07	1.06
	(6.2)	(5.4)	(3.5)	(6.5)	(8.8)	(11.7)	(17.9)	(22.2)
$\hat{d}_n^{(l)}$	2.22	2.08	2.93	1.17	1.09	1.05	1.02	1.01
	(4.3)	(3.8)	(2.8)	(4.6)	(6.2)	(8.2)	(12.5)	(15.5)
$\hat{d}_n^{(h)}$	1.14	1.27	7.46	1.37	1.34	1.30	1.21	1.13
	(7.1)	(6.2)	(4.4)	(7.6)	(10.5)	(14.4)	(23.1)	(26.5)
$\hat{d}_n^{(e)}$	4.92	1.49	7.40	1.47	1.42	1.34	1.21	1.13
	(4.1)	(4.5)	(4.0)	(7.8)	(10.8)	(14.7)	(23.1)	(26.4)
d_n^F	2.41	1.40	4.43	1.20	1.68	2.07	1.99	1.73
$d_n^\star$	1.00	1.00	1.00	1.10	1.10	1.09	1.07	1.07
	(6.2)	(5.5)	(1.0)	(6.7)	(8.9)	(11.8)	(18.4)	(23.2)
$\tilde{d}_n$	(6.3)	(5.3)	(1.3)	(5.4)	(7.1)	(9.2)	(13.7)	(16.6)

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

Table 2.2: Bias, Variance, and MSE in Location Model with Spatial AR(1) Error

	Bias		Variance		MSE (RMSE/True Value)			
	n=400	n=1024	n=400	n=1024	n=400		n=1200	
				$\rho = 0$				
$SHAC_0$	-0.014	-0.007	0.023	0.011	0.023	(0.15)	0.011	(0.10)
$SHAC_l$	-0.010	-0.004	0.013	0.006	0.013	(0.12)	0.006	(0.08)
$SHAC_h$	-0.025	-0.012	0.033	0.016	0.034	(0.18)	0.016	(0.13)
$SHAC_e$	-0.021	-0.009	0.034	0.016	0.034	(0.19)	0.016	(0.13)
$SHAC_F$	-0.018	-0.020	0.020	0.019	0.020	(0.14)	0.020	(0.14)
INID	-0.001	0.001	0.005	0.002	0.005	(0.07)	0.002	(0.04)
				$\rho = 0.3$				
$SHAC_0$	-0.337	-0.255	0.211	0.123	0.324	(0.28)	0.188	(0.21)
$SHAC_l$	-0.444	-0.354	0.124	0.071	0.321	(0.28)	0.196	(0.22)
$SHAC_h$	-0.319	-0.235	0.275	0.164	0.377	(0.30)	0.219	(0.23)
$SHAC_e$	-0.340	-0.251	0.287	0.172	0.402	(0.31)	0.235	(0.24)
$SHAC_F$	-0.519	-0.322	0.060	0.068	0.329	(0.28)	0.172	(0.20)
INID	-1.001	-1.000	0.005	0.002	1.006	(0.49)	1.000	(0.49)
				$\rho = 0.5$				
$SHAC_0$	-0.949	-0.703	1.192	0.714	2.093	(0.36)	1.209	(0.27)
$SHAC_l$	-1.156	-0.927	0.664	0.389	2.000	(0.35)	1.248	(0.28)
$SHAC_h$	-0.947	-0.670	1.553	0.968	2.450	(0.39)	1.417	(0.30)
$SHAC_e$	-0.994	-0.698	1.598	1.007	2.586	(0.40)	1.495	(0.30)
$SHAC_F$	-1.703	-1.128	0.174	0.217	3.074	(0.44)	1.490	(0.30)
INID	-2.859	-2.855	0.008	0.004	8.182	(0.71)	8.153	(0.71)
				$\rho = 0.7$				
$SHAC_0$	-3.701	-2.690	12.325	7.937	26.02	(0.45)	15.18	(0.35)
$SHAC_l$	-4.166	-3.337	7.166	4.313	24.52	(0.44)	15.45	(0.35)
$SHAC_h$	-3.860	-2.673	15.234	10.715	30.13	(0.49)	17.86	(0.38)
$SHAC_e$	-3.985	-2.743	15.325	10.983	31.21	(0.50)	18.51	(0.38)
$SHAC_F$	-6.876	-5.054	0.871	1.219	48.15	(0.62)	26.76	(0.46)
INID	-9.731	-9.705	0.024	0.010	94.71	(0.87)	94.20	(0.87)
				$\rho = 0.9$				
$SHAC_0$	-54.32	-39.60	1010.1	899.0	3960.9	(0.62)	2467.1	(0.49)
$SHAC_l$	-54.86	-43.78	781.5	541.4	3791.1	(0.61)	2457.7	(0.49)
$SHAC_h$	-59.31	-42.76	956.6	1077.9	4474.0	(0.66)	2906.4	(0.53)
$SHAC_e$	-59.80	-42.90	938.4	1071.9	4514.9	(0.66)	2912.1	(0.53)
$SHAC_F$	-85.73	-73.70	26.7	42.7	7376.6	(0.84)	5474.0	(0.73)
INID	-98.49	-97.92	0.5	0.2	9700.0	(0.97)	9587.7	(0.97)

Table 2.3: ECPs in Location Model with Spatial AR(1) Error

	99%		95%		90%	
	n=400	n=1024	n=400	n=1024	n=400	n=1024
			$\rho = 0$			
$SHAC_0$	98.8	99.0	95.0	95.0	90.4	90.3
$SHAC_l$	98.7	99.1	95.1	95.1	90.9	90.7
$SHAC_h$	98.8	99.0	94.8	95.2	90.1	90.1
$SHAC_e$	98.7	98.9	95.0	95.0	89.9	90.0
$SHAC_F$	98.9	98.9	95.4	94.9	90.5	90.0
INID	98.9	99.0	95.8	95.7	91.3	91.3
			$\rho = 0.3$			
$SHAC_0$	97.8	98.4	92.2	93.4	87.1	87.4
$SHAC_l$	97.5	98.2	92.0	92.9	86.2	86.5
$SHAC_h$	97.7	98.4	92.1	93.5	86.8	87.5
$SHAC_e$	97.4	98.1	91.9	93.0	87.0	87.0
$SHAC_F$	97.2	98.2	91.4	93.2	85.7	86.8
INID	94.0	94.0	84.6	84.0	76.4	77.6
			$\rho = 0.5$			
$SHAC_0$	96.7	97.6	90.2	91.9	83.3	86.1
$SHAC_l$	96.7	97.5	90.4	91.6	82.9	85.2
$SHAC_h$	96.5	97.7	90.3	91.6	82.8	85.8
$SHAC_e$	96.3	97.7	89.6	91.5	82.2	85.7
$SHAC_F$	95.0	96.9	87.4	90.7	79.1	83.6
INID	84.3	83.8	70.8	72.3	63.9	64.6
			$\rho = 0.7$			
$SHAC_0$	95.5	96.9	86.6	89.7	80.0	83.3
$SHAC_l$	95.0	96.5	87.0	89.1	79.6	82.9
$SHAC_h$	94.2	96.7	85.3	89.1	78.7	83.6
$SHAC_e$	93.4	96.6	84.7	88.9	78.1	83.2
$SHAC_F$	89.6	94.2	77.8	85.3	69.2	78.1
INID	66.5	66.7	53.3	54.2	46.2	47.7
			$\rho = 0.9$			
$SHAC_0$	86.7	93.5	77.3	84.9	69.1	78.5
$SHAC_l$	87.6	93.4	78.0	84.4	70.4	78.1
$SHAC_h$	84.2	91.4	73.9	83.8	66.7	76.9
$SHAC_e$	83.9	91.4	73.7	83.6	66.4	76.5
$SHAC_F$	68.8	82.1	56.8	70.4	48.9	62.5
INID	35.2	37.4	27.2	28.8	23.2	24.8

Table 2.4: Performance under Misspecified Approximating Parametric Model

	SAR(2)		SAR(3)		SAR(4)	
	$SHAC_0$	$SHAC_F$	$SHAC_0$	$SHAC_F$	$SHAC_0$	$SHAC_F$
			$\rho = 0.2$			
Bias	-0.293	-0.400	-0.316	-0.431	-0.322	-0.437
MSE	0.206	0.204	0.226	0.231	0.230	0.237
	(0.26)	(0.26)	(0.27)	(0.27)	(0.27)	(0.27)
99%	97.7	97.5	97.7	97.3	97.7	97.3
95%	92.4	91.6	92.3	91.5	91.9	91.5
90%	87.5	86.3	87.1	85.9	87.1	85.8
			$\rho = 0.3$			
Bias	-0.656	-1.020	-0.796	-1.231	-0.852	-1.307
MSE	0.809	1.123	1.091	1.607	1.207	1.802
	(0.33)	(0.39)	(0.35)	(0.43)	(0.36)	(0.44)
99%	97.0	95.9	96.5	95.1	96.4	94.7
95%	90.7	88.0	90.5	87.5	90.3	87.1
90%	84.2	80.6	83.0	79.3	82.8	78.6
			$\rho = 0.4$			
Bias	-1.707	-2.850	-2.760	-4.495	-3.448	-5.480
MSE	4.634	8.335	10.88	20.52	16.16	30.41
	(0.41)	(0.56)	(0.46)	(0.64)	(0.49)	(0.67)
99%	95.7	91.9	94.3	88.6	93.5	86.5
95%	88.6	81.4	86.2	77.0	84.2	73.6
90%	81.2	73.4	78.9	68.6	78.0	66.9
			$\rho = 0.5$			
Bias	-7.253	-11.95	-39.50	-57.33	-197.7	-249.9
MSE	72.33	144.0	1845.9	3294.3	41887	62499
	(0.53)	(0.74)	(0.66)	(0.88)	(0.77)	(0.95)
99%	91.7	81.1	84.7	63.6	74.3	43.8
95%	82.5	68.3	74.5	50.2	64.3	34.5
90%	76.0	60.7	67.4	42.8	55.7	28.1

(1) Number in parenthesis represents the ratio of the RMSE to the true value.
(2) $d_n^F = 4$.

Table 2.5: ECPs with the Bootstrap and Standard Normal Approximations

	Normal	i.i.d. Bootstrap	Wild Bootstrap
		$\rho = 0.0$	
99%	98.8	99.2	99.0
95%	95.0	95.7	95.2
90%	90.4	90.9	90.7
		$\rho = 0.3$	
99%	97.8	98.5	98.5
95%	92.2	93.4	93.7
90%	87.1	88.9	89.2
		$\rho = 0.5$	
99%	96.7	98.5	98.4
95%	90.2	92.8	93.0
90%	83.3	87.3	87.1
		$\rho = 0.7$	
99%	95.5	97.6	98.0
95%	86.6	92.2	91.9
90%	80.0	84.8	85.4
		$\rho = 0.9$	
99%	86.7	97.0	96.7
95%	77.3	88.3	87.4
90%	69.1	81.6	81.0
		$\rho = 0.95$	
99%	79.5	94.0	93.5
95%	69.2	85.8	83.2
90%	62.1	78.8	75.9

Table 2.6: Performance in the Presence of Irregulairty and Sparsity

	$n = 300, (20 \times 20)$		$n = 400, (25 \times 25)$	
	SHAC_0	SHAC_F	SHAC_0	SHAC_F
		$\rho = 0$		
Bias	-0.016	-0.021	-0.012	-0.027
MSE	0.023	0.021	0.015	0.023
	(0.15)	(0.15)	(0.12)	(0.15)
99%	99.1	98.8	99.1	99.1
95%	95.1	95.0	96.2	96.0
90%	89.7	89.5	92.0	92.0
		$\rho = 0.3$		
Bias	-0.256	-0.364	-0.207	-0.229
MSE	0.198	0.186	0.119	0.108
	(0.25)	(0.24)	(0.21)	(0.20)
99%	97.7	97.5	98.9	98.8
95%	92.6	91.7	94.8	94.5
90%	87.1	86.2	89.6	89.2
		$\rho = 0.5$		
Bias	-0.611	-1.107	-0.446	-0.667
MSE	1.008	1.364	0.531	0.588
	(0.32)	(0.37)	(0.27)	(0.28)
99%	96.9	95.7	98.2	97.9
95%	90.7	88.5	93.3	92.3
90%	86.0	80.8	86.9	85.0
		$\rho = 0.7$		
Bias	-1.989	-4.204	-1.251	-2.564
MSE	9.794	18.293	4.387	7.222
	(0.40)	(0.55)	(0.33)	(0.42)
99%	94.6	89.9	96.9	94.9
95%	87.9	80.4	90.2	84.6
90%	81.5	70.8	82.6	77.1
		$\rho = 0.9$		
Bias	-25.81	-49.63	-14.55	-32.66
MSE	1180.9	2480.2	457.9	1084.5
	(0.55)	(0.80)	(0.45)	(0.70)
99%	88.0	69.6	92.1	78.3
95%	78.8	57.1	83.1	63.8
90%	70.8	50.4	74.2	55.4

Table 2.7: Bias and MSE of $\hat{J}_{22,n}$ and ECPs with Different Weighting Matrices

n	ψ			ρ 0	0.3	0.5	0.7	0.9	0.95
400	0.3	$\hat{S}_n$	Bias	-0.023	-0.054	-0.096	-0.200	-0.813	-1.845
			MSE	0.029	0.041	0.075	0.234	4.076	27.43
			95%	93.8	93.4	92.8	92.7	91.7	91.4
		$\check{S}_n$	Bias	-0.019	-0.075	-0.147	-0.336	-1.706	-4.451
			MSE	0.024	0.062	0.151	0.535	7.850	42.93
			95%	93.9	92.6	91.5	89.8	85.9	82.6
	0.9	$\hat{S}_n$	Bias	-0.049	-0.286	-0.679	-1.968	-14.00	-40.78
			MSE	0.047	0.245	1.051	7.440	333.8	2893.6
			95%	92.7	90.9	89.5	86.8	81.9	79.3
		$\check{S}_n$	Bias	-0.039	-0.295	-0.686	-1.991	-14.62	-44.04
			MSE	0.039	0.245	1.049	7.584	357.3	3173.0
			95%	92.9	90.8	89.6	86.6	80.6	76.4
1024	0.3	$\hat{S}_n$	Bias	-0.009	-0.037	-0.071	-0.152	-0.620	-1.444
			MSE	0.012	0.018	0.035	0.111	1.771	11.90
			95%	95.1	95.1	94.9	94.8	94.4	93.6
		$\check{S}_n$	Bias	-0.007	-0.043	-0.087	-0.204	-1.068	-2.891
			MSE	0.010	0.034	0.086	0.300	4.361	24.89
			95%	95.2	94.6	94.0	93.8	90.9	89.5
	0.9	$\hat{S}_n$	Bias	-0.025	-0.210	-0.481	-1.363	-9.706	-29.49
			MSE	0.019	0.133	0.580	4.195	197.1	1829.4
			95%	95.8	94.4	93.3	91.6	89.1	87.0
		$\check{S}_n$	Bias	-0.020	-0.211	-0.477	-1.346	-9.850	-31.41
			MSE	0.016	0.132	0.580	4.289	215.6	2101.9
			95%	96.0	94.3	93.2	91.8	88.7	84.5

2.8 Appendix

Proof of Theorem 7

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

(a) Asymptotic Variance

Let $\varphi_{lkrs,n} = \sum_{i,j=1}^{n} r^{(r)}_{il,n} r^{(s)}_{jk,n} K\left(\frac{d^*_{ij,n}}{d_n}\right)$ and $N_n = \{\nu_{ij,n} | i,j = 1, \ldots, n\}$ be the set of measurement errors in distance. By Assumption F8(i), we have

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

where

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

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

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

Therefore,

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

using Assumptions F5 and F6.

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

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

In order to consider the boundary effects, we decompose $C_{2,n}$ as follows:

$$\frac{1}{nE\ell_n}E\left[\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}} K\left(\frac{d_{ij_{(i)},n}^*}{d_n}\right) K\left(\frac{d_{ab_{(a)},n}^*}{d_n}\right)\gamma_{ia,n}^{(rc)}\gamma_{j_{(i)}b_{(a)},n}^{(sd)}\right]$$
$$+ \frac{1}{nE\ell_n}E\left[\sum_{i\notin E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a\in E_n}\sum_{b_{(a)}=1}^{\ell_{a,n}} K\left(\frac{d_{ij_{(i)},n}^*}{d_n}\right) K\left(\frac{d_{ab_{(a)},n}^*}{d_n}\right)\gamma_{ia,n}^{(rc)}\gamma_{j_{(i)}b_{(a)},n}^{(sd)}\right]$$
$$+ \frac{1}{nE\ell_n}E\left[\sum_{i\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a\notin E_n}\sum_{b_{(a)}=1}^{\ell_{a,n}} K\left(\frac{d_{ij_{(i)},n}^*}{d_n}\right) K\left(\frac{d_{ab_{(a)},n}^*}{d_n}\right)\gamma_{ia,n}^{(rc)}\gamma_{j_{(i)}b_{(a)},n}^{(sd)}\right]$$
$$+ \frac{1}{nE\ell_n}E\left[\sum_{i,a\notin E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}} K\left(\frac{d_{ij_{(i)},n}^*}{d_n}\right) K\left(\frac{d_{ab_{(a)},n}^*}{d_n}\right)\gamma_{ia,n}^{(rc)}\gamma_{j_{(i)}b_{(a)},n}^{(sd)}\right]$$
$$= D_{1n} + D_{2n} + D_{3n} + D_{4n}.$$

In the following, we show that D_{1n} converges to $\bar{K}g_{rc}g_{sd}$ and the other terms become negligible as n increases.

In order to prove that D_{1n} converges to $\bar{K}g_{rc}g_{sd}$, it suffices to first show that

$$\lim_{n\to\infty}\frac{1}{nE\ell_n}E\left[\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}} K^2\left(\frac{d_{ij_{(i)},n}^*}{d_n}\right)\gamma_{ia,n}^{(rc)}\gamma_{j_{(i)}b_{(a)},n}^{(sd)}\right]$$
$$= \lim_{n\to\infty}\frac{1}{nE\ell_n}E\left[\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}} K^2\left(\frac{d_{ab_{(a)},n}^*}{d_n}\right)\gamma_{ia,n}^{(rc)}\gamma_{j_{(i)}b_{(a)},n}^{(sd)}\right]$$
$$= \bar{K}g_{rc}g_{sd}, \qquad \text{(A.2)}$$

and then show that

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

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

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

For the first term in (A.4), under the Assumption F11 (i),

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

Note that

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

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

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

That is, for any $\varepsilon > 0$, there exists a $N_0 > 0$ such that for $n \geq N_0$

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

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

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

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

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

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

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

As a result

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

using Assumption F11(ii).

For the second term in (A.4), the first step is to show

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

and the second step is to prove

$$
\frac{1}{nE\ell_n}E\left[\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}K^2\left(\frac{d^*_{ia,n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\left(\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\bar{\imath},b_{(a)},n}\right)\right]=o\left(1\right). \tag{A.9}
$$

For (A.8)

$$
\begin{aligned}
&\frac{1}{nE\ell_n}E\left[\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left(K^2\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)-K^2\left(\frac{d^*_{ia,n}}{d_n}\right)\right)\right. \\
&\left.\times\gamma^{(rc)}_{ia,n}\left(\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\bar{\imath},b_{(a)},n}\right)\right] \\
&\leq \frac{1}{nE\ell_n}E\left[\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left|\gamma^{(rc)}_{ia,n}\right|\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\bar{\imath},b_{(a)},n}\right|\right] \\
&= \frac{1}{n}\sum_{(i,a)\in\mathcal{F}_1}\left|\gamma^{(rc)}_{ia,n}\right|E\left[\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\bar{\imath},b_{(a)},n}\right|\right] \\
&+ \frac{1}{n}\sum_{(i,a)\in\mathcal{F}_2}\left|\gamma^{(rc)}_{ia,n}\right|E\left[\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\bar{\imath},b_{(a)},n}\right|\right] \\
&= M_{1n}+M_{2n},
\end{aligned}
$$

where

$$\mathcal{F}_1=\left\{(i,a):d_{ia,n}\leq d_n\ \&\ i,a\in E_n\right\},$$

and

$$\mathcal{F}_2=\left\{(i,a):d_{ia,n}> d_n\ \&\ i,a\in E_n\right\}.$$

For M_{1n}, we obtain

$$
\begin{aligned}
M_{1n} &\leq \left(\frac{E\ell_n}{n}\right)\frac{1}{E\ell_n}\sum_{(i,a)\in\mathcal{F}_1}\left|\gamma^{(rc)}_{ia,n}\right|E\left[\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right|\right. \\
&\left.+\frac{1}{(E\ell_n)^2}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{h_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left|\gamma^{(sd)}_{h_{(i)}b_{(a)},n}\right|\right] \\
&= \left(\frac{E\ell_n}{n}\right)\frac{1}{E\ell_n}\sum_{(i,a)\in\mathcal{F}_1}\left|\gamma^{(rc)}_{ia,n}\right|\left(\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{E\ell_n}\sum_{b_{(a)}=1}^{E\ell_n}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right|\right. \\
&\left.+\frac{1}{(E\ell_n)^2}\sum_{j_{(i)}=1}^{E\ell_n}\sum_{h_{(i)}=1}^{E\ell_n}\sum_{b_{(a)}=1}^{E\ell_n}\left|\gamma^{(sd)}_{h_{(i)}b_{(a)},n}\right|+o(1)\right) \\
&= O\left(\frac{E\ell_n}{n}\right),
\end{aligned}
$$

where the first equality holds with the same procedure as (A.7). For M_{2n},

$$\begin{aligned}
M_{2n} &\leq \frac{1}{d_n^q}\left(\frac{1}{n}\sum_{(i,a)\in\mathcal{F}_2}\left|\gamma_{ia,n}^{(rc)}\right| d_{ia}^q\right)\left(\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{E\ell_n}\sum_{b_{(a)}=1}^{E\ell_n}\left|\gamma_{j_{(i)}b_{(a)},n}^{(sd)}\right|\right. \\
&\left.+\frac{1}{(E\ell_n)^2}\sum_{j_{(i)}=1}^{E\ell_n}\sum_{h_{(i)}=1}^{E\ell_n}\sum_{b_{(a)}=1}^{E\ell_n}\left|\gamma_{h_{(i)}b_{(a)},n}^{(sd)}\right| + o(1)\right) \\
&= O\left(d_n^{-q}\right)
\end{aligned}$$

Therefore, we obtain

$$M_{1n} = o(1) \text{ and } M_{2n} = o(1).$$

The next step is to show (A.9).

$$\begin{aligned}
&\frac{1}{nE\ell_n}E\left[\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}} K^2\left(\frac{d_{ia,n}^*}{d_n}\right)\gamma_{ia,n}^{(rc)}\left(\gamma_{j_{(i)}b_{(a)},n}^{(sd)} - \gamma_{\bar{\imath},b_{(a)},n}^{(sd)}\right)\right] \\
&= E\left[\left(\frac{1}{n}\sum_{i,a\in E_n}K^2\left(\frac{d_{ia,n}^*}{d_n}\right)\gamma_{ia,n}^{(rc)}\right)\left(\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left(\gamma_{j_{(i)}b_{(a)},n}^{(sd)} - \gamma_{\bar{\imath},b_{(a)},n}^{(sd)}\right)\right)\right] \\
&= o(1).
\end{aligned}$$

For some generic constant C

$$\begin{aligned}
&\left|\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left(\gamma_{j_{(i)}b_{(a)},n}^{(sd)} - \gamma_{\bar{\imath},b_{(a)},n}^{(sd)}\right)\right| \\
&\leq \left|\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b=1}^{n}\gamma_{j_{(i)}b,n}^{(sd)}1\left\{d_{ab,n}^* < d_n\right\}\right| \\
&+\left|\frac{1}{(E\ell_n)^2}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b,h=1}^{n}\gamma_{hb,n}^{(sd)}1\left\{d_{ab,n}^* < d_n, d_{ih,n}^* < d_n\right\}\right| \\
&\leq \frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b=1}^{n}\left|\gamma_{j_{(i)}b,n}^{(sd)}\right| + \frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sup_b\sum_{h=1}^{n}\left|\gamma_{hb,n}^{(sd)}\right|\frac{1}{E\ell_n}\sum_{b=1}^{n}1\left\{d_{ab,n}^* < d_n\right\} \\
&\leq \frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b=1}^{n}\left|\gamma_{j_{(i)}b,n}^{(sd)}\right| + \left(\sup_b\sum_{h=1}^{n}\left|\gamma_{hb,n}^{(sd)}\right|\right)\frac{\ell_{i,n}\ell_{a,n}}{(E\ell_n)^2} \leq C\frac{\ell_{i,n}}{E\ell_n} + C\frac{\ell_{i,n}\ell_{a,n}}{(E\ell_n)^2}.
\end{aligned}$$

By Assumption F9, $E\lim_{n\to\infty}\frac{\ell_{i,n}}{E\ell_n}=\lim_{n\to\infty}\frac{E\ell_{i,n}}{E\ell_n}\le C.$

As $\left|n^{-1}\sum_{i,a\in E_n}K^2\left(\frac{d^*_{ia,n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\right|<\infty$ for all n, invoking the dominated convergence theorem and Assumption F10 yields

$$\begin{aligned}
&\lim_{n\to\infty}E\left[\left(\frac{1}{n}\sum_{i,a\in E_n}K^2\left(\frac{d^*_{ia,n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\right)\left(\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left(\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\bar{\imath},b_{(a)},n}\right)\right)\right]\\
&=E\left[P\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i,a\in E_n}K^2\left(\frac{d^*_{ia,n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\right)\right.\\
&\left.\times\left(\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left(\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\bar{\imath},b_{(a)},n}\right)\right)\right]\\
&=0
\end{aligned}$$

The last equality holds because

$$\begin{aligned}
&\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left(\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\gamma^{(sd)}_{\bar{\imath},b_{(a)},n}\right)\\
&=\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\frac{1}{(E\ell_n)^2}\sum_{h_{(i)}=1}^{\ell_{i,n}}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\gamma^{(sd)}_{h_{(i)},b_{(a)},n}\\
&=\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}-\left(\frac{1}{E\ell_n}+o_p(1)\right)\sum_{h_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\gamma^{(sd)}_{h_{(i)},b_{(a)},n}\\
&\xrightarrow{p}0.
\end{aligned}$$

Hence the second term in (A.4) is $o_p(1)$.

By a symmetric argument, we obtain the result that

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

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

$$\lim_{n\to\infty}E\left[\frac{1}{nE\ell_n}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left[K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)-K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\right]^2\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{jb,n}\right]=0. \tag{A.10}$$

But

$$E\left[\frac{1}{nE\ell_n}\sum_{i,a\in E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left[K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)-K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\right]^2\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{jb,n}\right]$$

$$=E\left[\frac{1}{nE\ell_n}\sum_{(i,j(i),a,b(a))\in\mathcal{I}_1}\left[K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)-K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\right]^2\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{jb,n}\right]$$

$$+\frac{1}{nE\ell_n}\sum_{(i,j(i),a,b(a))\in\mathcal{I}_2}\left[K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)-K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\right]^2\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{jb,n}$$

$$\equiv F_{1n}+F_{2n},$$

where

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

and

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

For F_{1n}, we have

$$F_{1n}\le E\left|\frac{1}{nE\ell_n}\sum_{(i,j(i),a,b(a))\in\mathcal{I}_1}\left[K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)-K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\right]^2\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{jb,n}\right|$$

$$\le\frac{c_L^2}{nE\ell_n}\sum_{(i,j(i),a,b(a))\in\mathcal{I}_1}\left|\frac{d^*_{ij_{(i)},n}}{d_n}-\frac{d^*_{ab_{(a)},n}}{d_n}\right|^2\left|\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{jb,n}\right|$$

$$\le\frac{4c_L^2c_n^2}{d_n^2}\left(\frac{1}{n}\sum_{i,a\in E_n}\left|\gamma^{(rc)}_{ia,n}\right|\right)\left(\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left|\gamma^{(rc)}_{jb,n}\right|\right)=O\left(\frac{c_n^2}{d_n^2}\right)$$

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

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

and

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

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

$$
\begin{aligned}
F_{2n} &\leq E\left|\frac{1}{nE\ell_n}\sum_{(i,j(i),a,b(a))\in\mathcal{I}_2}\left[K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right) - K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\right]^2 \gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right| \\
&\leq E\frac{4}{nE\ell_n}\sum_{(i,j(i),a,b(a))\in\mathcal{I}_2}\left|\left(d^*_{ia,n}\right)^q \gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\left(d^*_{ia,n}\right)^{-q}\right| \\
&= E\frac{4}{nE\ell_n}\sum_{i\in E_n}\sum_{a:d^*_{ia,n}\in[c_n,d_n]}\left(d^*_{ia,n}\right)^q \left|\gamma^{(rc)}_{ia,n}\right|\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right|\left(d^*_{ia,n}\right)^{-q} \\
&\leq 4E\left(\frac{1}{n}\sum_{i,a\in E_n}\left(d^*_{ia,n}\right)^q\left|\gamma^{(rc)}_{ia,n}\right|\right)\left(\frac{1}{E\ell_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{b_{(a)}=1}^{\ell_{a,n}}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right|\right)c_n^{-q} \\
&= o(c_n^{-q}).
\end{aligned}
$$

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

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

and (A.3) is proved.

Next, we show that D_{2n} is $o\left(1\right)$. For D_{2n},

$$
\begin{aligned}
&\frac{1}{nE\ell_n}E\left[\sum_{i\notin E_n}\sum_{j_{(i)}=1}^{\ell_{i,n}}\sum_{a\in E_n}\sum_{b_{(a)}=1}^{\ell_{a,n}}K\left(\frac{d^*_{ij_{(i)},n}}{d_n}\right)K\left(\frac{d^*_{ab_{(a)},n}}{d_n}\right)\gamma^{(rc)}_{ia,n}\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right] \\
&\leq E\left[\left(\frac{1}{n}\sum_{i\notin E_n}\sum_{a\in E_n}\left|\gamma^{(rc)}_{ia,n}\right|\right)\left(\frac{1}{E\ell_n}\sum_{b_{(a)}=1}^{\ell_{a,n}}\sum_{j_{(i)}=1}^{\ell_{i,n}}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right|\right)\right] \\
&= \left(\frac{1}{n}\sum_{i\notin E_n}\sum_{a\in E_n}\left|\gamma^{(rc)}_{ia,n}\right|\right)\left(\frac{1}{E\ell_n}\sum_{b_{(a)}=1}^{E\ell_n}\sum_{j_{(i)}=1}^{E\ell_n}\left|\gamma^{(sd)}_{j_{(i)}b_{(a)},n}\right| + o\left(1\right)\right) \\
&= o\left(1\right),
\end{aligned}
$$

by choosing the sequence of d_n in a way that $n_2/n \to 0$ as $n \to \infty$. We can show D_{3n} and D_{4n} are $o\left(1\right)$ in the similar way.

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

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

In terms of matrix form,

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

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

(b) Asymptotic Bias

By Assumption F8(ii) and the dominated convergence theorem, we have

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

Therefore,

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

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

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

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

$$\sqrt{\frac{n}{E\ell_n}}\left(\hat{J}_n - \tilde{J}_n\right) = \sqrt{\frac{n}{E\ell_n}}\frac{1}{n}\sum_{i,j=1}^{n} K\left(\frac{d^*_{ij,n}}{d_n}\right)\left[\hat{V}_{i,n}\hat{V}'_{j,n} - V_{i,n}V'_{j,n}\right]$$

$$\equiv 2L_{1,n}\sqrt{n}\left(\hat{\theta}-\theta_0\right) + \sqrt{n}\left(\hat{\theta}-\theta_0\right)' L_{2,n}\sqrt{n}\left(\hat{\theta}-\theta_0\right)$$

$$+ \sqrt{n}\left(\hat{\theta}-\theta_0\right)' L_{3,n}\sqrt{n}\left(\hat{\theta}-\theta_0\right) \tag{A.11}$$

where

$$L_{1,n} = \sqrt{\frac{E\ell_n}{n}}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} V_{j,n}\left(\frac{1}{E\ell_n}\sum_{i=1}^{n} K\left(\frac{d^*_{ij,n}}{d_n}\right)\frac{\partial}{\partial\theta'}V_{i,n}\right),$$

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

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

Therefore, under Assumption F12(i) it suffices to show that $L_{1,n} = o_p(1)$, $L_{2,n} = o_p(1)$ and $L_{3,n} = o_p(1)$.

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

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

$$\leq \left[\frac{1}{\sqrt{nE\ell_n}}\frac{1}{n}\sum_{i,j=1}^{n} 1(d^*_{ij,n}\leq d_n)\left\|\frac{\partial^2}{\partial\theta\partial\theta'}V_{i,n}(\overline{\theta})\right\|\left\|V_{j,n}(\overline{\theta})\right\|\right]^2$$

$$\leq \left[\sqrt{\frac{E\ell_n}{n}}\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}V_{i,n}(\overline{\theta})\right\|\left(\frac{1}{E\ell_n}\sum_{j=1}^{n} 1(d^*_{ij,n}\leq d_n)\left\|V_{j,n}(\overline{\theta})\right\|\right)\right]^2$$

$$\leq \frac{E\ell_n}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{\theta}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}V_{i,n}(\theta)\right\|^2\right)$$

$$\times\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{E\ell_n}\sum_{j=1}^{n} 1(d^*_{ij,n}\leq d_n)\left\|V_{j,n}(\overline{\theta})\right\|\right)^2$$

$$= \frac{E\ell_n}{n}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{E\ell_n}\sum_{j=1}^{n} 1(d^*_{ij,n}\leq d_n)\left\|V_{j,n}(\overline{\theta})\right\|\right)^2 O_p(1) \tag{A.12}$$

where the last equality follows from Assumption F12(iv). By Assumption F12(ii), we have

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

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

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

uniformly over i. Hence $L_{2,n}=o_p(1)$. Using the same procedure, we can show that $L_{3,n}=o_p(1)$ under Assumption F12(iii).

The next step is to show $L_{1,n}=o_p(1)$. By Markov inequality and Assumption F12(v):

$$
\begin{aligned}
&P\left(\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}V_{j,n}\left(\frac{1}{E\ell_n}\sum_{i=1}^{n}K\left(\frac{d^*_{ij,n}}{d_n}\right)\frac{\partial}{\partial\theta'}V_{i,n}\right)\right\|>\delta\right)\\
&\leq\frac{1}{\delta^2}E\left[\frac{1}{nE\ell_n^2}\sum_{i,j,a,b=1}^{n}K\left(\frac{d^*_{ij,n}}{d_n}\right)K\left(\frac{d^*_{ab,n}}{d_n}\right)V_{j,n}V_{b,n}\frac{\partial}{\partial\theta'}V_{i,n}\frac{\partial}{\partial\theta}V_{a,n}\right]\\
&\leq\frac{1}{\delta^2}\frac{1}{nE\ell_n^2}\sum_{i,j,a,b=1}^{n}\left|E\left(K\left(\frac{d^*_{ij,n}}{d_n}\right)K\left(\frac{d^*_{ab,n}}{d_n}\right)\right)\right|\left\|E\left(V_{j,n}V_{b,n}\frac{\partial}{\partial\theta'}V_{i,n}\frac{\partial}{\partial\theta}V_{a,n}\right)\right\|\\
&\leq\frac{1}{\delta^2}\frac{1}{nE\ell_n^2}\sum_{i,j,a,b=1}^{n}E\left[1(d^*_{ij,n}\leq d_n)1(d^*_{ab,n}\leq d_n)\right]\left\|E\left(V_{j,n}V_{b,n}\frac{\partial}{\partial\theta'}V_{i,n}\frac{\partial}{\partial\theta}V_{a,n}\right)\right\|\\
&\leq\frac{1}{\delta^2}E\left[\frac{\ell_{j,n}\ell_{b,n}}{E\ell_n^2}\frac{1}{n\ell_{j,n}\ell_{b,n}}\sum_{j,b=1}^{n}\sum_{i_{(j)}=1}^{\ell_{j,n}}\sum_{a_{(b)}=1}^{\ell_{b,n}}\left\|E\left(V_{j,n}V_{b,n}\frac{\partial}{\partial\theta'}V_{i_{(j)},n}\frac{\partial}{\partial\theta}V_{a_{(b)},n}\right)\right\|\right]\\
&\leq\frac{1}{\delta^2}\left(\frac{E\ell_{j,n}\ell_{b,n}}{E\ell_n^2}\right)\sup_{i,a,b}\sum_{j=1}^{n}\left\|E\left(V_{j,n}V_{b,n}\frac{\partial}{\partial\theta'}V_{i,n}\frac{\partial}{\partial\theta}V_{a,n}\right)\right\| \qquad \text{(A.13)}\\
&\leq\frac{C}{\delta^2}\frac{\left(E\ell_{j,n}^2\right)^{1/2}\left(E\ell_{b,n}^2\right)^{1/2}}{E\ell_n^2}\leq\frac{C'}{\delta^2}
\end{aligned}
$$

where the last inequality follows from Assumption F9. Combining this with the definition of L_{1n}, we obtain $L_{1,n}=O_p\left(\sqrt{\frac{E\ell_n}{n}}\right)=o_p(1)$.

(d) Asymptotic Truncated MSE

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

Lemma A4. *If $\{\xi_n\}$ is bounded sequence of random variables such that $\xi_n \xrightarrow{p} 0$, then $E\xi_n \to 0$.*

Lemma A5. *Let $\{X_n\}$ be a sequence of nonnegative rv's for which $\sup_{n\geq 1}EX_n^{1+\delta}<\infty$ for some $\delta>0$. Then, $\lim_{h\to\infty}\lim_{n\to\infty}(E\min\{X_n,h\}-EX_n)=0$.*

In our setting,

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

as $n \to \infty$. Here the $o_p(1)$ term follows from Theorem 1(c). Also $|\xi_n| \leq h$. By Lemma A4 , $E\xi_n \to 0$. Since this holds for all h, the first equality of Theorem 1(d) holds.

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

$$\lim_{h\to\infty}\lim_{n\to\infty}\left(MSE_h\left(\frac{n}{E\ell_n}, \tilde{J}_n, S_n\right) - MSE_h\left(\frac{n}{E\ell_n}, \tilde{J}_n, S\right)\right) = 0 \qquad \text{(A.14)}$$

$$\lim_{h\to\infty}\lim_{n\to\infty} MSE_h\left(\frac{n}{E\ell_n}, \tilde{J}_n, S\right) = \lim_{n\to\infty} MSE\left(\frac{n}{E\ell_n}, \tilde{J}_n, S\right). \qquad \text{(A.15)}$$

Under Assumption F12(ii), (A.14) holds by applying Lemma A4. Equation (A.15) holds by applying Lemma A5 with

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

It is easy to see that $\sup_{n\geq 1} EX_n^2 < \infty$, as required by Lemma A5, if $\forall r, s \leq p$ $E\left[\sqrt{\frac{n}{E\ell_n}}(\tilde{J}_{rs,n} - J_{rs,n})\right]^4 = O(1)$. Note that

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

where

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

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

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

$$D_{1n}=E\left[I_1^4\right],\ D_{2n}=E\left[I_1^3I_2\right],\ D_{3n}=E\left[I_1^2I_2^2\right],\ D_{4n}=E\left[I_1I_2^3\right], D_{5n}=E\left[I_2^4\right].$$

By Theorem 1(a) and (b), it is straightforward to show that D_{3n}, D_{4n} and D_{5n} are $O(1)$ under Assumption A.17(ii). The proofs for D_{1n} and D_{2n} are similar and we focus only on D_{1n} here. As denoted before, let $\varphi_{lkrs,n}=\sum_{i=1}^{n}\sum_{j=1}^{n}K(\frac{d^*_{ij,n}}{d_n})r^{(r)}_{il,n}r^{(s)}_{jk,n}$. Then,

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

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

$$\begin{aligned}
G_{1n} &= \frac{1}{n^2\left(E\ell_n\right)^2}E\sum_{l_1,l_2,l_3,l_4}\varphi_{l_1l_1rs,n}\varphi_{l_2l_2rs,n}\varphi_{l_3l_3rs,n}\varphi_{l_4l_4rs,n}E\tilde{\varepsilon}^2_{l_1,n}\tilde{\varepsilon}^2_{l_2,n}\tilde{\varepsilon}^2_{l_3,n}\tilde{\varepsilon}^2_{l_4,n}\\
&= \frac{1}{n^2\left(E\ell_n\right)^2}\left[E\left(\sum_l\varphi^4_{llrs,n}\right)+3E\sum_{l_1\neq l_2}\varphi^2_{l_1l_1rs,n}\varphi^2_{l_2l_2rs,n}\right]\\
&\le \frac{C}{n^2\left(E\ell_n\right)^2}E\left(\sum_l\varphi^2_{llrs,n}\right)^2=O\left[\frac{1}{\left(E\ell_n\right)^2}\right]=o(1)
\end{aligned}$$

using the fact that

$$\begin{aligned}
\sum_l\varphi^2_{llrs,n} &= \sum_l\left(\sum_{i=1}^n\sum_{j=1}^nK(\frac{d^*_{ij,n}}{d_n})r^{(r)}_{il,n}r^{(s)}_{jl,n}\right)^2\\
&\le \sum_l\left(\sum_{i=1}^n\left|r^{(r)}_{il,n}\right|\right)^2\left(\sum_{j=1}^n\left|r^{(r)}_{jl,n}\right|\right)^2=O(n)
\end{aligned}$$

by Assumption F6.

For G_{2n}, we have:

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

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

$$\frac{1}{n^2 (E\ell_n)^2} E \sum_{l_1=1}^{np} \sum_{l_2 \neq l_1}^{np} \varphi_{l_1 l_2 rs,n}^4 < \infty, \tag{A.16}$$

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

$$\frac{1}{n^2 (E\ell_n)^2} E \sum_{l_1=1}^{np} \sum_{l_2=1}^{np} \sum_{l_3=1}^{np} \sum_{l_4=1}^{np} \varphi_{l_1 l_2 rs,n} \varphi_{l_2 l_4 rs,n} \varphi_{l_4 l_3 rs,n} \varphi_{l_3 l_1 rs,n} < \infty. \tag{A.18}$$

Equation (A.16) is true because

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

where the last equality follows from Assumption F6.

For equation (A.17), we have

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

and

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

using equations (A.2) and (A.3). Hence (A.17) holds.

Finally, for equation (A.18), we note that

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

and

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

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

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

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

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

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

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

we have

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

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

Proof of Corollary 1

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

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

by Theorem 1(d), we obtain

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

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

Proof of Theorem 2

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

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

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

where

$$\begin{aligned}
&M_{1n}\\
&=\sqrt{\frac{n}{E\ddot{\ell}_n}}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left(K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)-K\left(\frac{d^*_{ij_{(i)},n}}{\ddot{d}_n}\right)\right)\left(\hat{V}_{i,n}\hat{V}'_{j_{(i)},n}-V_{i,n}V'_{j_{(i)},n}\right)\right],\\
&M_{2n}\\
&=\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\left(K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)-K\left(\frac{d^*_{ij_{(i)},n}}{\ddot{d}_n}\right)\right)V_{i,n}V'_{j_{(i)},n}\right),\\
&M_{3n}\\
&=\sqrt{\frac{n}{E\ddot{\ell}_n}}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j_{(i)}=\ddot{\ell}_{i,n}+1}^{n}K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)\hat{V}_{i,n}\hat{V}'_{j_{(i)}}\right).
\end{aligned}$$

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

Note that

$$
\begin{aligned}
&\|M_{1n}\| \\
&= \sqrt{\frac{E\ddot{\ell}_n}{n}} \left\| \frac{1}{E\ddot{\ell}_n} \sum_{i=1}^{n} \sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}} \left(K\left(\frac{d^*_{ij,n}}{\hat{d}_n}\right) - K\left(\frac{d^*_{ij,n}}{\ddot{d}_n}\right)\right) \left(\frac{\partial V_{i,n}(\bar{\theta})}{\partial \theta'} V_{j_{(i)},n}(\bar{\theta}) \right.\right. \\
&\left.\left. + V_{i,n}(\bar{\theta}) \frac{\partial V_{j_{(i)},n}(\bar{\theta})}{\partial \theta'} \right) \left(\hat{\theta} - \theta\right) \right\| \\
&\leq \sqrt{\frac{E\ddot{\ell}_n}{n}} O_p\left(1\right) \frac{1}{E\ddot{\ell}_n \sqrt{n}} \sum_{i=1}^{n} \sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}} \left| \frac{d^*_{ij,n}}{\hat{d}_n} - \frac{d^*_{ij,n}}{\ddot{d}_n} \right| \left(\left\| \frac{\partial}{\partial \theta} V_{i,n}(\bar{\theta}) V_{j_{(i)},n}(\bar{\theta}) \right\| \right. \\
&\left. + \left\| V_{i,n}(\bar{\theta}) \frac{\partial}{\partial \theta} V_{j_{(i)},n}(\bar{\theta}) \right\| \right) \\
&\leq o_p\left(1\right) \sqrt{n} \left| \frac{\ddot{d}_n}{\hat{d}_n} - 1 \right| \frac{1}{E\ddot{\ell}_n n} \sum_{i=1}^{n} \sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}} \left(\left\| \frac{\partial}{\partial \theta} V_{i,n}(\bar{\theta}) V_{j_{(i)},n}(\bar{\theta}) \right\| \right. \qquad \text{(A.19)} \\
&\left. + \left\| V_{i,n}(\bar{\theta}) \frac{\partial}{\partial \theta} V_{j_{(i)},n}(\bar{\theta}) \right\| \right) \\
&= o_p\left(1\right) \frac{\ddot{\ell}_{i,n}}{E\ddot{\ell}_n} \frac{1}{\ddot{\ell}_{i,n} n} \sum_{i=1}^{n} \sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}} \left(\left\| \frac{\partial}{\partial \theta} V_{i,n}(\bar{\theta}) V_{j_{(i)},n}(\bar{\theta}) \right\| + \left\| V_{i,n}(\bar{\theta}) \frac{\partial}{\partial \theta} V_{j_{(i)},n}(\bar{\theta}) \right\| \right), \qquad \text{(A.20)}
\end{aligned}
$$

where the first inequality uses Assumption F11(ii) and the $O_p\left(1\right)$ and $o_p\left(1\right)$ terms hold as $\sqrt{n}\left(\hat{\theta} - \theta\right) = O_p(1)$, $\sqrt{n}\left(\ddot{d}_n/\hat{d}_n - 1\right) = O_p(1)$. Since $\ddot{\ell}_{i,n}/E\ddot{\ell}_n = O_p(1)$, it now suffices to show that

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

Using Assumption F12(ii) and (iii), we have

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

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

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

$$
\begin{aligned}
&P\left(\left\|\frac{1}{\sqrt{nE\ddot{\ell}_n}}\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{in}}\left(K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)-K\left(\frac{d^*_{ij_{(i)},n}}{\ddot{d}_n}\right)\right)V_{i,n}V_{j_{(i)},n}\right\|>\delta\right)\\
&\leq\frac{1}{\delta^2}\frac{1}{nE\ddot{\ell}_n}E\sum_{i=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\sum_{a=1}^{n}\sum_{b_{(a)}=1}^{\ddot{\ell}_{a,n}}\left\|E\left(V_{i,n}V_{j_{(i)},n}V_{a,n}V_{b_{(a)},n}\right)\right\|\\
&\times\left|\left(K\left(\frac{d^*_{ij_{(i)},n}}{\hat{d}_n}\right)-K\left(\frac{d^*_{ij_{(i)},n}}{\ddot{d}_n}\right)\right)\left(K\left(\frac{d^*_{ab_{(a)},n}}{\hat{d}_n}\right)-K\left(\frac{d^*_{ab_{(a)},n}}{\ddot{d}_n}\right)\right)\right|\\
&\leq\frac{C}{\delta^2}\frac{1}{nE\ddot{\ell}_n}E\sum_{i,a=1}^{n}\sum_{j_{(i)}=1}^{\ddot{\ell}_{i,n}}\sum_{b_{(a)}=1}^{\ddot{\ell}_{a,n}}\left\|E\left(V_{i,n}V_{j_{(i)},n}V_{a,n}V_{b_{(a)},n}\right)\right\|\frac{d^*_{ij_{(i)},n}}{\ddot{d}_n}\frac{d^*_{ab_{(a)},n}}{\ddot{d}_n}\left(\frac{\ddot{d}_n}{\hat{d}_n}-1\right)^2\\
&\leq\frac{C}{\delta^2}\frac{1}{E\ddot{\ell}_n}\frac{1}{n^2}\sum_{i,j,a,b=1}^{n}\left\|E\left(V_{i,n}V_{j,n}V_{a,n}V_{b,n}\right)\right\|. \qquad \text{(A.21)}
\end{aligned}
$$

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

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i,j,a,b=1}^{n}\left\|E\left(V_{i,n}V_{j,n}V_{a,n}V_{b,n}\right)\right\| \\
&=\frac{1}{n^2}\sum_{i,j,a,b=1}^{n}\left|E\left(\sum_{\ell_1=1}^{np}r_{i\ell_1,n}\varepsilon_{\ell_1}\right)\left(\sum_{\ell_2=1}^{np}r_{j\ell_2,n}\varepsilon_{\ell 2}\right)\left(\sum_{\ell_3=1}^{np}r_{a\ell_3,n}\varepsilon_{\ell 3}\right)\right. \\
&\left.\times\left(\sum_{\ell_4=1}^{np}r_{b\ell_2,n}\varepsilon_{\ell 4}\right)\right| \\
&=\frac{1}{n^2}\sum_{i,j,a,b=1}^{n}\sum_{\ell=1}^{np}\left|r_{i\ell,n}r_{j\ell,n}r_{a\ell,n}r_{b\ell,n}\right|E\varepsilon_{\ell}^{4}+\frac{1}{n^2}\sum_{i,j,a,b=1}^{n}\left|\gamma_{i,a}\right|\left|\gamma_{j,b}\right| \\
&+\frac{1}{n^2}\sum_{i,j,a,b=1}^{n}\left|\gamma_{i,j}\right|\left|\gamma_{a,b}\right|+\frac{1}{n^2}\sum_{i,j,a,b=1}^{n}\left|\gamma_{i,b}\right|\left|\gamma_{j,a}\right| \\
&\leq C\left(\sum_{i=1}^{n}\left|r_{i\ell,n}\right|\right)^4\frac{p}{n}+3\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\gamma_{i,j}\right|\right)^2=O\left(1\right), \qquad \text{(A.22)}
\end{aligned}$$

so the upper bound in (A.21) is $o\left(1\right)$, which implies that $M_{2n}=o_p\left(1\right)$.

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

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

But

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

using equation (A.22). Hence $M_{3n} = o_p(1)$. Consequently

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

The first equality of Theorem 2(b) holds by applying Lemma A4 in the same way as in proof of the first equality of Theorem 1(d). Then the second equality of Theorem 2(b) holds by Theorem 1(d).

Proof of Corollary 2

By Corollary 3 and Theorem 2(b),

$$\lim_{h\to\infty}\lim_{n\to\infty}\left(MSE_h\left(n^{2q/(2q+\eta)}, \hat{J}_n(\dot{d}_n), S_n\right) - MSE_h\left(n^{2q/(2q+\eta)}, \hat{J}_n(\hat{d}_n), S_n\right)\right)$$
$$= \lim_{h\to\infty}\lim_{n\to\infty}\left(MSE_h\left(n^{2q/(2q+\eta)}, \hat{J}_n(d_n), S_n\right) - MSE_h\left(n^{2q/(2q+\eta)}, \hat{J}_n(\ddot{d}_n), S_n\right)\right) \tag{A.23}$$

Since $\ddot{g} = g$ and $\ddot{g}^{(q)} = g^{(q)}$, $\ddot{d}_n = d_n^\star$. Corollary 1 implies that the expression in (A.23) is ≥ 0 with the inequality being strict unless $d_n = d_n^\star + o(n^{1/(2q+\eta)})$.

2.9 Bibliography

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Chapter 3

k-step Bootstrap Bias Correction for Fixed Effects Estimators in Nonlinear Panel Models

Panel data consists of repeated observations from different individuals across time. One virtue of this data structure is that we can control for unobserved time-invariant individual heterogeneity in an econometric model. When individual effects are correlated with explanatory variables, we may use the fixed effects estimator which treats each unobserved individual effect as a parameter to be estimated. However, this approach usually suffers from inconsistency when the time series sample size (T) is short. This is known as the *incidental parameters problem*, first noted by Neyman and Scott (1948). Furthermore, even though T grows at the same rate as n, the fixed effects estimators are asymptotically biased so that the inference drawn from them may give misleading results.

This paper proposes a k-step parametric bootstrap bias corrected maximum likelihood (ML) estimator of nonlinear static panel models. In the k-step bootstrap procedure, we approximate the standard bootstrap estimator by taking k-steps of a Newton-Raphson (NR) iterative scheme. We employ the original estimate as the starting point for the NR steps. We estimate the asymptotic bias using the k-step bootstrap method and subtract this from the original biased estimator. We prove that the standard and k-step bootstrap bias corrected estimators are asymptotically normal and centered at the true parameter if T grows faster than $\sqrt[3]{n}$. This condition is important in practice because many economic data sets nowadays are composed of small T and large n and therefore the usefulness of the bias corrected estimation particularly depends on how much of the bias is corrected in small T. Our Monte Carlo experiments show that in finite samples, the k-step bootstrap bias corrected estimators reduce the bias remarkably even for small T. This bias correction does not increase the asymptotic variance and thus bias correction substantially improves statistical inference. In addition to bias correcting the parameter estimators, we also apply the k-step bootstrap bias correction to the average marginal effect estimation.

The substantial advantage of our approach over alternatives is that our method enables us not only to correct the asymptotic bias but also to improve the coverage accuracy of the associated confidence intervals (CI). We construct the CI's using a double k-step bootstrap procedure. In Monte Carlo experiments we find that in finite samples the error in coverage probability of our CI's is smaller than those of the other standard alternatives especially when T is small. This is true for the estimators of the model parameters as well as the estimators of the average marginal effects.

Another clear advantage is ease of computation. Standard bootstrap methods in nonlinear models are usually very time-intensive because it is required to solve R nonlinear optimization problems to obtain R bootstrap estimates. R usually needs to be fairly large for the bootstrap method to be reliable. Unless the optimization problem is simple, this would be a very time-intensive task. Particularly, as the fixed effects approach treats the individual effects as parameters, there are many parameters to be estimated and computational intensiveness can be particularly serious in this type of models. For example, in our empirical application (not reported here), there are 1461 individuals, which means there are more than 1461 parameters to be estimated. In addition, the double bootstrap procedure which is used for constructing CI's in this method also increases computational intensiveness substantially. In order to overcome this problem, we introduce the k-step bootstrap estimation which only involves computing the Hessian and the score functions. We show that when $n \to \infty$ the stochastic difference between the standard and k-step bootstrap estimators is $O_p(T^{-2^{k-1}})$. When $k \geq 2$, this difference is of smaller order than the bias term we intend to remove. As a result, we can use the k-step bootstrap in place of the standard bootstrap to achieve bias reduction.

Several papers have discussed the difficulties involved in controlling for this incidental parameters problem in nonlinear panel models and have suggested bias correction methods. Lancaster (2000) and Arellano and Hahn (2006) give an overview on the subject. Anderson (1970) and Honoré and Kyriazidou (2000) propose estimators which do not depend on individual effects in some specific cases. However, their approaches are the exception rather than the rule and so usually provide no guidance in general cases to eliminate the bias from nonlinear panel models. More generally, Hahn and Newey(2004) (denoted HN hereinafter) and Fernández-Val (2009) propose jackknife and analytic procedures for nonlinear static models, while Hahn and Kuersteiner (2004) propose analytic estimators in nonlinear dynamic models. Both expand the estimator in orders of T and estimate the leading bias term using the sample analogue. Bester and Hansen (2008) propose a penalized objective function approach to solve this problem.

There is also a large literature on bootstrap bias corrected estimation. Hall (1992) introduces general bootstrap algorithms for bias correction and for the construction of CI's which we adapt in this paper. Hahn, Kuersteiner and Newey (2004) analyze the asymptotic properties of a bootstrap bias corrected ML estimator in cross sectional data and show that it is higher order efficient. Pace and Salvan (2006) suggest a bootstrap bias corrected estimator when there are nuisance parameters, but their algorithm is different from ours. While we estimate the asymptotic bias of the fixed effects estimator directly by bootstrap, they use the bootstrap procedure to adjust the profile likelihood function from which they obtain their bias corrected estimator. The k-step bootstrap procedure first appears in Davidson and Mackinnon (1999) and Andrews (2002, 2005), in which they prove its higher-order equivalence to the standard bootstrap for extremum estimators.

The paper is organized as follows. Section 3.1 reviews the incidental parameters problem in nonlinear panel models and introduces the analytic form of the bias term which is demonstrated in HN. Section 3.2 describes the bootstrap bias correction procedure. Section 3.3 explains the k-step bootstrap bias correction. Section 3.4 establishes the asymptotic properties of our estimators. Section 3.5 discusses bias correction for average marginal effects. The Monte Carlo simulation results are reported in Section 3.6. The last section concludes.

3.1 Incidental Parameters Problem

In this section, we introduce the incidental parameters problem and present the asymptotic bias of the fixed effects estimator in nonlinear panel models.

Consider a nonlinear panel data model:

$$z_{it} \sim f(z; \theta, \alpha_i)$$

where θ is $(L_\theta \times 1)$ vector of parameters of interest and α_i is a scalar individual heterogeneity and f is a probability density function with parameters θ and α_i. For any given parameter value, $\{z_{it}\}$ are independently distributed across $i = 1, 2, ..., n$ and $t = 1, 2, ..., T$. The model includes discrete choice models and censored and truncation models as special cases.

Denote the true values of θ and $\alpha \equiv (\alpha_1, ..., \alpha_n)$ by θ_0 and $\alpha_0 = (\alpha_{10}, ..., \alpha_{n0})$ respectively and let $l(\theta, \alpha_i; z_{it}) \equiv \log f(z_{it}; \theta, \alpha_i)$. The objective function for the fixed effects estimator, $\hat{\theta}_{nT}$, is the concentrated log-likelihood function based on

$\hat{\alpha}_i$. That is, we obtain $\hat{\theta}_{nT}$ by solving

$$\hat{\theta}_{nT} = \arg\max_{\theta} \sum_{i=1}^{n} \sum_{t=1}^{T} l(\theta, \hat{\alpha}_i(\theta); z_{it}), \tag{A.1}$$

where

$$\hat{\alpha}_i(\theta) = \arg\max_{\alpha_i} \sum_{t=1}^{T} l(\theta, \alpha_i; z_{it}) \tag{A.2}$$

$$\hat{\alpha}\,(\theta) \equiv (\hat{\alpha}_1(\theta), ..., \hat{\alpha}_i(\theta), ..., \hat{\alpha}_N(\theta))$$

and the maximization is taken over a compact set.

Equation (A.2) implies that the estimation of α_i uses only T time series observations $(z_{i1}, \ldots, z_{iT})$. Therefore, given that T is fixed, $\hat{\alpha}_i$ does not converge to α_{i0} even though $n \to \infty$. This estimation error of $\hat{\alpha}_i$ causes $\hat{\theta}_{nT}$ to be inconsistent, which means $\text{Plim}_{n\to\infty} \hat{\theta}_{nT} \neq \theta_0$. This is known as the incidental parameters problem first noted by Neyman and Scott (1948). From the asymptotic properties of extremum estimators (e.g. Amemiya (1985)), as $n \to \infty$ with T fixed

$$\hat{\theta}_{nT} \xrightarrow{p} \theta_T, \;\; \theta_T \equiv \; \arg\max_{\theta} \overline{E}\left[\sum_{t=1}^{T} l(\theta, \hat{\alpha}_i(\theta); z_{it})\right] \tag{A.3}$$

where

$$\overline{E}\left[\sum_{t=1}^{T} l(\theta, \hat{\alpha}_i(\theta); z_{it})\right] \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\left[\sum_{t=1}^{T} l(\theta, \hat{\alpha}_i(\theta); z_{it})\right].$$

Since θ_0 maximizes $E[\sum_{i=1}^{n}\sum_{t=1}^{T} l(\theta, \alpha_{i0}; z_{it})]$, usually $\theta_T \neq \theta_0$. If the likelihood function is smooth enough, we can show by stochastic expansion that

$$\theta_T = \theta_0 + \frac{B}{T} + O\left(\frac{1}{T^2}\right) \tag{A.4}$$

for some B. This implies that $\theta_T \to \theta_0$ as $T \to \infty$. However, it is still asymptotically biased if T grows at the same rate as n. That is, as $n, T \to \infty$ and $n/T \to \rho$,

$$\begin{aligned}\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) &= \sqrt{nT}(\hat{\theta}_{nT} - \theta_T) + \sqrt{nT}\frac{B}{T} + O_p\left(\sqrt{\frac{n}{T^3}}\right) \\ &\xrightarrow{d} N(0, \Omega) + B\sqrt{\rho} \overset{d}{=} N(B\sqrt{\rho}, \Omega),\end{aligned} \tag{A.5}$$

for some variance matrix Ω. $\sqrt{nT}(\hat{\theta}_{nT} - \theta_T)$ converges to a normal distribution centered at zero, since θ_T is the probability limit of $\hat{\theta}_{nT}$. However, the second term, $\sqrt{nT}B/T$, does not vanish but converges to $B\sqrt{\rho}$. Hence, statistical inference drawn from this will result in misleading conclusions even when T is as large as n.

HN establish the analytic form of the leading bias of $\hat{\theta}_{nT}$ using stochastic expansion. For notational convenience, we define

$$u_{it}(\theta, \alpha_i) \equiv \frac{\partial}{\partial \theta} l(\theta, \alpha_i; z_{it}) \quad \text{and} \quad v_{it}(\theta, \alpha_i) \equiv \frac{\partial}{\partial \alpha_i} l(\theta, \alpha_i; z_{it})$$

and let additional subscripts denote partial derivatives,
i.e. $v_{it\alpha}(\theta, \alpha_i) \equiv \frac{\partial^2}{\partial \alpha_i^2} l(\theta, \alpha_i; z_{it})$. We suppress the arguments of the functions such as u_{it}, when they are evaluated at the true value (θ_0, α_{i0}). HN show that in equation (A.5)

$$B = -H^{-1}(\theta_0, \alpha_0) b\left(\theta_0, \alpha_0\right), \tag{A.6}$$

where

$$H(\theta_0, \alpha_0) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E\left[U_{it} U_{it}'\right],$$
$$b\left(\theta_0, \alpha_0\right) = \frac{1}{2} \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \frac{E\left[V_{2it} U_{it}\right]}{E\left[v_{it}^2\right]},$$

and

$$U_{it} \equiv u_{it} - \frac{E[u_{it} v_{it}]}{E[v_{it}^2]} \cdot v_{it}, \quad V_{2it} \equiv v_{it}^2 + v_{it\alpha}.$$

3.2 Bootstrap Bias Correction

In this section, we provide the bias corrected estimator using a parametric bootstrap procedure. The parametric bootstrap is different from the nonparametric bootstrap in that the former utilizes the parametric structure of the DGP by replacing the original parameters with their estimators to generate bootstrap samples, while the latter generates them from the empirical distribution function.

Let $F^* \equiv F_{\hat{\theta}_{nT},\hat{\alpha}}$ denote the distribution function of bootstrap samples. We obtain F^* from F by replacing θ_0 and α_0, with $\hat{\theta}_{nT}$ and $\hat{\alpha} \equiv \hat{\alpha}(\hat{\theta}_{nT})$. Therefore, in the bootstrap world, $\hat{\theta}_{nT}$ and $\hat{\alpha}$ are the true parameters. Let $\{z^*_{it}\}$ denote the bootstrap sample drawn at random from F^*. Based on $\{z^*_{it}\}$, we can obtain the bootstrap estimators, $\hat{\theta}^*_{nT}$ and $\hat{\alpha}^*_i$ by ML estimation. That is,

$$\hat{\theta}^*_{nT} = \arg\max_{\theta} \sum_{i=1}^{n} \sum_{t=1}^{T} l(\theta, \hat{\alpha}^*_i(\theta); z^*_{it}), \tag{A.7}$$

where

$$\hat{\alpha}^*_i(\theta) = \arg\max_{\alpha_i} \sum_{t=1}^{T} l(\theta, \alpha_i; z^*_{it}), \tag{A.8}$$

and as before the maximization is taken over a compact set.

The intuition behind the bootstrap bias correction is that the bias of a bootstrap estimator is a good approximation to that of a true parameter estimator. Under some regularity conditions, as $n \to \infty$ and $T \to \infty$,

$$E^*(\hat{\theta}^*_{nT}) - \hat{\theta}_{nT} = \frac{B}{T} + O_p\left(\frac{1}{T^2}\right)$$

where[1]

$$E^*(\hat{\theta}^*_{nT}) = \lim_{R\to\infty} \frac{1}{R} \sum_{r=1}^{R} \hat{\theta}^{*(r)}_{nT},$$

and E^* is the expectation operator with respect to F^*. Therefore, the bootstrap bias corrected estimator can be defined as

$$\tilde{\theta}_{nT} = 2\hat{\theta}_{nT} - E^*\left(\hat{\theta}^*_{nT}\right). \tag{A.9}$$

The above estimator reduces the order of the magnitude of a bias from $O_p\left(T^{-1}\right)$ to $O_p\left(T^{-2}\right)$. To show this, we employ the same definitions of HN in the bootstrap world, i.e.

$$u^*_{it} \equiv \frac{\partial}{\partial\theta} l(\hat{\theta}_{nT}, \hat{\alpha}_i; z^*_{it}), v^*_{it} \equiv \frac{\partial}{\partial\alpha_i} l(\hat{\theta}_{nT}, \hat{\alpha}_i; z^*_{it}) \text{ and } v^*_{it\alpha} \equiv \frac{\partial^2}{\partial\alpha_i^2} l(\hat{\theta}_{nT}, \hat{\alpha}_i; z^*_{it}).$$

[1]Note that $E^*(\hat{\theta}^*_{nT})$ may not be finite, in which case we can introduce truncation to prevent it from going to infinity. See equation (A.17) for such a modification. When the truncation threshold goes to infinity at an appropriate rate, the introduction of truncation does not affect the limiting distribution. For simiplicity of exposition, we do not explicitly incorporate truncation here but do so for the k-step bootstrap estimator.

Then,

$$\underset{n\to\infty}{P^{*}\lim}\left(\hat{\theta}^{*}_{nT}-\hat{\theta}_{nT}-\frac{B^{*}}{T}\right)=O_{p}\left(\frac{1}{T^{2}}\right), \tag{A.10}$$

and

$$B^{*}=-\left[H^{*}(\hat{\theta}_{nT},\hat{\alpha}(\hat{\theta}_{nT}))\right]^{-1}b^{*}\left(\hat{\theta}_{nT},\hat{\alpha}(\hat{\theta}_{nT})\right)$$

where

$$H^{*}\left(\hat{\theta}_{nT},\hat{\alpha}(\hat{\theta}_{nT})\right)=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E^{*}\left[U^{*}_{it}U^{*\prime}_{it}\right],$$

$$b^{*}\left(\hat{\theta}_{nT},\hat{\alpha}(\hat{\theta}_{nT})\right)=\frac{1}{2}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{E^{*}\left[U^{*}_{it}V^{*}_{2it}\right]}{E^{*}\left[(v^{*}_{it})^{2}\right]}$$

and

$$U^{*}_{it}\equiv u^{*}_{it}-\frac{E^{*}[u^{*}_{it}v^{*}_{it}]}{E^{*}\left[(v^{*}_{it})^{2}\right]}\cdot v^{*}_{it},\quad V^{*}_{2it}\equiv v^{*2}_{it}+v^{*}_{it\alpha}.$$

The conditional distribution of the bootstrap sample given the data or $(\hat{\theta}_{nT},\hat{\alpha}(\hat{\theta}_{nT}))$ is the same as the distribution of the original sample except that the former uses $(\hat{\theta}_{nT},\hat{\alpha}(\hat{\theta}_{nT}))$ rather than (θ_0,α_0) as true parameters[2]. Therefore, we have

$$H^{*}\left(\hat{\theta}_{nT},\hat{\alpha}(\hat{\theta}_{nT})\right)=H(\theta,\alpha)|_{(\theta,\alpha)=\left(\hat{\theta}_{nT},\hat{\alpha}(\hat{\theta}_{nT})\right)}$$

$$b^{*}\left(\hat{\theta}_{nT},\hat{\alpha}(\hat{\theta}_{nT})\right)=b(\theta,\alpha)|_{(\theta,\alpha)=\left(\hat{\theta}_{nT},\hat{\alpha}(\hat{\theta}_{nT})\right)}$$

By stochastic expansion, we show in Appendix I that

$$H(\theta,\alpha)|_{(\theta,\alpha)=\left(\hat{\theta}_{nT},\hat{\alpha}(\hat{\theta}_{nT})\right)}=H(\theta_0,\alpha_0)\left[1+O_p\left(\frac{1}{T}\right)\right],$$

$$b(\theta,\alpha)|_{(\theta,\alpha)=\left(\hat{\theta}_{nT},\hat{\alpha}(\hat{\theta}_{nT})\right)}=b(\theta_0,\alpha_0)\left[1+O_p\left(\frac{1}{T}\right)\right].$$

Combining the above two equations with the definition of B^{*}, we obtain

$$B^{*}=B+O_p\left(\frac{1}{T}\right). \tag{A.11}$$

As a final step, we show that under the assumptions given in Section 3.4

$$\sqrt{nT}\left(\underset{n\to\infty}{P^{*}\lim}\hat{\theta}^{*}_{nT}-E^{*}(\hat{\theta}^{*}_{nT})\right)=o_p\left(1\right). \tag{A.12}$$

[2]When $(\hat{\theta}_{nT},\hat{\alpha}(\hat{\theta}_{nT}))$ is regarded as a vector, it is understood to be $(\hat{\theta}^{\prime}_{nT},\hat{\alpha}(\hat{\theta}_{nT})^{\prime})^{\prime}$. For notational simplicity, we omit the transpose notation if confusion is unlikely.

From (A.10), (A.11) and (A.12), for $T/\sqrt[3]{n} \to \infty$,

$$\begin{aligned}\sqrt{nT}(\tilde{\theta}_{nT} - \theta_0) &= \sqrt{nT}(\hat{\theta}_{nT} - \theta_T) + \sqrt{nT}\left[(\theta_T - \theta_0) - (P^* \lim_{n\to\infty} \hat{\theta}^*_{nT} - \hat{\theta}_{nT})\right] \\ &+ \sqrt{nT}\left[P^* \lim_{n\to\infty} \hat{\theta}^*_{nT} - E^*(\hat{\theta}^*_{nT})\right] \\ &= \sqrt{nT}(\hat{\theta}_{nT} - \theta_T) + o_p(1) \xrightarrow{d} N(0, \Omega). \qquad \text{(A.13)}\end{aligned}$$

This implies that the bootstrap bias corrected estimator removes the dominant bias and is asymptotically unbiased.

3.3 k-step Bootstrap Bias Correction

In this section, we define the k-step bootstrap bias corrected estimator and demonstrate its higher order equivalence to the standard bootstrap estimator.

The k-step procedure approximates $\hat{\theta}^*_{nT}$ by the NR iterative procedure. Let $\hat{\theta}^*_{nT,k}$ and $\hat{\alpha}^*_{i,k}$ denote the k-step bootstrap estimator. We define $\hat{\theta}^*_{nT,k}$ and $\hat{\alpha}^*_k$ recursively in the following way:

$$\begin{pmatrix} \hat{\theta}^*_{nT,k} \\ \hat{\alpha}^*_k \end{pmatrix} = \begin{pmatrix} \hat{\theta}^*_{nT,k-1} \\ \hat{\alpha}^*_{k-1} \end{pmatrix} - H^{-1}_{k-1} S_{k-1} \qquad \text{(A.14)}$$

where[3]

$$H_{k-1} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left. \frac{\partial^2 \log l(\theta, \alpha_i; z^*_{it})}{\partial(\theta', \alpha')\,\partial(\theta', \alpha')'} \right|_{\theta=\hat{\theta}^*_{nT,k-1}, \alpha=\hat{\alpha}^*_{k-1}} \qquad \text{(A.15)}$$

$$S_{k-1} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left. \frac{\partial \log l(\theta, \alpha_i; z^*_{it})}{\partial(\theta', \alpha')'} \right|_{\theta=\hat{\theta}^*_{nT,k-1}, \alpha=\hat{\alpha}^*_{k-1}}$$

and the start-up estimator $\hat{\theta}^*_{nT,0} = \hat{\theta}_{nT}$, $\hat{\alpha}^*_0 = \hat{\alpha}$.

In the appendix of proofs, we show that

$$P^* \lim_{n\to\infty} \hat{\theta}^*_{nT,k} = P^* \lim_{n\to\infty} \hat{\theta}^*_{nT} + O_p\left(\frac{1}{T^{2^{k-1}}}\right). \qquad \text{(A.16)}$$

[3] The Hessian matrix we used is called the observed Hessian. We note that some terms in $\partial^2 \log l(\theta, \alpha_i; z^*_{it}) / \partial(\theta', \alpha')\,\partial(\theta', \alpha')'$ have zero expectation. Dropping these terms in equation (A.15), we obtain the expected Hessian. Our asymptotic results remain valid for the expected Hessian, as the dropped terms are of smaller order.

This implies the quadratic convergence of $\hat{\theta}^*_{nT,k}$ to $\hat{\theta}^*_{nT}$ as k increases. In particular, when $k \geq 2$, $P^* \lim_{n\to\infty} \hat{\theta}^*_{nT,k} = P^* \lim_{n\to\infty} \hat{\theta}^*_{nT} + O_p\left(1/T^2\right)$. So in large samples, the approximation error in using the k-step bootstrap instead of the standard bootstrap is of smaller order than the bias term that we intend to remove. Therefore, condition $k \geq 2$ is necessary for the k-step bootstrap to achieve effective bias reduction.

To implement the k-step bootstrap, we have to invert the Hessian matrix. Depending on the observations we sample, H_{j-1} may be close to be singular in practice, in which case $\hat{\theta}^*_{nT,j}$ goes to infinity. As a result, the mean of $\hat{\theta}^*_{nT,k}$ may not be finite. To circumvent the undue influence of the second derivative of the objective function on our estimator, we introduce the truncated version, $\check{\theta}^*_{nT,k}$. The truncated estimator is defined as

$$\check{\theta}^*_{nT,k} \equiv \hat{\theta}_{nT} + \left(\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right) 1\left(\sqrt{nT}\left|\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right| \leq M_{nT}\right). \tag{A.17}$$

$\check{\theta}^*_{nT,k}$ yields the same value as $\hat{\theta}^*_{nT,k}$ when the difference of $\hat{\theta}^*_{nT,k}$ from $\hat{\theta}_{nT}$ is bounded by $M_{nT}/\sqrt{nT}$, but does not blow up when it has an infinite value. Similarly, we define

$$\check{\alpha}^*_{i,k} \equiv \hat{\alpha}_i + \left(\hat{\alpha}^*_{i,k} - \hat{\alpha}_i\right) 1\left(\sqrt{nT}\left|\hat{\alpha}^*_{i,k} - \hat{\alpha}_i\right| \leq M_{nT}\right). \tag{A.18}$$

We can set M_{nT} large enough that this truncation does not affect the asymptotic properties. We show that when $M_{nT} \to \infty$ such that $\sqrt{n/T} = o(M_{nT})$, we have:

$$P^* \lim_{n\to\infty} \check{\theta}^*_{nT,k} = P^* \lim_{n\to\infty} \hat{\theta}^*_{nT,k} + O_p\left(\frac{1}{T^2}\right).$$

As a final step, we show that

$$\sqrt{nT}\left(P^* \lim_{n\to\infty} \check{\theta}^*_{nT,k} - E^*(\check{\theta}^*_{nT,k})\right) = o_p\left(1\right). \tag{A.19}$$

From (A.16) and (A.19) the limiting distribution of the bootstrap bias corrected estimator, $\tilde{\theta}_{nT}$, which is defined in (A.9), will be invariant even though we replace $\hat{\theta}^*_{nT}$ with $\check{\theta}^*_{nT,k}$. Hence, we can define our truncated k-step bootstrap biased corrected estimator as

$$\tilde{\theta}_{nT,k} \equiv 2\hat{\theta}_{nT} - E^*\left(\check{\theta}^*_{nT,k}\right). \tag{A.20}$$

Then for $T/\sqrt[3]{n} \to \infty$ and all $k \geq 2$,

$$\begin{aligned}\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0) &= \sqrt{nT}(\hat{\theta}_{nT} - \theta_T) + \sqrt{nT}\left[(\theta_T - \theta_0) - (P^* \lim_{n\to\infty} \check{\theta}^*_{nT,k} - \hat{\theta}_{nT})\right] \\ &\quad + \sqrt{nT}\left(E^*(\check{\theta}^*_{nT,k}) - P^* \lim_{n\to\infty} \check{\theta}^*_{nT,k}\right) \\ &= \sqrt{nT}(\hat{\theta}_{nT} - \theta_T) + o_p(1) \xrightarrow{d} N(0, \Omega). \qquad \text{(A.21)}\end{aligned}$$

3.4 Asymptotic Properties

In this section, we state the assumptions and rigorously establish the asymptotic properties of the standard bootstrap and k-step bootstrap estimators.

For ease of exposition, we write $l(\theta, \alpha; z_{it}) \equiv l(\theta, \alpha_i; z_{it})$ so that $l(\cdot, \cdot; z_{it})$ is regarded as a function of θ and α. We maintain the following assumptions:

Assumption 1 $n, T \to \infty$ such that $n = o(T^3)$ and $T = O(n)$.

Assumption 2 (i) $l(\theta, \alpha; z_{it})$ is continuous in $(\theta, \alpha) \in \Theta$; (ii) the parameter space Θ is compact; (iii) (θ_0, α_0) is an interior point in Θ.

Assumption 3 For each $\eta > 0$, there exists $\delta > 0$ such that

$$\inf_i \left[G_i(\theta_0, \alpha_0) - \sup_{\{(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_0)|>\eta\}} G_i(\theta, \alpha) \right] \geq \delta > 0,$$

where

$$G_i(\theta, \alpha) \equiv E\frac{1}{T}\sum_{t=1}^{T} l(\theta, \alpha; z_{it}).$$

Assumption 4 (i) $l(\theta, \alpha; z_{it})$ is continuously differentiable to six orders; (ii) there exists some $M(z_{it})$ such that

$$\left| \frac{\partial^{m_1+m_2} l(\theta, \alpha; z_{it})}{\partial \theta_j^{m_1} \partial \alpha_i^{m_2}} \right| \leq M(z_{it}), \quad 0 \leq m_1 + m_2 \leq 6$$

(iii) For some $Q > 64$, $E\left[M(z_{it})^Q\right] < C$ for a constant C and all $i = 1, 2, ..., N$.

Assumption 5 (i) $\overline{E}[\mathcal{I}_i] \equiv \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E[U_{it}U_{it}']$ exists and is positive definite (ii) $\min_i E[v_{it}^2] > 0$.

Assumption 6 Let $B_n = (n^{-1}\sum_{i=1}^{n} E[U_{it}U_{it}'])^{-1}\left(n^{-1}\sum_{i=1}^{n}\frac{E[V_{2it}U_{it}]}{E\left[v_{it}^2\right]}\right)$, then $B_n = B + O\left(1/\sqrt{n}\right)$.

Assumption 1 shows that our estimator is applicable as long as T grows faster than $\sqrt[3]{n}$. This implies that our asymptotic theory is valid with relatively small T and large n, which is often the case in micro panel data sets. Assumption 2 is a standard regularity assumption. Assumption 3 is the identification assumption for extremum estimators. Assumption 4 is the same as Condition 4 in Newey and Hahn (2004). It is stronger than the moment assumption for extremum estimators and under this assumption the asymptotic bias depends on the second order expansion and higher order terms go to 0 under Assumption 1. Assumption 5 allows us to invoke the central limit theorem. Assumption 6 ensures that the limiting bias term B is close to its finite sample analogue B_n. This assumption holds trivially if z_{it} are iid across i.

Theorem 9. *Under Assumptions 1-6,*

$$\sqrt{nT}(\tilde{\theta}_{nT} - \theta_0) \xrightarrow{d} N(0, (\overline{E}[\mathcal{I}_i])^{-1}).$$

For the proof see Appendix I.

Proposition 3. *Under Assumptions 1-6, for all $k \geq 1$*

$$P^*\lim_{n\to\infty}\hat{\theta}^*_{nT,k} = P^*\lim_{n\to\infty}\hat{\theta}^*_{nT} + O_p\left(\frac{1}{T^{2^{k-1}}}\right).$$

For the proof see Appendix II.

Theorem 10. *Under Assumptions 1-6, for all $k \geq 2$*

$$\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0) \xrightarrow{d} N(0, (\overline{E}[\mathcal{I}_i])^{-1}).$$

For the proof see Appendix III.

3.5 Bias Correction for Average Marginal Effects

In this section, we suggest bias corrected estimators of the average marginal effects using the k-step bootstrap procedure. In nonlinear models, the average marginal effect may be as interesting as the model parameters because it summarizes the effect over certain sub-population, which is often the quantity of interest in empirical studies.

The first average marginal effect, which we refer to as "the fixed effect average" or simply the average marginal effect, is the marginal effect averaged over α_i. It is defined as

$$\mu(w) = \frac{1}{n}\sum_{i=1}^{n} m(w, \theta_0, \alpha_{i0})$$

where w is the value of the covariate vector where the average effect is desired. For example, in a probit model, $m(w, \theta_0, \alpha_{i0}) = \theta_{0(j)}\phi(x'\theta_0 + \alpha_{i0})$ where $\theta_{0(j)}$ and $\phi(\cdot)$ are the coefficient on the j-th regressor of interest and the standard normal density function respectively.

The bias uncorrected estimator of $\mu(w)$ is

$$\hat{\mu}_{nT}(w) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m(w, \hat{\theta}_{nT}, \hat{\alpha}_i). \tag{A.22}$$

As in the case for the estimation of model parameters, we can construct a k-step bootstrap bias corrected estimator of the fixed effect average by estimating the bias with the difference between $\hat{\mu}_{nT}(w)$ and its bootstrap estimator. Our k-step bootstrap bias corrected estimator of the fixed effects average is

$$\tilde{\mu}_{nT,k}(w) = \frac{2}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m(w, \hat{\theta}_{nT}, \hat{\alpha}_i) - E^*\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m(w, \check{\theta}^*_{nT,k}, \check{\alpha}^*_{i,k})\right]. \tag{A.23}$$

The second average marginal effect, which we refer to as "the overall average marginal effect", is the marginal effect averaged over both α_i and the covariates. It is defined as

$$\nu = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m(w_{it}, \theta_0, \alpha_{i0}).$$

See also Fernández-Val (2009). Similarly to equations (A.22) and (A.23), the original and bias corrected estimators of ν are

$$\hat{\nu}_{nT} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m(w_{it}, \hat{\theta}_{nT}, \hat{\alpha}_i),$$

$$\tilde{\nu}_{nT,k}(w) = \frac{2}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m(w_{it}, \hat{\theta}_{nT}, \hat{\alpha}_i) - E^*\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m(w_{it}, \check{\theta}^*_{nT,k}, \check{\alpha}^*_{i,k})\right].$$

3.6 Monte Carlo Study

In this section, we report our Monte Carlo experiment results, which show that k-step bootstrap bias correction reduces the bias significantly in finite samples and also improves the coverage accuracy of CI's.

For our Monte Carlo experiment, we employ the design used in Heckman (1981), Greene (2004), HN, and Fernández-Val(2009). It is based on the following probit model:

$$\begin{aligned} &Y_{it} = 1\{X_{it}\theta_0 + \alpha_i - \epsilon_{it} \geq 0\}; \quad \epsilon_{it} \sim N(0,1),\ \alpha_i \sim N(0,1), \\ &X_{it} = t/10 + X_{i,t-1}/2 + u_{it}; \quad u_{it} \sim U(-1/2, 1/2), \\ &n = 100;\ T = 4, 8, 12;\ \theta_0 = 1. \end{aligned}$$

As discussed in HN, this model does not fit completely within our framework. First, X_{it} is correlated overtime. The correlation does not cause any problem as we can use the conditional MLE approach and all the asymptotic results remain valid. Second, there is no correlation between X_{it} and α_i. This is different from the usual condition under which the fixed effects estimator is used. However the incidental parameters problem is still present as it has nothing to do with whether there is a correlation between X_{it} and α_i. The bias of the fixed effects estimator can be severe for fixed effects models as well as for random effects models. The effectiveness of different bias reduction methods can be well evaluated with our data generating process. Another reason to use this design is that it is widely cited and used in other simulation studies, which helps us compare our estimator with the alternatives.

The uncorrected estimator of model parameters is

$$
\begin{aligned}
&(\hat{\theta}_{nT}, \hat{\alpha}) \\
&= \arg\max_{\theta,\alpha} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} [y_{it} \log \Phi(x_{it}\theta + \alpha_i) + (1 - y_{it}) \log(1 - \Phi(x_{it}\theta + \alpha_i)]
\end{aligned}
$$

and the estimators of the average marginal effects are

$$
\hat{\mu}_{nT}(\bar{x}) = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{nT} \phi(\bar{x}\hat{\theta}_{nT} + \hat{\alpha}_i), \;\; \hat{\nu}_{nT} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\theta}_{nT} \phi(x_{it}\hat{\theta}_{nT} + \hat{\alpha}_i).
$$

where $\Phi(\cdot)$ is the standard normal distribution function and $\bar{x}$ is the sample mean of $\{x_{it}, i = 1, 2, ..., N, t = 1, ..., T\}$.

For the k-step bootstrap, we generate bootstrap samples based on $\hat{\theta}_{nT}$ and $\{\hat{\alpha}_i\}_{i=1}^{n}$ and estimate $\hat{\theta}^{*}_{nT,k}$ using (A.14) with the bootstrap samples. We repeat this procedure 1000 times ($R = 1000$). Then, we obtain the bias corrected k-step bootstrap estimator from (A.20). As discussed before, for each k value, we can use either observed Hessian or expected Hessian in the NR step, leading to two versions of the k-step procedure. Each simulation is repeated 1000 times.

We compare the performance of our bias-corrected estimator with four alternative bias correction estimators: the jackknife and the analytic bias corrected estimators by Hahn and Newey (2004) and the analytical bias-corrected estimator by Fernández-Val (2009). The jackknife bias-corrected estimator is denoted 'Jackknife'. For HN analytic estimators, there are two versions: the analytic bias-corrected estimator using Bartlett equalities, denoted 'BC1'; the analytic bias-corrected estimator based on general estimating equations, denoted 'BC2'. Fernández-Val's estimator is denoted as 'BC3'.

For each estimator, we report its mean, median, standard deviation, root mean squared errors, and the empirical sizes of two-sided nominal 5% and 10% tests. The tests are based on symmetric CI's, that is, we reject the null hypothesis if the parameter value under the null falls outside the CI's. For the jackknife and analytical bias correction procedures, the interval estimator or the testing method are the same as that given in the respective papers. For the k-step procedure, the CI's are based the double bootstrap procedure.

To describe the double bootstrap procedure, we focus on an element of θ. Hence, without loss of generality, we can consider the case that θ_0 is a scalar. By iterating the bootstrap procedure, we define:

$$\tilde{\theta}^*_{nT,k} \equiv 2\check{\theta}^*_{nT,k} - E^{**}(\check{\theta}^{**}_{nT,k}),$$

where $E^{**}(\check{\theta}^{**}_{nT,k})$ is defined on the double bootstrap, that is, the k-step bootstrap using $(\check{\theta}^*_{nT,k}, \check{\alpha}^*_{i,k})$ as the true model parameters. Similarly

$$\tilde{\alpha}^*_{i,k} \equiv 2\check{\alpha}^*_{i,k} - E^{**}(\check{\alpha}^{**}_{i,k}).$$

Let

$$t\text{-stat} = \frac{\sqrt{nT}\left|\tilde{\theta}_{nT,k} - \theta_0\right|}{SE\left(\tilde{\theta}_{nT,k}, \tilde{\alpha}^*_k\right)}$$

be the t-statistic for θ_0 where

$$\left[SE\left(\tilde{\theta}_{nT,k}, \tilde{\alpha}_k\right)\right]^2 = \left(-\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} U_{it\theta}\left(\tilde{\theta}_{nT,k}, \tilde{\alpha}_{i,k}\right)\right)^{-1}, \ \tilde{\alpha}_{i,k} \equiv 2\hat{\alpha}_i - E^*(\check{\alpha}^*_{i,k}).$$

Let

$$t^*\text{-stat} = \frac{\sqrt{nT}\left|\tilde{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right|}{SE\left(\tilde{\theta}^*_{nT,k}, \tilde{\alpha}^*_k\right)}$$

be the corresponding t-statistic in the bootstrap world where

$$\left[SE\left(\tilde{\theta}^*_{nT,k}, \tilde{\alpha}^*_k\right)\right]^2 = \left(-\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} U^*_{it\theta}\left(\tilde{\theta}^*_{nT,k}, \tilde{\alpha}^*_{i,k}\right)\right)^{-1}, \tilde{\alpha}^*_{i,k} \equiv 2\check{\alpha}^*_{i,k} - E^{**}(\check{\alpha}^{**}_{i,k}).$$

Then the bootstrap CI is :

$$\left[\tilde{\theta}_{nT,k} - T^*_{1-\alpha/2}\frac{1}{\sqrt{nT}}SE(\tilde{\theta}_{nT,k}, \tilde{\alpha}_k), \quad \tilde{\theta}_{nT,k} + T^*_{1-\alpha/2}\frac{1}{\sqrt{nT}}SE(\tilde{\theta}_{nT,k}, \tilde{\alpha}_k)\right]$$

where $T^*_{1-\alpha/2}$ is the $(1-\alpha/2)\times 100\%$ percentile of t^*-stat. Our double bootstrap two-sided test is based on the above CI. We can use the same procedure to construct CIs for the average marginal effect and the overall average marginal effect. In our simulation experiment, we set the number of double bootstrap samples to be 100. We do so in order to reduce the computational burden. In empirical applications, we should use a larger number.

Table 1 shows the performance of the k-step bootstrap for different values of k. According to this result, the k-step bootstrap procedure reduces the bias significantly when $k \geq 2$. Results not reported here show that the one-step procedure is not effective in bias reduction. This result is consistent with our Theorem 2, which demonstrates the order of bias is reduced from $O_p(1/T)$ to $O_p(1/T^2)$ when $k \geq 2$. In terms of the MSE, the 2-step bootstrap with observed Hessian, the 3-step bootstrap with observed Hessian and the 3-step bootstrap with expected Hessian are efficient in general.

Table 2 compares different bias correction methods. We choose the 2-step bootstrap with observed Hessian as our benchmark. First, we see that the estimator without bias correction is severely biased when T is small. As T gets larger, the bias gets smaller, but there is still no improvement in the coverage accuracy of CI's. When $T = 4$, the bias of the uncorrected estimator is 42%. When $T = 12$, the bias is reduced to 13% . But the rejection probability is still 29% for the 5% two-sided test. Second, the k-step bootstrap performs better in finite samples than the other methods regardless of the size of T. In particular when $T = 4$, the outperforming of the k-step bootstrap procedure is remarkable, while as T increases other estimators become as accurate as ours. When $T = 4$ the bias of our estimator is 6% and RMSE is 0.249, while the bias of the jackknife method is 25% and its RMSE is 0.373. The analytic method by Fernández-Val (2009) also has the bias of 6% but its RMSE is 0.281 which implies that its variance is larger than ours. The k-step procedure achieves the smallest RMSE among all bias correction procedures. Third, in term of coverage accuracy, the CI's based on the double

k-step bootstrap outperform other CIs in an overall sense.

Table 3 shows the ratio of the estimator of the average marginal effect to the true value. As HN and Fernández-Val (2009) show, the bias of the uncorrected estimator is negligible, even when $T = 4$. Its bias is less than 2% and in terms of RMSE, it performs as good as the bias corrected ones. However its CI's are not accurate especially when we have small T. When $T = 4$, its error in coverage probability for the 95% CI is about 5%. Inaccurate CI's are not just the problem of the bias uncorrected estimator. Jackknife and the analytic bias correction do not reduce the coverage error either. When $T = 4$, the errors in coverage probability for the 95% CI from jackknife and analytic estimators are 11% and 4-6% respectively. In contrast, the coverage error of the 95% CI constructed from the k-step double bootstrap is only 1.5%. We find that the our estimator improves the accuracy of the CI's which is not the case in the other standard alternatives.

Table 4 gives the Monte Carlo results for the ratio of the estimator of the overall average marginal effect to the true value. This is similar to the fixed effect average in Table 3 except that the average is taken over both the fixed effects and the covariate. As in the previous case, we find little evidence that bias correction is necessary in terms of RMSE. Actually, the RMSE of the bias uncorrected estimator is smaller than that of the jackknife estimator in general. It also shows that in contrast to other estimators our double k-step bootstrap procedure improves the coverage accuracy of the CI's particularly when T is small.

3.7 Conclusion

In this paper, we propose the k-step bootstrap bias correction for the fixed effects estimator in nonlinear static panel models and establish the asymptotic properties of our bias corrected estimator. In simulation experiments, we show that the k-step bias correction procedure is often more effective than the alternatives. When T is small, the procedure achieves substantial bias reduction and has the smallest RMSE among the competing procedures. The confidence interval based on the double k-step bootstrap has a smaller coverage error than other CI's. This is true for both model parameters and average marginal effects. The asymptotic properties of our CIs and the possible higher order refinement are not studied here. It is an interesting topic for future research.

3.8 Acknowledgments

Chapter 3, which is coauthored with Yixiao Sun, has been submitted for publication.

Table 3.1: k-step Bootstrap Estimators

Estimator	Mean	Median	SD	RMSE
		$T = 4$		
k=2, O	0.94	0.94	0.242	0.249
k=3, O	0.84	0.84	0.192	0.253
k=3, E	0.83	0.84	0.194	0.256
k=2, E	1.02	1.01	0.270	0.271
		$T = 8$		
k=2, O	0.98	0.98	0.114	0.115
k=3, E	0.97	0.97	0.113	0.117
k=2, E	1.02	1.02	0.122	0.123
k=3, O	0.96	0.96	0.118	0.124
		$T = 12$		
k=3, E	0.99	0.99	0.078	0.078
k=3, O	0.97	0.97	0.077	0.080
k=2, O	0.99	0.99	0.081	0.082
k=2, E	1.01	1.00	0.087	0.087

Notes: Cross section sample size $n = 100$ and the true value $\theta_0 = 1$. We use "E" to indicate the use of the expected Hessian in the k-step bootstrap while we use "O" to indicate the use of the observed Hessian in the k-step bootstrap. The estimators are ordered according to their RMSE.

Table 3.2: Bias Corrected Estimators of θ

Estimator	Mean	Median	SD	p;.05	p;.10	RMSE
			$T = 4$			
Probit	1.42	1.40	0.385	0.30	0.40	0.569
2-step Bootstrap	0.94	0.94	0.242	0.04	0.06	0.249
Jackknife	0.75	0.75	0.277	0.11	0.19	0.373
BC1	1.11	1.10	0.304	0.04	0.11	0.323
BC2	1.20	1.19	0.333	0.09	0.16	0.388
BC3	1.06	1.06	0.275	0.02	0.06	0.281
			$T = 8$			
Probit	1.18	1.18	0.132	0.28	0.39	0.238
2-step Bootstrap	0.98	0.98	0.114	0.04	0.10	0.115
Jackknife	0.95	0.96	0.118	0.05	0.11	0.128
BC1	1.05	1.05	0.134	0.05	0.11	0.143
BC2	1.05	1.05	0.132	0.05	0.10	0.141
BC3	1.02	1.02	0.124	0.03	0.07	0.126
			$T = 12$			
Probit	1.13	1.12	0.090	0.29	0.40	0.161
2-step Bootstrap	0.99	0.99	0.081	0.05	0.11	0.082
Jackknife	0.98	0.98	0.080	0.05	0.10	0.083
BC1	1.04	1.04	0.087	0.07	0.13	0.096
BC2	1.03	1.03	0.085	0.06	0.11	0.090
BC3	1.01	1.01	0.082	0.04	0.09	0.083

Notes: Cross section sample size $n = 100$ and the true value $\theta_0 = 1$. Jackknife denotes HN Jackknife bias corrected estimator; BC1 denotes HN bias corrected estimator based on Bartlett equalities; BC2 denotes HN bias corrected estimator based on general estimating equations; BC3 denotes Fernández-Val (2009) bias corrected estimator which uses expected quantities in the estimation of the bias. p denotes empirical rejection probability.

Table 3.3: Bias Corrected Estimators of the Average Marginal Effect μ

Estimator	Mean	Median	SD	p;.05	p;.10	RMSE
			$T = 4$			
Probit	0.98	0.98	0.256	0.097	0.157	0.256
2-step bootstrap	0.98	0.98	0.256	0.065	0.122	0.256
Jackknife	1.06	1.05	0.307	0.159	0.224	0.313
BC1	1.00	0.99	0.265	0.113	0.178	0.265
BC2	1.05	1.05	0.266	0.117	0.185	0.271
BC3	0.94	0.94	0.240	0.090	0.155	0.247
			$T = 8$			
Probit	1.02	1.01	0.116	0.065	0.122	0.117
2-step bootstrap	1.00	1.00	0.113	0.043	0.096	0.117
Jackknife	1.00	0.99	0.130	0.086	0.153	0.130
BC1	1.02	1.02	0.133	0.090	0.153	0.134
BC2	1.02	1.02	0.131	0.087	0.154	0.133
BC3	1.00	1.00	0.117	0.058	0.107	0.117
			$T = 12$			
Probit	1.02	1.01	0.072	0.072	0.119	0.074
2-step bootstrap	1.00	1.00	0.070	0.052	0.094	0.070
Jackknife	1.00	1.00	0.074	0.05	0.093	0.074
BC1	1.02	1.02	0.075	0.061	0.122	0.078
BC2	1.02	1.02	0.074	0.059	0.112	0.077
BC3	1.01	1.01	0.074	0.049	0.096	0.075

Notes: Cross section sample size $n = 100$ and the true value $\mu_0 = 1$.

Table 3.4: Bias Corrected Estimators of the Overall Average Marginal Effect ν

Estimator	Mean	Median	SD	p;.05	p;.10	RMSE
			$T = 4$			
Probit	0.99	1.00	0.248	0.11	0.17	0.249
2-step Bootstrap	0.99	0.98	0.248	0.04	0.09	0.248
Jackknife	1.02	1.02	0.285	0.12	0.19	0.286
BC1	1.00	1.00	0.261	0.12	0.18	0.261
BC2	1.04	1.04	0.255	0.12	0.19	0.258
BC3	0.94	0.94	0.226	0.08	0.13	0.234
			$T = 8$			
Probit	0.99	0.99	0.100	0.07	0.13	0.100
2-step Bootstrap	0.99	0.99	0.100	0.04	0.08	0.100
Jackknife	1.01	1.01	0.107	0.07	0.14	0.107
BC1	1.01	1.01	0.110	0.09	0.15	0.110
BC2	1.00	1.00	0.105	0.07	0.13	0.105
BC3	0.97	0.97	0.103	0.08	0.13	0.107
			$T = 12$			
Probit	0.99	0.99	0.063	0.09	0.16	0.065
2-step Bootstrap	0.99	0.99	0.064	0.06	0.12	0.065
Jackknife	1.00	1.00	0.064	0.05	0.11	0.064
BC1	1.00	1.00	0.065	0.06	0.11	0.065
BC2	0.99	0.99	0.062	0.05	0.10	0.063
BC3	0.98	0.98	0.062	0.05	0.11	0.065

Notes: Cross section sample size $n = 100$ and the true value $\nu_0 = 1$

3.9 Appendix

I. Proof of Theorem 1

Throughout the proof, we assume that we have truncated $\hat{\theta}^*_{nT}$ so that $\hat{\theta}^*_{nT} - \hat{\theta}_{nT}$ is bounded in absolute value by $M_{nT}/\sqrt{nT}$. The technical details for showing that truncation has a negligible effect on the asymptotic properties of $\sqrt{nT}(\tilde{\theta}_{nT} - \theta_0)$ are the same as those for $\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0)$. We present the details for the latter in Appendix III and omit them here. Truncation allows us to convert probability orders into moment orders.

The bootstrap bias corrected estimator is defined as

$$\tilde{\theta}_{nT} = 2\hat{\theta}_{nT} - E^*\left(\hat{\theta}^*_{nT}\right).$$

Therefore,

$$\sqrt{nT}\left(\tilde{\theta}_{nT} - \theta_0\right) = \sqrt{nT}\left(\hat{\theta}_{nT} - \theta_T\right) + \sqrt{nT}\left(\theta_T - \theta_0 - \left[E^*\left(\hat{\theta}^*_{nT}\right) - \hat{\theta}_{nT}\right]\right).$$

HN has shown that

$$\sqrt{nT}\left(\hat{\theta}_{nT} - \theta_T\right) \xrightarrow{d} N(0, \overline{E}[\mathcal{I}_i]^{-1}),$$

where

$$\mathcal{I}_i \equiv E[U_{it}U'_{it}].$$

Therefore, in order to prove Theorem 1, it suffices to show

$$\sqrt{nT}\left(\theta_T - \theta_0 - \left[E^*\left(\hat{\theta}^*_{nT}\right) - \hat{\theta}_{nT}\right]\right) = o_p(1). \tag{A.24}$$

The LHS of (A.24) can be decomposed into two parts :

$$\begin{aligned}&\sqrt{nT}\left(\theta_T - \theta_0 - \left[E^*\left(\hat{\theta}^*_{nT}\right) - \hat{\theta}_{nT}\right]\right)\\ &= \sqrt{nT}\left(\theta_T - \theta_0 - \left[P^*\lim_{n\to\infty}\hat{\theta}^*_{nT} - \hat{\theta}_{nT}\right]\right) + \sqrt{nT}\left(P^*\lim_{n\to\infty}\hat{\theta}^*_{nT} - E^*\left(\hat{\theta}^*_{nT}\right)\right).\end{aligned}$$

Hence, it suffices to show

$$\theta_T - \theta_0 - \left[P^*\lim_{n\to\infty}\hat{\theta}^*_{nT} - \hat{\theta}_{nT}\right] = O_p\left(\frac{1}{T^2}\right), \tag{A.25}$$

$$P^*\lim_{n\to\infty}\hat{\theta}^*_{nT} - E^*\left(\hat{\theta}^*_{nT}\right) = o_p\left(\frac{1}{\sqrt{nT}}\right). \tag{A.26}$$

1. Prove $\theta_T - \theta_0 - \left[P^* \lim_{n\to\infty} \hat{\theta}^*_{nT} - \hat{\theta}_{nT} \right] = O_p\left(\frac{1}{T^2}\right)$

Let $F \equiv (F_1, \ldots, F_n)$ and $\hat{F} \equiv (\hat{F}_1, \ldots, \hat{F}_n)$, where $F_i := F_{\theta_0, \alpha_{i0}}$ is the distribution function of stratum i and $\hat{F}_i$ is its empirical distribution function. Define $F(\epsilon) \equiv F + \epsilon\sqrt{T}\left(\hat{F} - F\right)$ for $\epsilon \in [0, 1/\sqrt{T}]$. For each fixed θ and ϵ, let $\alpha_i(\theta, F_i(\epsilon))$ and $\theta(F(\epsilon))$ be the solutions to the estimating equations

$$0 = \int V_i\left(\theta, \alpha_i(\theta, F_i(\epsilon))\right) dF_i(\epsilon),$$

$$0 = \sum_{i=1}^{n} \int U_i\left(\theta(F(\epsilon)), \alpha_i(\theta(F(\epsilon)), F_i(\epsilon))\right) dF_i(\epsilon).$$

A Taylor series expansion gives

$$\hat{\theta}_{nT} - \theta_0 = \frac{1}{\sqrt{T}}\theta^{\epsilon}(0) + \frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2 \theta^{\epsilon\epsilon}(0) + \frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3 \theta^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}) \tag{A.27}$$

where $\theta^{\epsilon}(\epsilon) \equiv d\theta(F(\epsilon))/d\epsilon$, $\theta^{\epsilon}(\epsilon) \equiv d^2\theta(F(\epsilon))/d\epsilon^2, \ldots,$ and $\tilde{\epsilon}$ is between 0 and $1/\sqrt{T}$. HN show that

$$\sqrt{nT}\frac{1}{\sqrt{T}}\theta^{\epsilon}(0) \xrightarrow{d} N\left(0, \left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\right),$$

$$\left(\frac{1}{\sqrt{T}}\right)^2 \theta^{\epsilon\epsilon}(0) = \frac{1}{T}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(-\frac{1}{n}\sum_{i=1}^{n}\frac{E\left[V_{2it}U_{it}\right]}{E\left[v_{it}^2\right]}\right) + O_p\left(\frac{1}{T^2}\right),$$

$$\left(\frac{1}{\sqrt{T}}\right)^3 \theta^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}) = o_p\left(\frac{1}{T^2}\right).$$

Therefore,

$$\theta_T - \theta_0 = \frac{B}{T} + O\left(\frac{1}{T^2}\right), \tag{A.28}$$

where

$$B = \lim_{n\to\infty}\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(-\frac{1}{n}\sum_{i=1}^{n}\frac{E\left[V_{2it}U_{it}\right]}{E\left[v_{it}^2\right]}\right).$$

Similarly, in the bootstrap world, for each fixed θ and ϵ, let $\alpha_i(\theta, F_i^*(\epsilon))$ and $\theta(F^*(\epsilon))$ be the solutions to the estimating equations

$$0 = \int V_i^*\left(\theta, \alpha_i(\theta, F_i^*(\epsilon))\right) dF_i^*(\epsilon),$$

$$0 = \sum_{i=1}^{n} \int U_i^*\left(\theta(F^*(\epsilon)), \alpha_i(\theta(F^*(\epsilon)), F_i^*(\epsilon))\right) dF_i^*(\epsilon),$$

where $F_i^*(\epsilon) = F_{\hat{\theta}_{nT},\hat{\alpha}_i(\hat{\theta}_{nT})} + \epsilon\sqrt{T}\left(\hat{F}_i^* - F_{\hat{\theta}_{nT},\hat{\alpha}_i(\hat{\theta}_{nT})}\right)$, $F_{\hat{\theta}_{nT},\hat{\alpha}_i(\hat{\theta}_{nT})}$ is the distribution function of stratum i in the bootstrap sample and $\hat{F}_i^*$ is the corresponding empirical distribution. Note that $F_i^*(0)$ is the same as $F_i = F_{\theta_0,\alpha_{i0}}$ except that the true parameter is $(\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT}))$ rather than (θ_0, α_{i0}).

Similar to equation (A.27), in the bootstrap world, we have

$$\hat{\theta}_{nT}^* - \hat{\theta}_{nT} = \frac{1}{\sqrt{T}}\hat{\theta}^{\epsilon}(0) + \frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2 \hat{\theta}^{\epsilon\epsilon}(0) + \frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3 \hat{\theta}^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}^*)$$

where $\hat{\theta}^{\epsilon}(\epsilon) \equiv d\theta(F^*(\epsilon))/d\epsilon$, $\hat{\theta}^{\epsilon}(\epsilon) \equiv d^2\theta(F^*(\epsilon))/d\epsilon^2, \ldots$, and $\tilde{\epsilon}^*$ is between 0 and $1/\sqrt{T}$. Also,

$$\sqrt{nT}\frac{1}{\sqrt{T}}\hat{\theta}^{\epsilon}(0) \xrightarrow{d} N\left(0, \left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i^*\right)^{-1}\right), \tag{A.29}$$

$$\left(\frac{1}{\sqrt{T}}\right)^2 \hat{\theta}^{\epsilon\epsilon}(0) = \frac{1}{T}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i^*\right)^{-1}\left(-\frac{1}{n}\sum_{i=1}^{n}\frac{E^*\left[V_{2it}^*U_{it}^*\right]}{E^*\left[v_{it}^{*2}\right]}\right) + O_{p^*}\left(\frac{1}{T^2}\right), \tag{A.30}$$

$$\left(\frac{1}{\sqrt{T}}\right)^3 \hat{\theta}^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}^*) = o_{p^*}\left(\frac{1}{T^2}\right). \tag{A.31}$$

Using the same argument as in HN and with some calculations, we have, under Assumption 4(ii):

$$P^*\lim_{n\to\infty}\hat{\theta}_{nT}^* - \hat{\theta}_{nT} = \frac{B^*}{T} + O_p\left(\frac{1}{T^2}\right) \tag{A.32}$$

where

$$B^* = \lim_{n\to\infty}\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i^*\right)^{-1}\left[-\frac{1}{n}\sum_{i=1}^{n}\frac{E^*\left[V_{2it}^*U_{it}^*\right]}{E^*\left[v_{it}^{*2}\right]}\right].$$

We proceed to show that $B^* = B + O_p(1/T)$. First, we have

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i^* - \frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right| \\ &= \left|\frac{1}{n}\sum_{i=1}^{n}\int U_{it}\left(\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})\right)U_{it}\left(\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})\right)' dF_{\hat{\theta}_{nT},\hat{\alpha}_i(\hat{\theta}_{nT})} - \frac{1}{n}\sum_{i=1}^{n}E\left(U_{it}U_{it}'\right)\right| \\ &\le A_{nT} + B_{nT}\end{aligned} \tag{A.33}$$

where

$$A_{nT}=\left|\frac{1}{n}\sum_{i=1}^{n}\int\left(H_{it}\left(\hat{\theta}_{nT},\hat{\alpha}_i(\hat{\theta}_{nT})\right)-H_{it}\left(\theta_0,\alpha_{i0}\right)\right)dF_{\hat{\theta}_{nT},\hat{\alpha}_i(\hat{\theta}_{nT})}\right|$$

$$B_{nT}=\left|\frac{1}{n}\sum_{i=1}^{n}\int H_{it}\left(\theta_0,\alpha_{i0}\right)\left(dF_{\hat{\theta}_{nT},\hat{\alpha}_i(\hat{\theta}_{nT})}-dF_i\right)\right|$$

and

$$H_{it}\left(\theta,\alpha_i\right)=U_{it}\left(\theta,\alpha_i\right)U_{it}\left(\theta,\alpha_i\right)'.$$

To evaluate the stochastic order of A_{nT} and B_{nT}, we use the following result:

$$\hat{\alpha}_i(\hat{\theta}_{nT})-\alpha_{i0}=\frac{\beta_i}{T}+\frac{1}{T}\sum_{t=1}^{T}\psi_{it}+O_p\left(\frac{1}{T^2}\right)$$

uniformly over $i=1,2,...,n$ where

$$\psi_{it}=-\left(E[v_{it\alpha}]\right)^{-1}v_{it},\ \ \beta_i=-\left(E[v_{it\alpha}]\right)^{-1}\left\{E[v_{it\alpha}\psi_{it}]+\frac{1}{2}E[v_{it\alpha\alpha}]E[\psi_{it}^2]\right\}.$$

This result is given in HN and follows from the standard higher order expansion.

For A_{nT}, suppose that $\bar{\theta}_1$ is a parameter value between θ_0 and $\hat{\theta}_{nT}$ and $\bar{\alpha}_{i1}$ is a value between α_{i0} and $\hat{\alpha}_i$. Then

$$\begin{aligned}
A_{nT}&\leq\left|\hat{\theta}_{nT}-\theta_0\right|\left|\frac{1}{n}\sum_{i=1}^{n}\int\frac{\partial H_{it}\left(\bar{\theta}_1,\bar{\alpha}_{i1}\right)}{\partial\theta}dF_{\hat{\theta}_{nT},\hat{\alpha}_i(\hat{\theta}_{nT})}\right|\\
&+\left|\frac{1}{n}\sum_{i=1}^{n}(\hat{\alpha}_i(\hat{\theta}_{nT})-\alpha_{i0})\int\frac{\partial H_{it}\left(\bar{\theta}_1,\bar{\alpha}_{i1}\right)}{\partial\alpha_i}dF_{\hat{\theta}_{nT},\hat{\alpha}_i(\hat{\theta}_{nT})}\right|\\
&\leq\left|\hat{\theta}_{nT}-\theta_0\right|\left|\frac{1}{n}\sum_{i=1}^{n}\int M^2(z_{it})dF_{\theta,\alpha_i}\right|_{\theta=\hat{\theta}_{nT},\alpha=\alpha(\hat{\theta}_{nT})}\\
&+\left|\frac{1}{n}\sum_{i=1}^{n}(\hat{\alpha}_i(\hat{\theta}_{nT})-\alpha_{i0})\left[\int\frac{\partial H_{it}\left(\theta,\alpha\right)}{\partial\alpha_i}dF_{\theta,\alpha_i}\right]_{\theta=\hat{\theta}_{nT},\alpha=\alpha(\hat{\theta}_{nT})}\right|(1+o\left(1\right))\\
&=O_p\left(\frac{1}{T}\right)+\left|\frac{1}{n}\sum_{i=1}^{n}(\hat{\alpha}_i(\hat{\theta}_{nT})-\alpha_{i0})\left[\int\frac{\partial H_{it}\left(\theta,\alpha\right)}{\partial\alpha_i}dF_{\theta,\alpha_i}\right]_{\theta=\hat{\theta}_{nT},\alpha=\alpha(\hat{\theta}_{nT})}\right|\\
&\times\left(1+o\left(1\right)\right)
\end{aligned}\qquad\text{(A.34)}$$

using Assumptions 1 and 4. Next

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}(\hat{\alpha}_i(\hat{\theta}_{nT})-\alpha_{i0})\left[\int\frac{\partial H_{it}\left(\theta,\alpha\right)}{\partial\alpha_i}dF_{\theta,\alpha_i}\right]_{\theta=\hat{\theta}_{nT},\alpha=\alpha(\hat{\theta}_{nT})}\right| \\
&\leq\left|\frac{1}{nT}\sum_{i=1}^{n}\beta_i E\left[\frac{\partial H_{it}\left(\theta_0,\alpha_{i0}\right)}{\partial\alpha_i}\right]\left(1+o\left(1\right)\right)\right| \\
&+\left|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\psi_{it}E\left[\frac{\partial H_{it}\left(\theta_0,\alpha_{i0}\right)}{\partial\alpha_i}\right]\left(1+o\left(1\right)\right)\right|+O_p\left(\frac{1}{T^2}\right) \\
&=O_p\left(\frac{1}{T}\right)+O_p\left(\frac{1}{\sqrt{nT}}\right)+O_p\left(\frac{1}{T^2}\right)=O_p\left(\frac{1}{T}\right)
\end{aligned} \tag{A.35}$$

using LLN and CLT. We have thus proved

$$A_{nT}=O_p\left(\frac{1}{T}\right). \tag{A.36}$$

For B_{nT}, suppose that $\bar{\theta}_2$ is between θ_0 and $\hat{\theta}_{nT}$ and that $\bar{\alpha}_{i2}$ is between α_{i0} and $\hat{\alpha}_i$. Then

$$\begin{aligned}
&B_{nT} \\
&=\left|\frac{1}{n}\sum_{i=1}^{n}\int H_{it}\left(\theta_0,\alpha_{i0}\right)\left(f(z;\hat{\theta}_{nT},\hat{\alpha}_i(\hat{\theta}_{nT}))-f(z;\theta_0,\alpha_{i0})\right)dz\right| \\
&:=I_1+I_2
\end{aligned} \tag{A.37}$$

where

$$\begin{aligned}
I_1&=\left|\frac{1}{n}\sum_{i=1}^{n}\left[(\hat{\theta}_{nT}-\theta_0)\int H_{it}\left(\theta_0,\alpha_{i0}\right)u_{it}(\bar{\theta}_2,\bar{\alpha}_{i2})dF_{\bar{\theta}_2,\bar{\alpha}_{i2}}\right]\right| \\
&=\left|(\hat{\theta}_{nT}-\theta_0)\right|\left|\frac{1}{n}\sum_{i=1}^{n}\left(\int H_{it}\left(\theta_0,\alpha_{i0}\right)u_{it}dF_i+o_p(1)\right)\right| \\
&\leq\sup_i E\left[M(z_{it})^3\right]\left|(\hat{\theta}_{nT}-\theta_0)\right|=O_p\left(\frac{1}{T}\right),
\end{aligned} \tag{A.38}$$

and

$$\begin{aligned}
I_2&=\left|\frac{1}{n}\sum_{i=1}^{n}\left[(\hat{\alpha}_i(\hat{\theta}_{nT})-\alpha_{i0})\int H_{it}\left(\theta_0,\alpha_{i0}\right)v_{it}(\bar{\theta}_2,\bar{\alpha}_{i2})dF_{\bar{\theta}_2,\bar{\alpha}_{i2}}\right]\right| \\
&=\left|\frac{1}{n}\sum_{i=1}^{n}(\hat{\alpha}_i(\hat{\theta}_{nT})-\alpha_{i0})\left[\int H_{it}\left(\theta_0,\alpha_{i0}\right)v_{it}(\theta_0,\alpha_{i0})dF_i+o_p\left(1\right)\right]\right|=O_p\left(\frac{1}{T}\right)
\end{aligned} \tag{A.39}$$

using the argument similar to (A.35). Therefore,

$$B_{nT} = O_p\left(\frac{1}{T}\right). \tag{A.40}$$

Combining (A.36) and (A.40) yields:

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i^* = \frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i + O_p\left(\frac{1}{T}\right). \tag{A.41}$$

Using the same procedure, we can show that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{E^*\left[V_{2it}^* U_{it}^*\right]}{E^*\left[v_{it}^{*2}\right]} = \frac{1}{n}\sum_{i=1}^{n}\frac{E\left[V_{2it} U_{it}\right]}{E\left[v_{it}^{2}\right]} + O_p\left(\frac{1}{T}\right). \tag{A.42}$$

Therefore, from (A.41) and (A.42),

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i^*\right)^{-1}\left(-\frac{1}{n}\sum_{i=1}^{n}\frac{E^*\left[V_{2it}^* U_{it}^*\right]}{E^*\left[v_{it}^{*2}\right]}\right)\\
&= \lim_{n\to\infty}\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(-\frac{1}{n}\sum_{i=1}^{n}\frac{E\left[V_{2it} U_{it}\right]}{E\left[v_{it}^{2}\right]}\right) + O_p\left(\frac{1}{T}\right).
\end{aligned} \tag{A.43}$$

That is

$$B^* = B + O_p\left(\frac{1}{T}\right)$$

completing the proof of (A.25).

2. Prove $\sqrt{nT}\left(P^*\lim_{n\to\infty}\hat{\theta}_{nT}^* - E^*\left(\hat{\theta}_{nT}^*\right)\right) \xrightarrow{p} 0$

We write

$$\begin{aligned}
&\sqrt{nT}\left(P^*\lim_{n\to\infty}\hat{\theta}_{nT}^* - E^*\left(\hat{\theta}_{nT}^*\right)\right)\\
&= \sqrt{nT}\left(E^*\left(P^*\lim_{n\to\infty}\hat{\theta}_{nT}^*\right) - E^*\left(\hat{\theta}_{nT}^*\right)\right) = \sqrt{nT}\left(\lim_{n\to\infty}E^*\left(\hat{\theta}_{nT}^*\right) - E^*\left(\hat{\theta}_{nT}^*\right)\right)\\
&= \sqrt{nT}\left(\lim_{n\to\infty}E^*\left(\frac{1}{\sqrt{T}}\hat{\theta}^{\epsilon}(0)\right) - E^*\left(\frac{1}{\sqrt{T}}\hat{\theta}^{\epsilon}(0)\right)\right)\\
&+ \frac{\sqrt{nT}}{2}\left(\lim_{n\to\infty}E^*\left(\frac{1}{T}\hat{\theta}^{\epsilon\epsilon}(0)\right) - E^*\left(\frac{1}{T}\hat{\theta}^{\epsilon\epsilon}(0)\right)\right) + o_p(1).
\end{aligned}$$

The second equality holds by the dominated convergence theorem and the last equality follows from an argument similar to HN. Therefore, it suffices to show that

$$\sqrt{nT}\left(\frac{1}{\sqrt{T}}\lim_{n\to\infty}E^*\left(\hat{\theta}^{\epsilon}(0)\right)-\frac{1}{\sqrt{T}}E^*\left(\hat{\theta}^{\epsilon}(0)\right)\right)=0, \tag{A.44}$$

$$\sqrt{nT}\left(\frac{1}{T}\lim_{n\to\infty}E^*\left(\hat{\theta}^{\epsilon\epsilon}(0)\right)-\frac{1}{T}E^*\left(\hat{\theta}^{\epsilon\epsilon}(0)\right)\right)=o_p(1). \tag{A.45}$$

Equation (A.44) holds because

$$\begin{aligned}
&\sqrt{nT}\left(\frac{1}{\sqrt{T}}\lim_{n\to\infty}E^*\left(\hat{\theta}^{\epsilon}(0)\right)-\frac{1}{\sqrt{T}}E^*\left(\hat{\theta}^{\epsilon}(0)\right)\right)\\
&=\sqrt{n}\left(\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i^*\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}E^*\left(U_i^*\right)-\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i^*\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}E^*\left(U_i^*\right)\right)\\
&=0,
\end{aligned}$$

using $E^*\left(U_i^*\right)=0$.

Equation (A.45) holds because

$$\begin{aligned}
&\sqrt{nT}\left(\frac{1}{T}\lim_{n\to\infty}E^*\left(\hat{\theta}^{\epsilon\epsilon}(0)\right)-\frac{1}{T}E^*\left(\hat{\theta}^{\epsilon\epsilon}(0)\right)\right)\\
&=\sqrt{\frac{n}{T}}\left[\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i^*\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{E^*(V_{2it}^*U_{it}^*)}{E^*(v_{it}^*)}\right)\right.\\
&\left.-\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i^*\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{E^*(V_{2it}^*U_{ii}^*)}{E^*(v_{it}^{*2})}\right)\right]+O_p\left(\sqrt{\frac{n}{T^3}}\right)
\end{aligned} \tag{A.46}$$

where the $O_p(\cdot)$ terms come from $O_{p^*}(\cdot)$ in (A.30) and the leading term in $(A.46)$

is

$$\sqrt{\frac{n}{T}}\left[\lim_{n\to\infty}\left(\sum_{i=1}^{n}E_{\theta,\alpha_i}U_{it}\left(\theta,\alpha_i\right)U_{it}'\left(\theta,\alpha_i\right)\right)^{-1}\left(\sum_{i=1}^{n}\frac{E_{\theta,\alpha_i}(V_{2it}\left(\theta,\alpha_i\right)U_{it}\left(\theta,\alpha_i\right))}{E_{\theta,\alpha_i}(v_{it}^2\left(\theta,\alpha_i\right))}\right)\right.$$
$$-\left(\sum_{i=1}^{n}E_{\theta,\alpha_i}U_{it}\left(\theta,\alpha_i\right)U_{it}'\left(\theta,\alpha_i\right)\right)^{-1}$$
$$\left.\left.\times\left(\sum_{i=1}^{n}\frac{E_{\theta,\alpha_i}(V_{2it}\left(\theta,\alpha_i\right)U_{it}\left(\theta,\alpha_i\right))}{E_{\theta,\alpha_i}(v_{it}^2\left(\theta,\alpha_i\right))}\right)\right]\right|_{(\theta,\alpha)=\left(\hat{\theta}_{nT},\hat{\alpha}\left(\hat{\theta}_{nT}\right)\right)}$$
$$=\sqrt{\frac{n}{T}}\left(O_p\left(\frac{1}{\sqrt{n}}\right)\right)=o_p(1)$$

using Assumption 6. In the above equation, we use the subscript θ,α_i on E_{θ,α_i} to emphasize that it is the expectation under F_{θ,α_i}. This completes the proof of Theorem 1. ■

II. Proof of Proposition 2

It is easy to show that

$$P^*\lim_{n\to\infty}\left(\hat{\theta}_{nT}^*-\hat{\theta}_{nT}\right)^2=O_p\left(\frac{1}{nT}\right)=O_p\left(\frac{1}{T^2}\right), \tag{A.47}$$

and

$$P^*\lim_{n\to\infty}\left(\hat{\alpha}_i^*-\hat{\alpha}_i\right)^2=O_p\left(\frac{1}{T}\right). \tag{A.48}$$

By the definition of the k-step bootstrap estimator:

$$\begin{pmatrix}\hat{\theta}_{nT,k}^*\\ \hat{\alpha}_k^*\end{pmatrix}=\begin{pmatrix}\hat{\theta}_{nT,k-1}^*\\ \hat{\alpha}_{k-1}^*\end{pmatrix}-H_{k-1}^{-1}S_{k-1}.$$

For notational compactness, let $\beta=(\theta',\alpha')'$, $\hat{\beta}^*=(\hat{\theta}_{nT}^{*\prime},\hat{\alpha}^{*\prime})'$, $\hat{\beta}=(\hat{\theta}_{nT}',\hat{\alpha}')'$. Then

$$\hat{\beta}_k^*=\hat{\beta}_{k-1}^*-\left[H(\hat{\beta}_{k-1}^*;z_{it}^*)\right]^{-1}S\left(\hat{\beta}_{k-1}^*;z_{it}^*\right)$$

for $\hat{\beta}_0^*=\hat{\beta}$. Using a Taylor expansion and the first order condition:

$$S\left(\hat{\beta}^*;z_{it}\right)=0,$$

we have

$$\begin{aligned}
&\hat{\beta}_k^* - \hat{\beta}^* \\
&= \hat{\beta}_{k-1}^* - \left[H(\hat{\beta}_{k-1}^*; z_{it}^*)\right]^{-1} S\left(\hat{\beta}_{k-1}^*; z_{it}^*\right) - \hat{\beta}^* \\
&= \left[H(\hat{\beta}_{k-1}^*; z_{it}^*)\right]^{-1} \left[S\left(\hat{\beta}^*; z_{it}\right) - S\left(\hat{\beta}_{k-1}^*; z_{it}^*\right) - H(\hat{\beta}_{k-1}^*; z_{it}^*)\left(\hat{\beta}^* - \hat{\beta}_{k-1}^*\right)\right] \\
&= \frac{1}{2}\left[H(\hat{\beta}_{k-1}^*; z_{it}^*)\right]^{-1} \xi
\end{aligned}$$

where $\hat{\beta}_{k-1}^\dagger$ lies between $\hat{\beta}^*$ and $\hat{\beta}_{k-1}^*$, $\xi = \left(\xi_1, ..., \xi_u, ..., \xi_{L_\beta}\right)$ is a vector with u-th element

$$\xi_u = \left(\hat{\beta}^* - \hat{\beta}_{k-1}^*\right)' H_{\beta_u}(\hat{\beta}_{k-1}^\dagger; z_{it}^*)\left(\hat{\beta}^* - \hat{\beta}_{k-1}^*\right)$$

and

$$\begin{aligned}
H_{\beta_u}(\beta; z_{it}^*) &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial}{\partial\beta_u}\frac{\partial^2 \log(\beta; z_{it}^*)}{\partial\beta\partial\beta'} \\
&:= \begin{pmatrix} H_{\theta\theta\beta_u}(\beta; z_{it}^*) & H_{\theta\alpha\beta_u}(\beta; z_{it}^*) \\ H_{\alpha\theta\beta_u}(\beta; z_{it}^*) & H_{\alpha\alpha\beta_u}(\beta; z_{it}^*) \end{pmatrix}.
\end{aligned}$$

A more explicit expression for ξ_u is

$$\begin{aligned}
\xi_u &= \left(\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^*\right)' H_{\theta\theta\beta_u}\left(\hat{\beta}_{k-1}^\dagger; z_{it}^*\right)\left(\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^*\right) \\
&\quad + 2\left(\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^*\right)' H_{\theta\alpha\beta_u}\left(\hat{\beta}_{k-1}^\dagger; z_{it}^*\right)\left(\hat{\alpha}_{k-1}^* - \hat{\alpha}^*\right) \\
&\quad + \left(\hat{\alpha}_{k-1}^* - \hat{\alpha}^*\right)' H_{\alpha\alpha\beta_u}\left(\hat{\beta}_{k-1}^\dagger; z_{it}^*\right)\left(\left(\hat{\alpha}_{k-1}^* - \hat{\alpha}^*\right)\right).
\end{aligned}$$

Hence

$$\begin{aligned}
\|\xi_u\| &\leq \left\|2\left(\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^*\right)' H_{\theta\theta\beta_u}\left(\hat{\beta}_{k-1}^\dagger; z_{it}^*\right)\left(\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^*\right)\right\| \\
&\quad + \left\|2\left(\hat{\alpha}_{k-1}^* - \hat{\alpha}^*\right)' H_{\alpha\alpha\beta_u}\left(\hat{\beta}_{k-1}^\dagger; z_{it}^*\right)\left(\hat{\alpha}_{k-1}^* - \hat{\alpha}^*\right)\right\| \\
&\leq 2\left\|H_{\theta\theta\beta_u}\left(\hat{\beta}_{k-1}^\dagger; z_{it}^*\right)\right\|\left\|\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^*\right\|^2 \\
&\quad + \left\|\frac{2}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left|H_{\alpha_i\alpha_i\beta_u}\left(\hat{\beta}_{k-1}^\dagger; z_{it}^*\right)\right|\left(\hat{\alpha}_{i,k-1}^* - \hat{\alpha}_i^*\right)^2\right\|
\end{aligned}$$

where $\|\cdot\|$ is the Euclidean norm, that is, for a symmetric matrix A, $\|A\|^2 = trace(AA')$. So

$$\left\|\hat{\beta}_k^* - \hat{\beta}^*\right\| \leq \zeta_{\theta,k-1}^* \left\|\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^*\right\|^2 + \zeta_{\alpha,k-1}^*, \tag{A.49}$$

where

$$\zeta_{\theta,k-1}^* = \left\|\left(H(\beta_{k-1}^*; z_{it}^*\right)^{-1}\right\| \sum_{u=1}^{L_\beta} \left\|H_{\theta\theta\beta_u}(\hat{\beta}_{k-1}^\dagger; z_{it}^*)\right\|, \tag{A.50}$$

$$\zeta_{\alpha,k-1}^* = \left\|\left(H(\beta_{k-1}^*; z_{it}^*\right)^{-1}\right\| \sum_{u=1}^{L_\beta} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left|H_{\alpha_i\alpha_i\beta_u}\left(\hat{\beta}_{k-1}^\dagger; z_{it}^*\right)\right| \left(\hat{\alpha}_{i,k-1}^* - \hat{\alpha}_i^*\right)^2. \tag{A.51}$$

For $k = 1$, we have

$$\begin{aligned} &P\left(TP^*\lim_{n\to\infty}\left\|\hat{\beta}_k^* - \hat{\beta}^*\right\| > 2C_1\right) \\ &\leq P\left(TP^*\lim_{n\to\infty}\zeta_{\theta,k-1}^* \left\|\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^*\right\|^2 > C_1\right) + P\left(TP^*\lim_{n\to\infty}\zeta_{\alpha,k-1}^* > C_1\right) \end{aligned} \tag{A.52}$$

The first probability in (A.52) satisfies

$$\begin{aligned} &P\left(TP^*\lim_{n\to\infty}\zeta_{\theta,k-1}^* \left\|\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^*\right\|^2 > C_1\right) \\ &\leq P\left(TP^*\lim_{n\to\infty}\zeta_{\theta,k-1}^* \left\|\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^*\right\|^2 > C_1, P^*\lim_{n\to\infty}\zeta_{\theta,k-1}^* < C_2\right) \\ &+ P\left(P^*\lim_{n\to\infty}\zeta_{\theta,k-1}^* \geq C_2\right) \\ &\leq P\left(TP^*\lim_{n\to\infty}\left\|\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^*\right\|^2 > C_1/\left(C_2\right)\right) + P\left(P^*\lim_{n\to\infty}\zeta_{\theta,k-1}^* \geq C_2\right) \\ &= P\left(P^*\lim_{n\to\infty}\zeta_{\theta,k-1}^* \geq C_2\right) + o(1) \end{aligned} \tag{A.53}$$

using $(A.47)$. We proceed to bound $P\left(P^*\lim_{n\to\infty}\zeta^*_{\theta,k-1} \geq C_2\right)$. By definition,

$$
\begin{aligned}
&P\left(P^*\lim_{n\to\infty}\zeta^*_{nT} \geq C_2\right)\\
&= P\left(\left\|P^*\lim_{n\to\infty}\left(H(\hat{\beta}^*_{k-1}; z^*_{it}\right)^{-1}\right\| P^*\lim_{n\to\infty}\sum_{u=1}^{L_\beta}\left\|H_{\theta\theta\beta_u}(\hat{\beta}^\dagger_{k-1}; z^*_{it})\right\| \geq C_2\right)\\
&\leq P\left(P^*\lim_{n\to\infty}\sum_{u=1}^{L_\beta}\left\|n^{-1}H_{\theta\theta\beta_u}(\hat{\beta}^\dagger_{k-1}; z^*_{it})\right\| \geq \sqrt{C_2}\right)\\
&+ P\left(\left\|P^*\lim_{n\to\infty}\left(n^{-1}H(\hat{\beta}_{k-1}; z^*_{it}\right)^{-1}\right\| \geq \sqrt{C_2}\right)\\
&:= A + B, \qquad \text{(A.54)}
\end{aligned}
$$

where

$$A = P\left(P^*\lim_{n\to\infty}\sum_{u=1}^{L_\beta}\left\|n^{-1}H_{\theta\theta\beta_u}(\hat{\beta}^\dagger_{k-1}; z^*_{it})\right\| \geq \sqrt{C_2}\right), \qquad \text{(A.55)}$$

and

$$B = P\left(\left\|P^*\lim_{n\to\infty}\left(n^{-1}H(\hat{\beta}_{k-1}; z^*_{it}\right)^{-1}\right\| \geq \sqrt{C_2}\right).$$

Note that $\hat{\beta}^\dagger_{k-1}$ is between $\hat{\beta}$ and $\hat{\beta}^*$ and

$$\hat{\beta}^* = \hat{\beta} + o_{p^*}(1). \qquad \text{(A.56)}$$

Using a uniform law of large numbers under the probability measure P^* and the dominated convergence theorem, we have

$$
\begin{aligned}
&P^*\lim_{n\to\infty}\frac{1}{n}\sum_{u=1}^{L_\beta}\left\|H_{\theta\theta\beta_u}(\hat{\beta}^\dagger_{k-1}; z^*_{it})\right\|\\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{u=1}^{L_\beta}P^*\lim_{n\to\infty}\left\|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial}{\partial\beta_u}\frac{\partial^2\log(\hat{\beta}^\dagger_{k-1}; z^*_{it})}{\partial\theta\partial\theta'}\right\|\\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{u=1}^{L_\beta}\left\|\lim_{n\to\infty}E^*\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial}{\partial\beta_u}\frac{\partial^2\log(\hat{\beta}; z^*_{it})}{\partial\theta\partial\theta'}\right\|\\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{u=1}^{L_\beta}\left\|\overline{E}_\beta\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\beta_u}\frac{\partial^2\log(\beta; z_{it})}{\partial\theta\partial\theta'}\right]\right\|_{\beta=\hat{\beta}} \qquad \text{(A.57)}
\end{aligned}
$$

where the last equality follows because P^* conditional on the data (i.e. $\hat{\theta}_{nT}, \hat{\alpha}_i$) is the same as P but with different model parameters. Hence,

$$A = P\left(\frac{1}{n}\sum_{u=1}^{L_\beta}\left\|\overline{E}_\beta\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\beta_u}\frac{\partial^2\log(\beta;z_{it}^*)}{\partial\theta\partial\theta'}\right]\right\|_{\beta=\hat{\beta}} \geq \sqrt{C_2}\right)$$

which can be made arbitrarily small if we choose a large C_2.

Using the same argument, we can show that, when C_2 is large enough, $B = o(1)$ as n and T go to ∞. We have therefore proved

$$P\left(P^*\lim_{n\to\infty}\zeta_{nT}^* \geq C_2\right) = o(1). \tag{A.58}$$

when C_2 is large enough.

To show that the second probability in (A.52) is $o\,(1)$, it suffices to prove

$$P\left(\frac{T}{n}\sum_{u=1}^{L_\beta}P^*\lim_{n\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left|H_{\alpha_i\alpha_i\beta_u}\left(\hat{\beta}_{k-1}^{\dagger};z_{it}^*\right)\right|\left(\hat{\alpha}_{i,k-1}^*-\hat{\alpha}_i^*\right)^2 > \sqrt{C_2}\right)$$
$$= o\,(1)\,.$$

But

$$\frac{1}{n}\sum_{u=1}^{L_\beta}P^*\lim_{n\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left|H_{\alpha_i\alpha_i\beta_u}\left(\hat{\beta}_{k-1}^{\dagger};z_{it}^*\right)\right|\left(\hat{\alpha}_{i,k-1}^*-\hat{\alpha}_i^*\right)^2$$
$$= \frac{1}{n}\sum_{u=1}^{L_\beta}P^*\lim_{n\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left|H_{\alpha_i\alpha_i\beta_u}\left(\hat{\beta}_{k-1}^{\dagger};z_{it}^*\right)\right|\left(\frac{\beta_i^*}{T}+\frac{1}{T}\sum_{t=1}^{T}\psi_{it}^*\right)^2$$
$$\times\left(1+o_p\,(1)\right)$$
$$= \frac{1}{n}\sum_{u=1}^{L_\beta}P^*\lim_{n\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left|H_{\alpha_i\alpha_i\beta_u}\left(\hat{\beta}_{k-1}^{\dagger};z_{it}^*\right)\right|\left[\frac{\left(\beta_i^*\right)^2}{T^2}+\frac{1}{T}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_{it}^*\right)^2\right]$$
$$\times\left(1+o_p\,(1)\right)$$
$$= \frac{1}{n}\sum_{u=1}^{L_\beta}P^*\lim_{n\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left|H_{\alpha_i\alpha_i\beta_u}\left(\hat{\beta}_{k-1}^{\dagger};z_{it}^*\right)\right|\left[\frac{\left(\beta_i^*\right)^2}{T^2}+\frac{1}{T^2}\sum_{t=1}^{T}\left(\psi_{it}^*\right)^2\right]$$
$$\times\left(1+o_p\,(1)\right)$$
$$= O_p\left(\frac{1}{T}\right)$$

where the last equality follows from the ULLN.

Combining (A.53) with (A.58), we have, for $k = 1$.

$$P\left(TP^*\lim_{n\to\infty}\left|\hat{\beta}_k^* - \hat{\beta}^*\right| > C\right) = o(1)$$

when C is large enough. That is, when $k = 1$

$$P^*\lim_{n\to\infty}\hat{\beta}_k^* = P^*\lim_{n\to\infty}\hat{\beta}^* + O_p\left(\frac{1}{T}\right). \tag{A.59}$$

For $k \geq 2$, we note that

$$\left\|\hat{\beta}_k^* - \hat{\beta}^*\right\| \leq \eta_{nT}^*\left\|\hat{\beta}_{k-1}^* - \hat{\beta}^*\right\|^2$$

where

$$\eta_{nT}^* = \frac{1}{2}\max_k\left\|\left[H(\hat{\beta}_{k-1}^*; z_{it}^*)\right]^{-1} H_{\beta_u}(\hat{\beta}_{k-1}^{\dagger}; z_{it}^*)\right\|$$

Using the recursive relationship repeatedly, we have, for $k \geq 2$,

$$\left\|\hat{\beta}_k^* - \hat{\beta}^*\right\| \leq (\eta_{nT}^*)^{\phi}\left\|\hat{\beta}_1^* - \hat{\beta}^*\right\|^{2^{k-1}} \tag{A.60}$$

where $\phi = \sum_{j=2}^{k} 2^{j-1}$.

Using a similar argument, we can show that $P^*\lim_{n\to\infty}\eta_{nT}^* = O_p(1)$. Combining this with (A.59) and (A.60), we have

$$P^*\lim_{n\to\infty}\left(\hat{\beta}_k^* - \hat{\beta}^*\right) = O_p\left(\left(\frac{1}{T}\right)^{2^{k-1}}\right),$$

which implies that

$$P^*\lim_{n\to\infty}\left(\hat{\theta}_{nT,k}^* - \hat{\theta}_{nT}^*\right) = O_p\left(\left(\frac{1}{T}\right)^{2^{k-1}}\right)$$

as desired. ■

III. Proof of Theorem 3

Define our truncated k-step bootstrap bias corrected estimator to be:

$$\check{\theta}_{nT,k} \equiv 2\hat{\theta}_{nT} - E^*\left(\check{\theta}_{nT,k}^*\right).$$

Then

$$
\begin{aligned}
&\sqrt{nT}\left(\check{\theta}_{nT,k}-\theta_0\right)\\
&=\sqrt{nT}\left(\hat{\theta}_{nT}-\theta_T\right)+\sqrt{nT}\left(\theta_T-\theta_0-\left[E^*\left(\check{\theta}^*_{nT,k}\right)-\hat{\theta}_{nT}\right]\right)\\
&=\sqrt{nT}\left(\hat{\theta}_{nT}-\theta_T\right)+\sqrt{nT}\left(\theta_T-\theta_0-\left[P^*\lim_{n\to\infty}\hat{\theta}^*_{nT}-\hat{\theta}_{nT}\right]\right)\\
&+\sqrt{nT}\left(P^*\lim_{n\to\infty}\hat{\theta}^*_{nT}-P^*\lim_{n\to\infty}\hat{\theta}^*_{nT,k}\right)+\sqrt{nT}\left(P^*\lim_{n\to\infty}\hat{\theta}^*_{nT,k}-P^*\lim_{n\to\infty}\check{\theta}^*_{nT,k}\right)\\
&+\sqrt{nT}\left(P^*\lim_{n\to\infty}\check{\theta}^*_{nT,k}-E^*\left(\check{\theta}^*_{nT,k}\right)\right)
\end{aligned}
$$

As HN has shown that

$$\sqrt{nT}\left(\hat{\theta}_{nT}-\theta_T\right)\xrightarrow{d}N(0,\overline{E}[\mathcal{I}_i]^{-1})$$

and we have shown that

$$\sqrt{nT}\left(\theta_T-\theta_0-\left[P^*\lim_{n\to\infty}\hat{\theta}^*_{nT}-\hat{\theta}_{nT}\right]\right)=O_p\left(\frac{1}{T^2}\right)$$

and

$$P^*\lim_{n\to\infty}\hat{\theta}^*_{nT}-P^*\lim_{n\to\infty}\hat{\theta}^*_{nT,k}=O_p\left(\frac{1}{T^2}\right)$$

Therefore, it suffices to show

$$P^*\lim_{n\to\infty}\hat{\theta}^*_{nT,k}-P^*\lim_{n\to\infty}\check{\theta}^*_{nT,k}=O_p\left(\frac{1}{T^2}\right)\tag{A.61}$$

$$P^*\lim_{n\to\infty}\check{\theta}^*_{nT,k}-E^*\left(\check{\theta}^*_{nT,k}\right)=O_p\left(\frac{1}{T^2}\right)\tag{A.62}$$

1. Prove $P^*\lim_{n\to\infty}\hat{\theta}^*_{nT,k}-P^*\lim_{n\to\infty}\check{\theta}^*_{nT,k}=O_p\left(\frac{1}{T^2}\right)$

By definition (c.f. (A.17)) we have:

$$P^*\lim_{n\to\infty}\hat{\theta}^*_{nT,k}-P^*\lim_{n\to\infty}\check{\theta}^*_{nT,k}=P^*\lim_{n\to\infty}\left(\hat{\theta}_{nT}-\hat{\theta}^*_{nT,k}\right)1\left(\sqrt{nT}\left|\hat{\theta}^*_{nT,k}-\hat{\theta}_{nT}\right|>M_{nT}\right).$$

As

$$
\begin{aligned}
&\left|P^*\lim_{n\to\infty}\left(\hat{\theta}_{nT}-\hat{\theta}^*_{nT,k}\right)\right|\\
&\leq\left|P^*\lim_{n\to\infty}\left(\hat{\theta}_{nT}-\hat{\theta}^*_{nT}\right)\right|+\left|P^*\lim_{n\to\infty}\left(\hat{\theta}^*_{nT}-\hat{\theta}^*_{nT,k}\right)\right|=O_p\left(\frac{1}{T}\right),
\end{aligned}
$$

it suffices to show

$$P^* \lim_{n\to\infty} 1\left(\sqrt{nT}\left|\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right| > M_{nT}\right) = O_p\left(\frac{1}{T}\right).$$

For any given $\delta > 0$, we have

$$\begin{aligned}
&P\left(T \cdot P^* \lim_{n\to\infty} 1\left(\sqrt{nT}\left|\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right| > M_{nT}\right) > \delta\right) \\
&= P\left(P^* \lim_{n\to\infty} 1\left(\sqrt{nT}\left|\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right| > M_{nT}\right) = 1\right) \\
&= P\left(1\left(P^* \lim_{n\to\infty}\sqrt{nT}\left|\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right| > M_{nT}\right) = 1\right) \\
&= P\left(\left(P^* \lim_{n\to\infty}\sqrt{nT}\left|\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right|\right) > M_{nT}\right) \\
&\leq P\left(\left(P^* \lim_{n\to\infty}\sqrt{nT}\left|\hat{\theta}^*_{nT,k} - \hat{\theta}^*_{nT}\right| + P^* \lim_{n\to\infty}\sqrt{nT}\left|\hat{\theta}_{nT} - \hat{\theta}^*_{nT}\right|\right) > M_{nT}\right) \\
&= P\left(\sqrt{nT}\left[O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{T^2}\right)\right] > M_{nT}\right) \\
&= o\left(1\right)
\end{aligned}$$

using the condition that $\sqrt{\frac{n}{T}} = o\left(M_{nT}\right)$ and $M_{nT} \to \infty$. Here we have used $P^* \lim_{n\to\infty}\left|\hat{\theta}^*_{nT,k} - \hat{\theta}^*_{nT}\right| = O_p\left(\frac{1}{T^2}\right)$ and $P^* \lim_{n\to\infty}\sqrt{nT}\left|\hat{\theta}_{nT} - \hat{\theta}^*_{nT}\right|/M_{nT} = o_p\left(1\right)$.

Combining the above results yields

$$P^* \lim_{n\to\infty}\hat{\theta}^*_{nT,k} - P^* \lim_{n\to\infty}\check{\theta}^*_{nT,k} = O_p\left(\frac{1}{T^2}\right)$$

as desired.

2. Prove $\sqrt{nT}\left(P^* \lim_{n\to\infty}\check{\theta}^*_{nT,k} - E^*\left(\check{\theta}^*_{nT,k}\right)\right) = o_p(1)$

We write $\sqrt{nT}\left(P^* \lim_{n\to\infty}\check{\theta}^*_{nT,k} - E^*\left(\check{\theta}^*_{nT,k}\right)\right)$ as

$$\begin{aligned}
&\sqrt{nT}\left(P^* \lim_{n\to\infty}\check{\theta}^*_{nT,k} - E^*\left(\check{\theta}^*_{nT,k}\right)\right) \\
&= \sqrt{nT}\Bigg(P^* \lim_{n\to\infty}\left[\left(\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right) 1\left(\sqrt{nT}\left|\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right| \leq M_{nT}\right)\right] \\
&- E^*\left[\left(\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right) 1\left(\sqrt{nT}\left|\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right| \leq M_{nT}\right)\right]\Bigg)
\end{aligned}$$

We consider only the case that $k = 1$ as the proof for other cases is similar. Let

$$D_{nT}(\theta, \alpha) = \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} U_{it\theta}(\theta, \alpha) \right)^{-1} \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} U_{it}(\theta, \alpha) \right),$$

then

$$\begin{aligned}
&\sqrt{nT} \left(P^* \lim_{n\to\infty} \check{\theta}^*_{nT,k} - E^* \left(\check{\theta}^*_{nT,k} \right) \right) \\
&= -P^* \lim_{n\to\infty} \left[\sqrt{nT} D^*_{nT}(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT})) 1 \left(\left| \sqrt{nT} D^*_{nT}(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT})) \right| \leq M_{nT} \right) \right] \\
&+ E^* \left[\sqrt{nT} D^*_{nT}(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT})) 1 \left(\left| \sqrt{nT} D^*_{nT}(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT})) \right| \leq M_{nT} \right) \right] \\
&= E \left[\sqrt{nT} D_{nT}(\theta, \alpha) 1 \left(\sqrt{nT} \left| D_{nT}(\theta, \alpha)) \right| \leq M_{nT} \right) \right] \Big|_{\theta = \hat{\theta}_{nT}, \alpha = \hat{\alpha}\left(\hat{\theta}_{nT}\right)} \\
&- P \lim_{n\to\infty} \sqrt{nT} D_{nT}(\theta, \alpha) 1 \left(\sqrt{nT} \left| D_{nT}(\theta, \alpha)) \right| \leq M_{nT} \right) \Big|_{\theta = \hat{\theta}_{nT}, \alpha = \hat{\alpha}\left(\hat{\theta}_{nT}\right)} \\
&= o_p(1)
\end{aligned}$$

where the last equality follows from the dominated convergence theorem.

Finally,

$$\begin{aligned}
&\sqrt{nT} \left(\tilde{\theta}_{nT,k} - \theta_0 \right) \\
&= \sqrt{nT} \left(\hat{\theta}_{nT} - \theta_T \right) + \sqrt{nT} \left(\theta_T - \theta_0 - \left[E^* \left(\check{\theta}^*_{nT,k} \right) - \hat{\theta}_{nT} \right] \right) \\
&= \sqrt{nT} \left(\hat{\theta}_{nT} - \theta_T \right) + \sqrt{nT} \left(\theta_T - \theta_0 - \left[P^* \lim_{n\to\infty} \hat{\theta}^*_{nT} - \hat{\theta}_{nT} \right] \right) \\
&+ \sqrt{nT} \left(P^* \lim_{n\to\infty} \hat{\theta}^*_{nT} - P^* \lim_{n\to\infty} \hat{\theta}^*_{nT,k} \right) + \sqrt{nT} \left(P^* \lim_{n\to\infty} \hat{\theta}^*_{nT,k} - P^* \lim_{n\to\infty} \check{\theta}^*_{nT,k} \right) \\
&+ \sqrt{nT} \left[P^* \lim_{n\to\infty} \check{\theta}^*_{nT,k} - E^* \left(\check{\theta}^*_{nT,k} \right) \right] \\
&= \sqrt{nT} \left(\hat{\theta}_{nT} - \theta_T \right) + o_p(1) \xrightarrow{d} N(0, \bar{E} \left[\mathcal{I}_i \right]^{-1})
\end{aligned}$$

as $n/T^3 \to 0$. Thus, Theorem 3 is proved. ■

3.10 Bibliography

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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