University of Connecticut

Department of Economics Working Paper Series

Bootstrap Inference Under Cross Sectional Dependence

by

Timothy Conley
Western University

Sílvia Gonçalves
McGill University, CIREQ and CIRANO

Min Seong Kim
University of Connecticut

Benoit Perron
Université of Montréal, CIREQ and CIRANO

Working Paper 2022-14
July 2022

365 Fairfield Way, Unit 1063
Storrs, CT 06269-1063
Phone: (860) 486-3022
Fax: (860) 486-4463
http://www.econ.uconn.edu/

Bootstrap inference under cross sectional dependence*

Timothy Conley
Department of Economics
Western University

Sílvia Gonçalves
Department of Economics
McGill University, CIREQ and CIRANO

Min Seong Kim
Department of Economics
University of Connecticut

Benoit Perron
Département de sciences économiques
Université of Montréal, CIREQ and CIRANO

July 22, 2022

Abstract

In this paper, we introduce a method of generating bootstrap samples with unknown patterns of cross sectional/spatial dependence which we call the spatial dependent wild bootstrap. This method is a spatial counterpart to the wild dependent bootstrap of Shao (2010) and generates data by multiplying a vector of independently and identically distributed external variables by the eigendecomposition of a bootstrap kernel. We prove the validity of our method for studentized and unstudentized statistics under a linear array representation of the data. Simulation experiments document the potential for improved inference with our approach. We illustrate our method in a firm-level regression application investigating the relationship between firms' sales growth and the import activity in their local markets using unique firm-level and imports data for Canada.

JEL Nos.: C12, C32, C38, C52
Keywords: bootstrap, cross sectional dependence, spatial HAC, eigendecomposition, economic distance

---

*Email: tconley3@uwo.ca, silvia.goncalves@mcgill.ca, min_seong.kim@uconn.edu, and benoit.perron@umontreal.ca. We are grateful for comments from Andrès Santos and two anonymous referees as well as from Donald Andrews, Xiaohong Chen, Jungbin Hwang, Yuichi Kitamura, Matt Webb, and Giuseppe Cavaliere, as well as seminar and conference participants at Queen's, Yale, York (UK), CUHK in Shenzhen, and Michigan State, and MEG, CESG, LAMES, CIREQ and the SCSE. We also thank Aldo Sandoval Hernandez for excellent research assistance. We are grateful to Danny Leung for his help in accessing Statistics Canada data. The views expressed in this paper do not necessarily reflect those of Statistics Canada. All empirical results have been reviewed to ensure that no confidential information is disclosed. Finally, we acknowledge funding from SSHRC and FRQSC.

1 Introduction

Economic data is often characterized by dependence and heterogeneity and accounting for these features is important when performing statistical inference. This paper presents a bootstrap method that is robust to spatial (cross sectional) dependence and heterogeneity of unknown forms in the context of a linear regression model. Spatially dependent observations often need to be indexed in more than one dimension and are not naturally ordered nor regularly spaced. This makes the application of the bootstrap potentially challenging. For instance, the spatial block bootstrap (see e.g. Lahiri and Zhu (2006), and Nordman, Lahiri and Fridley (2007)) requires a careful partition of the data into blocks which may not be feasible in many applications.

Our approach in this paper is based on a variation of the wild bootstrap and does not require resampling blocks of observations. We propose a residual-based wild bootstrap using a regression model, where the external random variables used to perturb the residuals are cross sectionally dependent. Covariances between pairs of external random variables are equal to a kernel weight that depends on a distance measure. Shao (2010) proposed this method for the time series case with distances equal to time gaps, calling it a "dependent wild bootstrap." The theory for the spatial context is not trivial; however, it is as easy to apply as in the time series case, requiring only (potentially imperfect) measures of distances between all pairs of observations.

The economics involved in applications with spatial data often suggests "economic distance" measures which can be used to model spatial dependence which decays with distance. We exploit the availability of such distance measures to generate bootstrap observations with cross sectional dependence. Distance measures can vary depending on the application and multiple metrics are also easily allowed in our setup. This allows our method to apply to panel data settings. For instance, our example application in Section 7 illustrates using our method with firm-level data where correlations across firms arise from both overlap in their local markets and similarity in their technologies.

We prove the first order asymptotic validity of our "spatial dependent wild bootstrap" under a set of regularity assumptions that are similar to those used in the spatial HAC literature (Conley (1999), Kelejian and Prucha (2007), Kim and Sun (2011)). In particular, we assume that the score vector for each observation i is a linear transformation of a possibly infinite number of common i.i.d. random innovations. Modelling spatial dependence as a linear process is quite common in the spatial econometrics literature (see e.g. Kelejian and Prucha (2007), Kim and Sun (2011, 2013) and Robinson (2011)). It avoids having to index observations in a Euclidean space, as required with mixing conditions, and a special case of this model is the popular spatial autoregressive (SAR) process. Compared to Shao (2010), who assumes a stationary mixing time series, our assumptions allow for heterogenous spatial dependence in dimensions higher than one, but we rule out nonlinear forms of dependence.

We generate spatially dependent external random variables using the eigendecomposition of the bootstrap kernel matrix. This matrix contains weights given by a kernel function evaluated at the distance measure and is equal to the bootstrap covariance matrix of the $n \times 1$ vector of external random variables. Hence, it must be positive semi-definite, and a sufficient condition is that we choose a bootstrap kernel function whose Fourier transform is weakly positive. A similar assumption is imposed by Shao (2010) in the one-dimensional time series context. We discuss a class of kernels that satisfy this condition when spatial dependence is of dimension higher than one and the distance is Euclidean. We also propose a modification for cases where the bootstrap kernel matrix is not positive semi-definite. Our bootstrap method contains several existing methods as special cases. One is the regular wild bootstrap. The other is the cluster wild bootstrap, popularized by Cameron et al. (2008) and studied by Djogbenou et al. (2019).

We provide a theoretical justification for bootstrap hypothesis tests based on studentized statistics requiring the use of a spatial HAC estimator for the original and the bootstrap test statistics. We allow for kernels used to construct test statistics to be different than those used for generating the bootstrap data. This is important since the bootstrap kernel function needs to be positive semi-definite, but one may want to use other kernels to construct test statistics. We also allow for the use of restricted residuals when computing bootstrap critical values for hypothesis tests. The use of restricted rather than unrestricted residuals often results in better size control.

The structure of the paper is as follows. In Section 2 we describe the setup and review the spatial HAC literature. In Section 3 we introduce the spatial dependent wild bootstrap and prove the consistency of the bootstrap distribution under a set of regularity assumptions that rely on a linear array representation for the score vector. The results of this section can be used to justify the construction of bootstrap percentile intervals, which do not require studentization. In Section 4 we discuss hypothesis testing based on studentized test statistics. Section 5 discusses an extension of our method to nonlinear models. Section 6 illustrates the finite sample performance of the method in comparison to alternative asymptotic-based methods. In Section 7, we illustrate our method in a firm-level regression investigating the relationship between a firm's sales growth and the import activity in its local market, where two metrics characterize residuals' dependence. An appendix contains mathematical derivations.

2 Linear regression with spatial or space-time dependence

We consider the following linear regression model

$$y_i = x_i'\beta + u_i, \quad i = 1, \ldots, n,$$

where the $(p \times 1)$ vector of regressors, x_i, and error term u_i, might be spatially or space-time dependent. The OLS estimator of β is

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \sum_{i=1}^{n} x_i y_i.$$

Under some regularity conditions, we know that

$$\left(Q^{-1} J_n Q^{-1}\right)^{-1/2} \sqrt{n} \left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, I_p\right), \tag{1}$$

where $Q = p\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} x_i x_i'$ is a positive definite matrix and

$$J_n = Var\left(n^{-1/2} \sum_{i=1}^{n} x_i u_i\right) \equiv \frac{1}{n} \sum_{i,j=1}^{n} E\left(V_i V_j'\right), \text{ where } V_i \equiv x_i u_i.$$

We assume throughout that J_n is nonsingular uniformly in n. According to (1), the asymptotic covariance matrix of $\hat{\beta}$ is $C_n = Q^{-1} J_n Q^{-1}$, which we need to estimate for inference on β. A consistent estimator of Q is

$$\hat{Q}_n = \frac{1}{n} \sum_{i=1}^{n} x_i x_i'.$$

Estimating J_n in the presence of spatial (cross sectional) or space-time correlation is more challenging as all pairs of observations could potentially be correlated.

The literature on spatial HAC inference confronts this problem by using auxiliary data on distances to model covariances between observations and to construct a nonparametric estimator for J_n, see Conley (1999). The basic idea is that measurements of a distance between observations can serve to characterize covariance structures in a manner analogous to time lags in a time series setting. Observations that are deemed close are modelled as potentially highly dependent, but those far enough away are approximately independent.

The spatial HAC literature has considered estimators of the form:

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

where $\hat{V}_i = x_i \hat{u}_i$ and $K(\cdot)$ is a real-valued kernel function with $K(0) = 1$. The distance between i and j is denoted d_{ij} and d_n is a scale parameter (bandwidth). We require $d_{ij} \geq 0$, $d_{ii} = 0$, $d_{ij} = d_{ji}$, but not the triangular inequality $d_{ij} \leq d_{ik} + d_{kj}$. This approach can be viewed as an extension of smoothed periodogram spectral density estimators that have long been used in the time series literature, e.g. Bartlett (1955) where distances are analogous to time lags. It can be viewed as a generalization of what are commonly called cluster or group dependence estimators, see e.g. Liang and Zeger (1986) and Moulton (1986), where observations are taken to be correlated within a known set of groups or clusters but independent across groups/clusters. These cluster estimators are a special case of spatial HAC with a discrete distance metric reflecting group membership and a uniform kernel K.

Distances need not be based upon physical locations; they can be much more general measures of 'economic distance' as in Conley (1999). For example, Conley and Ligon (2002) use an economic distance measure based on the transportation cost between countries in the context of a cross-country growth regression. Other examples include economic distances based on the similarity of input and output structures as considered by Chen and Conley (2001) and Conley and Dupor (2003). In many applications, distances can be based on attributes, e.g. input shares for firms. In these cases, observations' locations can be indexed by a vector of attributes $s_i \in \mathbb{R}^\tau$, and the distance between two units, i and j, may correspond to the Euclidean distance between s_i and s_j. Our method will be applicable with non-Euclidean metrics as well, possibly with a modification (see Section 3.3).

The existing spatial HAC literature also allows for the presence of measurement error in d_{ij} (see e.g. Conley (1999), Conley and Molinari (2007), Kelejian and Prucha (2007) and Kim and Sun (2011)), i.e. $\hat{J}_n$ is based on $\tilde{d}_{ij}$ rather than on the "true" measures d_{ij}, where $\tilde{d}_{ij}$ is such that

$$\tilde{d}_{ij} = d_{ij} + \xi_{ij},$$

where ξ_{ij} is a measurement error. We will follow this literature and also allow for measurement errors in $\tilde{d}_{ij}$ when applying the dependent wild bootstrap. Regularity conditions on ξ_{ij} will be discussed in Section 3.2.

Our method is readily applicable to a panel data setting where distances between observations are in part a function of the observations lead/lag in time. When distances between pairs of observations are derived from locations, e.g. s_i and s_j, time can be viewed as just another element of these location vectors with K defined to be a product kernel with one time series and one spatial component as in Conley (1999) (see also Kim and Sun, 2013). In general, distance can depend on any fixed number of metrics, see e.g. Kelejian and Prucha (2007) who suggest a HAC estimator with M metrics based on $K\left(\min_{1\leq m\leq M}\left\{d_{ij,m}/d_n\right\}\right)$, where $d_{ij,m}$ denotes the m^{th} distance measure between i and j. For ease of exposition, we present our theory using a single distance measure.

3 Bootstrap inference

3.1 The bootstrap method

The bootstrap data generating process is described as follows. Let

$$y_i^* = x_i'\hat{\beta} + u_i^*, \ i = 1, \ldots, n, \tag{3}$$

and generate

$$u_i^* = \hat{u}_i \cdot \eta_i, \tag{4}$$

where $\hat{u}_i = y_i - x_i'\hat{\beta}$ and η_i is an external random variable chosen by the researcher.

In this section, we rely on the unrestricted estimator $\hat{\beta}$ to generate the bootstrap observations on the dependent variable and discuss bootstrap consistency results that do not impose any constraint on β. We will discuss hypothesis testing in the next section, where $\hat{\beta}$ can be a restricted OLS estimator which imposes the null hypothesis under consideration. This is a key advantage of our method since the bootstrap literature has shown that imposing the null on the bootstrap DGP can result in large size improvements. See e.g. Davidson and MacKinnon (1999) and Djogbenou, MacKinnon and Nielsen (2019).

The choice of η_i is crucial. The regular wild bootstrap generates η_i in an i.i.d. fashion such that $E^*(\eta_i) = 0$ and $Var^*(\eta_i) = 1$ for all i. This implies that the bootstrap errors u_i^* are independently distributed, conditional on the data, with mean zero and variance $\hat{u}_i^2$. Hence, the wild bootstrap preserves heteroskedasticity but destroys cross sectional (or space-time) dependence.

Our goal in this paper is to generalize the regular wild bootstrap method so as to preserve cross sectional or space-time dependence and heterogeneity with a general form. As usual, we require that $E^*(\eta_i) = 0$ and $Var^*(\eta_i) = 1$ for all i. However, we do not generate η_i independently across i. Instead, given a potentially mismeasured distance $\tilde{d}_{ij}$ between observations, we generate $\{\eta_i : i = 1, \ldots, n\}$ such that their covariance structure is given by

$$Cov^*(\eta_i, \eta_j) = K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right) \quad \text{for all } (i,j), \tag{5}$$

where $K^*(\cdot)$ denotes a real valued kernel function and d_n^* is a bandwidth parameter. The choice of K^* and d_n^* is discussed below, where formal assumptions on these quantities will be introduced. We will also provide an algorithm on how to generate η_i such as to verify (5). Before we do so, let us provide some intuition for why this bootstrap method can be robust to cross sectional dependence. Let

$$\hat{\beta}^* = \left(\sum_{i=1}^n x_i x_i'\right)^{-1} \sum_{i=1}^n x_i y_i^*$$

denote the bootstrap OLS estimator. Using the bootstrap data generating process above, we can easily show that the bootstrap covariance matrix of $\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right)$ is

$$Var^*\left(\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right)\right) = \hat{Q}_n^{-1}\hat{J}_{boot,n}\hat{Q}_n^{-1},$$

where

$$\hat{J}_{boot,n} = Var^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i u_i^*\right) = \frac{1}{n}\sum_{i,j=1}^n x_i Cov^*\left(u_i^*, u_j^*\right) x_j' = \frac{1}{n}\sum_{i,j=1}^n x_i \hat{u}_i x_j' \hat{u}_j Cov^*(\eta_i, \eta_j) = \frac{1}{n}\sum_{i,j=1}^n K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\hat{V}_i\hat{V}_j',$$

given that $Cov^*(\eta_i, \eta_j) = K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)$ by (5). This shows that the spatial dependent wild bootstrap algorithm induces a bootstrap covariance matrix $\hat{J}_{boot,n}$ which is just an example of a spatial HAC covariance estimator, where the kernel function is K^* and the bandwidth parameter is d_n^*. Given the link to the spatial HAC covariance matrix estimator, we can expect this bootstrap method to be valid under conditions similar to those used in the spatial HAC literature.

Next, we describe our requirements on the bootstrap spatial kernel function K^*. To do so, we introduce the notion of "pseudo-neigbhors". Given the bandwidth parameter d_n^*, an observation j is defined as a pseudo-neighbor of i if its measured distance to i is less than d_n^*. More specifically, let

$$\mathscr{B}_{i,n}^* = \left\{ j : \tilde{d}_{ij} \leq d_n^* \right\}, \; \ell_{i,n}^* = \sum_{j=1}^{n} 1 \left\{ j \in \mathscr{B}_{i,n}^* \right\} \text{ and } \ell_n^* = \frac{1}{n} \sum_{i=1}^{n} \ell_{i,n}^*. \tag{6}$$

Then, $\mathscr{B}_{i,n}^*$ is the set of pseudo-neighbors that unit i has based on d_n^*, $\ell_{i,n}^*$ is the size of $\mathscr{B}_{i,n}^*$ and ℓ_n^* is its average. Note that $\mathscr{B}_{i,n}^*$ is a random set due to measurement error in $\tilde{d}_{ij}$, implying that both $\ell_{i,n}^*$ and ℓ_n^* are random sequences.

The following condition specifies the requirements on the spatial kernel K^*.

Assumption 1 (i) The kernel function $K^* : \mathbb{R} \to [-1,1]$ satisfies $K^*(0) = 1$, and $K^*(z) = K^*(-z)$ for all $z \in \mathbb{R}$. (ii) $\frac{1}{E\ell_n^*} \sup_i E\left(\sum_{j \notin \mathscr{B}_{i,n}^*} \left| K^* \left(\frac{\tilde{d}_{ij}}{d_n^*} \right) \right| \right) = O(1)$ and $\frac{1}{E\ell_n^*} \sup_i \sum_{j \notin \mathscr{B}_{i,n}^*} \left| K^* \left(\frac{\tilde{d}_{ij}}{d_n^*} \right) \right| = O_P(1)$; (iii) The matrix $\mathbb{K}_n^* = \left[K^* \left(\tilde{d}_{ij}/d_n^* \right) \right]_{i,j=1}^n$ is symmetric and positive semi-definite for all n, almost surely.

Part (i) is a standard assumption in the HAC literature which is satisfied by standard kernels such as the rectangular, Bartlett, Parzen and Quadratic Spectral (QS) kernels. Parts (ii) and (iii) are new to our context. Assumption 1 (ii) is automatically satisfied by truncated kernels for which $K^*(z) = 0$ for $|z| \geq 1$, regardless of the distance used. It allows for kernels that do not truncate provided the tails of K^* decay sufficiently fast. Providing more primitive conditions for general distances is difficult, but we can do so for the Euclidean distance. For instance, with locations indexed on the line (such as a time series) and assuming away the presence of measurement error in distances, this condition is satisfied if $\int_{-\infty}^{\infty} |K^*(u)|\, du < \infty$**.** Standard kernels used in time series analysis satisfy this condition, including the QS kernel and the exponential (Gaussian) kernel. Shao (2010) excludes these kernels by assuming a truncated kernel. Similarly, if we map locations into a two-dimensional lattice, a sufficient condition for part (ii) is that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| K^* \left(\sqrt{x^2 + y^2} \right) \right| dxdy < \infty$ or equivalently $\int_0^1 r\, |K^*(r)|\, dr < \infty$, a condition that is satisfied by the Gaussian kernel.

Part (iii) is a high level condition that requires the matrix of weights $\mathbb{K}_n^* = \left[K^* \left(\tilde{d}_{ij}/d_n^* \right) \right]_{i,j=1}^n$ to be symmetric and positive semi-definite. The reason why we impose this condition is that $\mathbb{K}_n^*$ is the bootstrap variance matrix of $\eta = (\eta_1, \ldots, \eta_n)'$ and therefore needs to satisfy these conditions. When distances correspond to Euclidean distances between points in $\mathbb{R}^\tau$, a sufficient condition is that the kernel K^* be positive definite.

Bochner's Theorem provides a necessary and sufficient condition for a kernel K^* to be positive definite in $\mathbb{R}^\tau$: the Fourier transform of the kernel function K^* is weakly positive. For example, with locations indexed on the line, $\int_{-\infty}^{\infty} K^*(u)\, e^{-iur} du \geq 0$ for all $r \in \mathbb{R}$. This well-known condition is met by a number of kernel functions such as the Bartlett kernel or the Parzen kernel. Shao (2010) imposes this condition, see his equation (2), when studying the properties of the dependent wild bootstrap for the one-dimensional dependent context. In higher dimensions, an analogous condition applies; $K^*(x)$ is positive definite if it is of the form:

$$K^*(x) = \Gamma\left(\frac{\tau}{2}\right)\int_0^{\infty}\left(\frac{2}{rx}\right)^{(\tau-2)/2} J_{(\tau-2)/2}(rx)\,dF(r)\,,\ x \geq 0, \tag{7}$$

where F is a probability distribution function on $[0,\infty)$ and $J_{(\tau-2)/2}(\cdot)$ is a Bessel function of order $(\tau-2)/2$. This characterization follows from simplifying the integrals in the Fourier transform via polar coordinates to exploit the radial symmetry of $K^*(x)$. Discussions of positive definite kernels can be found in, e.g., Conley (1999), Chen and Conley (2001), Gneiting (2002), or Kelejian and Prucha (2007), see Yaglom (1987) for a textbook characterization of this class of functions.

The set of positive definite kernels depends on the dimension τ and it shrinks as τ grows, implying that a kernel which is positive definite in τ dimensions will be positive definite in any smaller number of dimensions. An example of this class of kernel functions from Kelejian and Prucha (2007) is:

$$K_v^*(x) = \begin{cases} (1-x)^v\,, & 0 \leq x \leq 1 \\ 0, & x > 1 \end{cases} \tag{8}$$

where $v \geq (\tau+1)/2$. This is similar to the sharp (or steep) origin kernel in Phillips, Sun and Jin (2007).

The set of kernels that are positive definite in any dimensional Euclidean space ($\tau = \infty$) can be represented as:

$$K^*(x) = \int_0^{\infty} \exp(-x^2r^2)dF(r)\,,\ x \geq 0. \tag{9}$$

An example kernel in this class is

$$K^*(x) = \exp(-x^2), \tag{10}$$

which is a Gaussian kernel (see e.g. Stein (1999), p.44). Choosing a kernel that is positive definite in any dimensional Euclidean space avoids the need to know the dimension τ. When distances are non-Euclidean, we do not know if there is a class of kernels guaranteed to be positive definite; we discuss one strategy to overcome this issue in Section 3.3.

A recent paper by Kojevnikov (2021) also considers a modification of the dependent wild bootstrap kernel function that satisfies Assumption 1(iii) in the context of a network dependent model. His weighting function depends on the topology of the network and requires a structure of locations, which we do not require.

Under Assumption 1, $\mathbb{K}_n^*$ is symmetric and positive semi-definite, which implies that there exists Φ_n such that

$$\mathbb{K}_n^* = \Phi_n\Lambda_n\Phi_n',$$

where Λ_n is a diagonal matrix with the nonnegative eigenvalues of $\mathbb{K}_n^*$ and the columns of Φ_n are the associated orthonormal eigenvectors ($\Phi_n'\Phi_n = I_n$). We can write

$$\Phi_n = [\phi_1, ..., \phi_n] \text{ with } \phi_k = \begin{bmatrix} \phi_{1k} \\ \vdots \\ \phi_{nk} \end{bmatrix} \text{ and } \Lambda_n = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \text{ with } \lambda_i \geq 0 \text{ for all } i.$$

Thus, we can generate η_i as follows. Letting $L_n = \Phi_n \Lambda_n^{1/2}$, we set

$$\underset{n\times 1}{\eta} = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_n \end{pmatrix} = L_n \cdot v, \quad v \sim \text{ i.i.d.}\,(0, I_n)\,, \tag{11}$$

where η_i is the i^{th} element of η. This algorithm implies that $E^*(\eta) = 0$ and $Var^*(\eta) = L_n L_n' = \mathbb{K}_n^*$.

An attractive feature of our bootstrap method is that it contains several existing methods as special cases. The simplest example is the wild bootstrap with $\mathbb{K}_n^* = I_n$ and $L_n = I_n$.

A more complex example is the cluster wild bootstrap proposed by Cameron, Gelbach and Miller (2008). This method is very popular in applied work and its theoretical properties have been recently studied by Djogbenou, Nielsen and MacKinnon (2019). The usual way of describing the cluster wild bootstrap is as follows. Suppose we can partition the sample of n observations into G groups of observations, so that the $n \times 1$ vector $\hat{u}$ can be partitioned as $\hat{u} = (\hat{u}_1, \ldots, \hat{u}_G)'$, where for each g, $\hat{u}_g = \left(\hat{u}_{1g}, \ldots, \hat{u}_{n_g,g}\right)'$, and $n = \sum_{g=1}^G n_g$. The cluster wild bootstrap generates residuals as follows:

$$\hat{u}_{jg}^* = \hat{u}_{jg} \cdot \varepsilon_g,$$

where ε_g is a common shock to all observations in cluster g.

One way to map this setup to ours is to order the observations by cluster. The weighting matrix $\mathbb{K}_n^*$ has typical element given by $\mathbb{K}_n^*(l,m) = 1\,(l \text{ and } m \text{ belong to same cluster})$, that is the (l,m) element is 1 if the two observations belong to the same cluster and 0 otherwise. This results in a block diagonal $\mathbb{K}_n^*$ with matrices of ones of dimensions $n_g \times n_g$ along the diagonal. In other words,

$$\mathbb{K}_n^* = \begin{pmatrix} \mathbb{K}_{1,n}^* & 0 & & 0 \\ 0 & \mathbb{K}_{2,n}^* & & \\ & & \ddots & 0 \\ 0 & & 0 & \mathbb{K}_{G,n}^* \end{pmatrix},$$

where $\mathbb{K}_{g,n}^* = \iota_{n_g}\iota_{n_g}'$, with $\iota_{n_g} = (1, \ldots, 1)'$ for each $g = 1, \ldots, G$. The eigendecomposition of $\mathbb{K}_n^*$ is equal to $L_n L_n'$, where L_n is a block diagonal matrix with $L_{g,n}$ on the main diagonal and $L_{g,n}$ is an $n_g \times n_g$ matrix whose first column is a vector of ones and the remaining columns are zero. Thus, setting $\eta = L_n \cdot v$ where $v \sim^{iid} (0, I_n)$ is equivalent to generating an $n \times 1$ vector of shocks partitioned as $\eta = (\eta_1', \ldots, \eta_G')'$, where for each g cluster, $\eta_g = (\varepsilon_g, \ldots, \varepsilon_g)'$ contains the same shock ε_g.

3.2 Bootstrap distribution consistency

In this section, we examine the properties of our bootstrap procedure. To establish the asymptotics, we assume that the $p \times 1$ vector of scores V_i has a linear array representation. In particular, we make the following assumption.

Assumption 2

(i) For $a = 1, \ldots, p$,

$$V_i^{(a)} = \sum_{l=1}^{\infty} r_{il}^{(a)} e_l, \tag{12}$$

where $V_i^{(a)}$ is the a-th component of V_i, e_l is a random innovation, and $r_{il}^{(a)}$ is a nonstochastic weight.

(ii) $e_l \sim^{iid} (0,1)$ and there exists a constant $M < \infty$ such that $E\left(e_l^4\right) < M$.

(iii) For each l, $\sum_{i=1}^{\infty} \left| r_{il}^{(a)} \right| < M$, and for each i, $\sum_{l=1}^{\infty} \left| r_{il}^{(a)} \right| < M$, for all $a = 1, \ldots, p$.

Assumption 2 is sufficient for proving that a central limit theorem applies to the scaled average of the scores, i.e. that $J_n^{-1/2} n^{-1/2} \sum_{i=1}^{n} V_i \xrightarrow{d} N(0, I_p)$. A linear transformation of i.i.d. random variables is often employed in the literature to characterize spatial (or spatiotemporal) processes. See, for example, Kelejian and Prucha (2007), Kim and Sun (2011, 2013), Robinson (2011), Robinson and Thawornkaiwong (2012), Lee and Robinson (2016) and Hidalgo and Schafgans (2017). In particular, our linear array model in (12) is the same as in Robinson (2011) (see also Hidalgo and Schafgans (2017) for a panel extension of this model), with the difference that we impose the linear array representation directly on the score vector rather than assuming that the error term u_i is a linear array.

As in the time series context, an alternative to a linear array representation would be to assume some mixing-type conditions in the cross sectional dimension, as e.g. in Conley (1999). This is also the approach of Shao (2010), who considers the one-dimensional (time series) case. Although mixing assumptions have the advantage of allowing for nonlinear forms of dependence, this type of conditions are harder to deal with in the cross sectional context than in the time series context and in particular more difficult to apply without directly indexing observations in Euclidean space as in Conley (1999). The linear array representation is general enough to cover most spatial models used in economics, including in particular the class of spatial autoregressive (SAR) models as a special case.[1] Because the coefficients $r_{il}^{(a)}$ are a function of i, we allow for heterogeneity in the second and higher order moments of $\{V_i\}$.

[1]Distance construction can be problematic for some SAR specifications. While SAR models are linear processes they do not necessarily have a covariance structure that can be characterized by a set of distances. In particular, simple graph distances in some SAR models will not fully characterize the implied covariance structure, see Martellosio (2012). We leave the characterization of SAR models for which an array of distances can be constructed for future work.

A key requirement for bootstrap validity is that the asymptotic bootstrap variance of $\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right)$ replicates the asymptotic variance of $\sqrt{n}\left(\hat{\beta} - \beta\right)$. This entails showing the consistency of the bootstrap variance $\hat{J}_{boot,n}$ towards J_n. For this result, we need to impose further restrictions on the cross sectional dependence of $\{V_i\}$. Define

$$K_q^* = \lim_{z \to 0} \frac{1 - K^*(z)}{|z|^q} \text{ for } q \in [0, \infty)$$

and let $q_0^* = \max\left\{q : K_q^* < \infty\right\}$ be the Parzen characteristic exponent of $K^*(z)$. For instance, $q_0^* = 1$ for the Bartlett and Kelejian and Prucha (2007) kernels and $q_0^* = 2$ for the QS, Parzen, and Gaussian kernels. Larger values of q_0^* imply smoother kernel functions at 0 and a smaller asymptotic bias for the HAC estimator, ceteris paribus (see Andrews (1991) for the time series HAC estimator and Kim and Sun (2011) for its spatial analogue).

As in the HAC literature, the asymptotic bias of the bootstrap variance estimator depends on the decaying rate of spatial dependence as a function of the distance metric and q_0^*. Our next assumption follows Kim and Sun (2011, 2013) and is used to control this bias.

Assumption 3 There exists a constant $C_{q_0^*} < \infty$ such that $\frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n ||E\left(V_i V_j'\right)|| d_{ij}^{q_0^*} < C_{q_0^*}$, for all n, where $||\cdot||$ denotes the Euclidean norm of a matrix.

Assumption 3 requires the degree of cross sectional dependence between V_i and V_j to decrease as a function of the "true" distance d_{ij} (and the degree of smoothness of the kernel function at zero as dictated by q_0^*). In the time series context, this assumption is implied by a standard smoothness condition on the spectral density function of $\{V_t\}$ evaluated at zero: $\sum_{j=-\infty}^{+\infty} ||E\left(V_t V_{t+j}'\right)|| \, |j|^{q_0^*} < \infty$ (see Andrews, 1991, eq. (3.4)).

Our next assumption imposes conditions on the measurement error ξ_{ij} that imply that the presence of measurement error in $\tilde{d}_{ij}$ does not change this absolute summability condition.

Assumption 4 (i) $\left\{\xi_{ij}\right\}$ are independent of $\{e_l\}$ and $\{x_i\}$; (ii) $\left|\xi_{ij}\right| \leq c_\xi$ for all $i, j = 1, \ldots, n$.

Assumption 4 assumes that measurement errors are bounded, which is standard in the spatial HAC literature, see e.g. Conley (1999) and Keleijan and Prucha (2007). The independence assumption implies that $\left\{\xi_{ij}\right\}$ is independent of $\{V_i\}$, greatly simplifying the proof.

Our next assumption controls the number of pseudo-neighbors that a given observation i is allowed to have and corresponds to Assumption 5 in Kim and Sun (2011). In particular, we require that each observation i has at most $cE\ell_n^*$ pseudo-neighbors, where c is an arbitrary (large) constant and $E\ell_n^*$ is the expected average number of pseudo-neighbors.

Assumption 5 For all i, $\ell_{i,n}^* \leq cE\ell_n^*$, a.s., for some constant $c > 0$.

Assumption 5 rules out the possibility that most observations are concentrated around some locations and not others. Note that in the time series context with regularly spaced observations $\ell_{i,n}^* \leq 2d_n^*$ (with equality for all $i \in [d_n^*, n - d_n^* + 1]$ if d_n^* is an integer) and $E\ell_n^* > d_n^*$, and this implies that $\ell_{i,n}^* \leq 2E\ell_n^*$ for all i. Thus, Assumption 5 is automatically satisfied in this case. We can also see that $E\ell_n^*$ and d_n^* are related to each other and both parameters can be thought of as bandwidth parameters. This is true more generally, with $E\ell_n^*$ clearly increasing with d_n^*. More specifically, as discussed by Kim and Sun (2011, p. 354), for locations on a regular lattice, it is natural to assume that $E\ell_n^*$ is proportional to $d_n^{*\tau}$, where τ is the dimension of the space. When $\tau = 1$, this implies that $E\ell_n^*$ is proportional to d_n^*, as discussed above, whereas for $\tau = 2$ we obtain $E\ell_n^* = \alpha d_n^{*2}$ for some constant α. Since $E\ell_n^*$ (and d_n^*) plays the role of the bandwidth parameter in the usual time series HAC literature, we will let $E\ell_n^* \to \infty$ as $n \to \infty$ but at a slower rate than n when deriving our results. This is because we will show that (as usual) a larger $E\ell_n^*$ (and hence a larger d_n^*) reduces the asymptotic bias but increases the variance of the bootstrap variance estimator at the rate $O\left(E\ell_n^*/n\right)$.

The final assumption imposes regularity conditions on the regressors. Note that Assumption 1 implies that a similar moment condition holds for the scores V_i, i.e. $E\left\|V_i\right\|^4 \leq M$.

Assumption 6 (i) There exists a constant $M < \infty$ such that $E\left\|x_i\right\|^4 \leq M$; (ii) $n^{-1}\sum_{i=1}^n x_i x_i' \xrightarrow{P} Q$, a positive definite matrix.

Theorem 3.1 *Suppose Assumptions 1-6 hold. If $d_n^*, E\ell_n^* \to \infty$ such that $E\ell_n^*/n^{1/2} = o(1)$ and $E^*\left|v_i\right|^4 \leq M$, then we have*

$$\sup_{x \in \mathbb{R}^p}\left|P^*\left(\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right) \leq x\right) - P\left(\sqrt{n}\left(\hat{\beta} - \beta\right) \leq x\right)\right| = o_p(1)$$

as $n \to \infty$.

Theorem 3.1 states the consistency of the bootstrap distribution of $\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right)$. The proof of Theorem 3.1 is in the Appendix. It follows by showing that $\hat{J}_{boot,n} - J_n \to^P 0$ and

$$\left(Q^{-1}J_nQ^{-1}\right)^{-1/2}\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right) \to^{d^*} N\left(0, I_p\right),$$

in probability, see the Appendix for the definition of $\to^{d^*}$ in probability.

The rate condition on the bandwidth parameter $E\ell_n^*/\sqrt{n} \to 0$ is stronger than the rate $E\ell_n^*/n \to 0$ used for showing the consistency of the (spatial HAC) variance estimator $\hat{J}_{boot,n}$ (analogously Theorem 1 of Kim and Sun (2011) and the proof of Lemma A.2). The stronger rate condition on $E\ell_n^*$ is used to prove that a bootstrap central limit theorem holds for the scaled average of the bootstrap scores, $n^{-1/2}\sum_{i=1}^n V_i^* \equiv n^{-1/2}\sum_{i=1}^n V_i \eta_i$.

For the one-dimensional context, Shao (2010) proves the validity of the dependent wild bootstrap for smooth functions of sample means of stationary mixing time series data that are possibly irregularly spaced in time. His rate condition on the bandwidth parameter (which is given by ℓ_n^* rather than $E\ell_n^*$ since there is no measurement error) is more stringent than ours, requiring that $\ell_n^*/n^{1/3} \to 0$ as $n \to \infty$. He also assumes that the external random variables η_i are ℓ_n^*-dependent, an assumption we do not make. As he remarks after his Theorem 3.1, this assumption makes the proof of the bootstrap central limit theorem easier as he relies on a blocking argument that exploits the ℓ_n^*-dependence of the process η_i. Our method of proof is different from his. In particular, we use the eigendecomposition of $\mathbb{K}_n^*$ to write $\eta_i = \sum_{k=1}^n \left(\sqrt{\lambda_k}\phi_{ik}\right) v_k$, where λ_k and ϕ_k are the k^{th} eigenvalue and eigenvector of $\mathbb{K}_n^*$, and $v_k \sim$ i.i.d.$(0,1)$ independently of the original sample. It follows that $n^{-1/2}\sum_{i=1}^n V_i^*$ can be written as $n^{-1/2}\sum_{k=1}^n \omega_k v_k$, where conditionally on the original sample, $\omega_k = \sqrt{\lambda_k} V'\phi_k$ is a known function and $v_k \sim$ i.i.d.$(0,1)$. Hence, we apply Lyapunov's CLT for independent heterogeneous arrays and rely on the rate condition $E\ell_n^*/n^{1/2} \to 0$ as $n \to \infty$ to verify the Lyapunov's condition. Note that despite the (conditional) independence of the array $\omega_k v_k$, the conditional bootstrap variance of $n^{-1/2}\sum_{i=1}^n V_i^*$ is still robust to spatial dependence. Indeed, we can show that this variance is equal to

$$n^{-1}\sum_{k=1}^n \omega_k^2 = V'\left(n^{-1}\sum_{k=1}^n \lambda_k \phi_k \phi_k'\right)V = n^{-1}V'\Phi_n\Lambda_n\Phi_n'V = n^{-1}V'\mathbb{K}_n^*V = n^{-1}\sum_{i,j=1}^n K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)V_iV_j' = J_{boot,n},$$

a spatial HAC variance estimator. Although this estimator is infeasible as it is based on $V_i = x_i u_i$, we show in Lemma A.2 (ii) that the difference between this estimator and its feasible version $\hat{J}_{boot,n}$ is asymptotically negligible under our assumptions.

We now provide an algorithm to compute valid bootstrap percentile intervals.

Algorithm: Bootstrap percentile intervals

(i) Compute

$$\hat{\beta} = \left(\sum_{i=1}^n x_i x_i'\right)^{-1}\sum_{i=1}^n x_i y_i, \quad \hat{u}_i = y_i - x_i'\hat{\beta}, \text{ and } \hat{V}_i = x_i\hat{u}_i, i = 1,\dots,n.$$

(ii) For a given bandwidth choice d_n^* (to be discussed later), compute the matrix $\mathbb{K}_n^* = \left[K^*\left(\tilde{d}_{ij}/d_n^*\right)\right]_{i,j=1}^n$ and its eigendecomposition $\mathbb{K}_n^* = \Phi_n\Lambda_n\Phi_n'$, where Λ_n is a diagonal matrix with the nonnegative eigenvalues of $\mathbb{K}_n^*$ and the columns of Φ_n are the associated orthonormal eigenvectors. Let $L_n = \Phi_n\Lambda_n^{1/2}$.

(iii) Generate an $n\times 1$ vector $v \sim (0, I_n)$. We recommend using a sequence of standard normal variables, but other distributions like independent Rademacher random variables ($+1$ or -1 with equal probability) could be used. Then generate a sequence of random variables $\{\eta_i : i = 1,\dots,n\}$ by multiplying this vector v by L_n:

$$\eta = L_n v.$$

(iv) Let

$$y_i^* = x_i'\tilde{\beta} + u_i^*, \text{ where } u_i^* = \tilde{u}_i\eta_i,$$

and compute the bootstrap OLS estimator

$$\hat{\beta}^* = \left(\sum_{i=1}^n x_i x_i'\right)^{-1} \sum_{i=1}^n x_i y_i^*.$$

(v) Repeat steps (iii)-(iv) B times. Symmetric $(1-\alpha)$ confidence intervals can be obtained as:

$$\left[\hat{\beta} - \hat{q}_{1-\alpha}, \hat{\beta} + \hat{q}_{1-\alpha}\right],$$

where $\hat{q}_{1-\alpha}$ is the $(1-\alpha)$ quantile of the distribution of $\left|\hat{\beta}^* - \hat{\beta}\right|$. If preferred, equal-tailed intervals can be obtained as:

$$\left[\hat{\beta} - \hat{p}_{1-\alpha/2}, \hat{\beta} - \hat{p}_{\alpha/2}\right],$$

where $\hat{p}_{\alpha/2}$ and $\hat{p}_{1-\alpha/2}$ are, respectively, the $\alpha/2$ and $1-\alpha/2$ quantiles of the distribution of $\left(\hat{\beta}^* - \hat{\beta}\right)$.

3.3 Extension to bootstrap kernels that are not positive semi-definite

The asymptotic validity of the spatial dependent wild bootstrap described in the previous section depends crucially on the positive semi-definiteness and symmetry properties of $\mathbb{K}_n^*$ (cf. Assumption 1(iii)). These assumptions guarantee that the bootstrap covariance matrix of the external random vector η is non-negative definite, thus ensuring that the bootstrap measure is a valid measure. Although the symmetry property of the distance metric (which we assume for $\tilde{d}_{ij}$) implies the symmetry of $\mathbb{K}_n^*$, the assumption that $\mathbb{K}_n^*$ is positive semi-definite can fail when $\tilde{d}_{ij}$ is non-Euclidean. In this case, choosing K^* as a non-negative definite function (as e.g., letting K^* be the Gaussian kernel) does not guarantee the non-negative definiteness of $\mathbb{K}_n^*$.

In this section, we provide an alternative bootstrap method that does not require $\mathbb{K}_n^*$ to be positive semi-definite[2]. The idea is to transform $\mathbb{K}_n^*$ into a positive semi-definite matrix $\mathbb{M}_n^*$ and choose η such that $E^*(\eta\eta') = \mathbb{M}_n^*$. For example, if we assume that $\hat{V}'\mathbb{K}_n^*\hat{V}$ is positive definite, we could choose:

$$\mathbb{M}_n^* = \mathbb{K}_n^*\hat{V}\left(\hat{V}'\mathbb{K}_n^*\hat{V}\right)^{-1}\hat{V}'\mathbb{K}_n^*.$$

Note that this matrix has rank equal to the rank of $\hat{V}$, which is p. The next theorem implies the validity of percentile confidence intervals constructed using the previous algorithm by defining $L_n = \Phi_n\Lambda_n^{1/2}$, where Λ_n is a diagonal matrix with the eigenvalues of $\mathbb{M}_n^*$ on the diagonal and Φ_n is the matrix of associated eigenvectors. In addition, in this case, the external variables must be generated from the standard normal distribution.

Theorem 3.2 *Suppose the assumptions of Theorem 3.1 hold but $\mathbb{K}_n^*$ is not necessarily positive semi-definite. Suppose $\hat{V}'\mathbb{K}_n^*\hat{V}$ is positive definite and define $\mathbb{M}_n^* = \mathbb{K}_n^*\hat{V}\left(\hat{V}'\mathbb{K}_n^*\hat{V}\right)^{-1}\hat{V}'\mathbb{K}_n^*$. If $v \sim N(0, I_n)$,*

$$\sup_{x\in\mathbb{R}^p}\left|P^*\left(\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right) \le x\right) - P\left(\sqrt{n}\left(\hat{\beta} - \beta\right) \le x\right)\right| = o_p(1)$$

as $n \to \infty$ and $d_n^* \to \infty$ such that $E\ell_n^*/n \to 0$.

[2] We thank Andrès Santos for this suggestion.

This result proves the consistency of the bootstrap distribution of $\sqrt{n}\left(\hat{\beta}^{*}-\hat{\beta}\right)$, thus justifying the construction of bootstrap percentile intervals for β even when $\mathbb{K}_n^*$ is not positive semi-definite.

Remark 1 *Assuming that $\hat{V}'\mathbb{K}_n^*\hat{V}$ is positive definite is weaker than assuming that $\mathbb{K}_n^*$ is positive definite. In practice, it is possible to ensure that $\hat{V}'\mathbb{K}_n^*\hat{V}$ is a positive definite matrix by replacing its negative eigenvalues by $\varepsilon > 0$, a small positive constant, as suggested by Politis (2011) (see also McMurry and Politis, 2010). Since under our remaining assumptions, $n^{-1}\hat{V}'\mathbb{K}_n^*\hat{V}$ converges in probability to J_n, and J_n is assumed to be positive definite uniformly in n, this regularization is asymptotically negligible.*

The Gaussianity assumption on v is crucial for proving Theorem 3.2, as we explain here. For simplicity, assume that $p = 1$, i.e. that $\hat{V}_i = x_i\hat{u}_i$ is a scalar. Then, $n^{-1}\hat{V}'\mathbb{K}_n^*\hat{V}$ is a scalar and $\mathbb{K}_n^*\hat{V}$ is $n \times 1$, which implies that the rank of $\mathbb{M}_n^*$ is 1. This implies that

$$\eta = L_n v = \sqrt{\lambda_1}\phi_1 v_1,$$

where λ_1 is the only non-zero eigenvalue of $\mathbb{M}_n^*$ and ϕ_1 is its corresponding eigenvector. It follows that $\eta_i = \sqrt{\lambda_1}\phi_{i1}v_1$, for all $i = 1,\ldots,n$. Since $u_i^* = \hat{u}_i\eta_i$, we have that

$$u_i^* = \hat{u}_i\eta_i = (\sqrt{\lambda_1}\phi_{i1}\hat{u}_i)v_1,$$

implying that the bootstrap scores $V_i^* \equiv x_i u_i^* = (\hat{V}_i\sqrt{\lambda_1}\phi_{i1})v_1$ are all proportional to v_1. The implication is that $n^{-1/2}\sum_{i=1}^n V_i^*$ does *not* satisfy a bootstrap central limit theorem. To obtain a Gaussian distribution, we need to generate v_1 as $N(0,1)$. Under this condition, conditionally on the original sample,

$$n^{-1/2}\sum_{i=1}^n V_i^* = (n^{-1/2}\hat{V}'\sqrt{\lambda_1}\phi_1)v_1 \sim N\left(0, n^{-1}\hat{V}'\mathbb{M}_n^*\hat{V}\right),$$

where by construction,

$$n^{-1}\hat{V}'\mathbb{M}_n^*\hat{V} = n^{-1}\hat{V}'\mathbb{K}_n^*\hat{V} \equiv \hat{J}_{boot,n}.$$

Thus, conditionally on the data,

$$\sqrt{n}\left(\hat{\beta}^{*}-\hat{\beta}\right) = \hat{Q}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i^* \sim N\left(0, \hat{Q}_n^{-1}\hat{J}_{boot,n}\hat{Q}_n^{-1}\right).$$

Theorem 3.2 follows because $\hat{Q}_n^{-1}\hat{J}_{boot,n}\hat{Q}_n^{-1} - C_n = o_p(1)$, where $C_n = Q^{-1}J_nQ^{-1}$ (apply Lemma A.2, noting that it does not require the positive definiteness assumption on $\mathbb{K}_n^*$).

When $p > 1$, the same argument applies except that the rank of $\mathbb{M}_n^*$ is now $p > 1$, implying that we can write

$$n^{-1/2}\sum_{i=1}^n V_i^* = \sum_{k=1}^p (n^{-1/2}\hat{V}'\sqrt{\lambda_k}\phi_k)v_k,$$

in which case the joint Gaussianity of $(v_1,\ldots,v_p)$ determines the Gaussianity of the scaled average of the bootstrap scores.

4 Hypothesis testing

The previous results justify the construction of bootstrap percentile confidence intervals. These are based on the bootstrap quantiles of the unstudentized statistic $\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right)$. In this section, we consider bootstrap tests based on studentized statistics. Specifically, we consider testing

$$H_0 : R\beta = r_0 \text{ vs } H_1 : R\beta \neq r_0, \tag{13}$$

where R is a $r \times p$ matrix with $r \leq p$ and r_0 is a $r \times 1$ vector.

For testing (13), we employ the Wald statistic given by

$$\mathscr{W}_n = \sqrt{n}\left(R\hat{\beta} - r_0\right)' \left[R\hat{Q}_n^{-1}\hat{J}_n\hat{Q}_n^{-1}R'\right]^{-1} \sqrt{n}\left(R\hat{\beta} - r_0\right), \tag{14}$$

a special case of which is the squared t statistic when $r = 1$. The Wald test statistic requires the use of a spatial HAC estimator given by $\hat{J}_n$. Our assumption is that this estimator is of the usual form

$$\hat{J}_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n}\right)\hat{V}_i\hat{V}_j',$$

where $K(\cdot)$ and d_n correspond to a spatial kernel function and a bandwidth parameter that are possibly different from the bootstrap kernel function $K^*(\cdot)$ and the bootstrap bandwidth parameter d_n^*.

For bootstrap testing, using restricted residuals is often preferable as this reduces the size distortions. In this case, the bootstrap data generating process can be described as follows. Let $\tilde{\beta}$ denote the restricted OLS estimator of β obtained under H_0. We generate bootstrap data as $y_i^* = x_i'\tilde{\beta} + u_i^*$, where $u_i^* = \tilde{u}_i\eta_i$, with $\tilde{u}_i = y_i - x_i'\tilde{\beta}$ and η_i generated as before. In what follows, we denote by $\ddot{\beta}$ either $\tilde{\beta}$ or $\hat{\beta}$, depending on whether we use the restricted or the unrestricted residuals.

The bootstrap Wald statistic is then defined as

$$\mathscr{W}_n^* = \sqrt{n}\left(R\hat{\beta}^* - R\ddot{\beta}\right)' \left[R\hat{Q}_n^{-1}\hat{J}_n^*\hat{Q}_n^{-1}R'\right]^{-1} \sqrt{n}\left(R\hat{\beta}^* - R\ddot{\beta}\right)$$

where

$$\hat{J}_n^* = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n}\right)\hat{V}_i^*\hat{V}_j^{*\prime},$$

with $\hat{V}_i^* = x_i\hat{u}_i^*$, $\hat{u}_i^* = y_i^* - x_i'\hat{\beta}^*$, and $\hat{\beta}^*$ the unrestricted OLS estimator from regressing y_i^* on x_i.

Similarly to the Wald test statistic, the bootstrap Wald statistic also requires the choice of a spatial kernel and a bandwidth parameter. Our approach in this paper is to use the same kernel and bandwidth for studentizing the two Wald test statistics. Hence, our approach is similar to the naive bootstrap approach considered by Gonçalves and Vogelsang (2011).

To establish the asymptotic validity of the bootstrap Wald test, we need to impose conditions on K and d_n. In order to do so, we define another set of pseudo-neighbors of i using the bandwidth d_n:

$$\mathscr{B}_{i,n} = \left\{ j : \tilde{d}_{ij} \leq d_n \right\}, \; \ell_{i,n} = \sum_{j=1}^{n} 1\left\{ j \in \mathscr{B}_{i,n} \right\} \text{ and } \ell_n = \frac{1}{n} \sum_{i=1}^{n} \ell_{i,n},$$

and make the following assumption.

Assumption 7 (i) $K : \mathbb{R} \rightarrow [-1,1]$ satisfies $K(0) = 1$, and $K(z) = K(-z)$ for all $z \in \mathbb{R}$. (ii) $\frac{1}{E\ell_n} \sup_i E\left(\sum_{j \notin \mathscr{B}_{i,n}} K\left(\frac{\tilde{d}_{ij}}{d_n} \right) \right) = O(1)$, $\frac{1}{E\ell_n} \sup_i \sum_{j \notin \mathscr{B}_{i,n}} K\left(\frac{\tilde{d}_{ij}}{d_n} \right) = O_P(1)$. (iii) There exists a constant $C_{q_0} < \infty$ such that $\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} ||E\left(V_i V_j' \right)|| d_{ij}^{q_0} < C_{q_0}$, for all n, where q_0 denotes the Parzen characteristic exponent of $K(z)$. (iv) For all i, $\ell_{i,n} \leq cE\ell_n$, a.s., for some constant $c > 0$.

Assumption 7 imposes conditions on K and d_n which are similar to those imposed on K^* and d_n^* by Assumption 1 (i) and (ii), and Assumptions 3 and 5. The main difference is that we do not require the kernel function K to be positive definite. If necessary, we can ensure that $R\hat{Q}_n^{-1}\hat{J}_n\hat{Q}_n^{-1}R'$ (and its bootstrap analogue) is positive definite by replacing its negative eigenvalues by $\varepsilon > 0$, a small positive constant, as discussed in Remark 1.

Theorem 4.1 *Suppose Assumptions 1-7 hold. If $E^* |v_i|^4 < M$ and $d_n, E\ell_n \rightarrow \infty$ and $d_n^*, E\ell_n^* \rightarrow \infty$ as $n \rightarrow \infty$ such that $E\ell_n/n = o(1)$ and $E\ell_n^*/n^{1/2} = o(1)$, then, under H_0, as $n \rightarrow \infty$,*

$$\sup_{x \in \mathbb{R}} \left| P^*\left(\mathscr{W}_n^* \leq x \right) - P\left(\mathscr{W}_n \leq x \right) \right| = o_P(1).$$

Assumption 7 (along with our remaining assumptions) is used to show the consistency of $\hat{J}_n^*$ for $\hat{J}_{boot,n} = Var^*\left(n^{-1/2} \sum_{i=1}^{n} x_i u_i^* \right)$. See Lemma A.4. Since $\hat{J}_{boot,n}$ converges to J_n, this implies that $\hat{J}_n^*$ is consistent for J_n, which together with Theorem 3.1 imply the result.

Theorem 4.1 shows that the bootstrap Wald test $\mathscr{W}_n^*$ mimics the null distribution of $\mathscr{W}_n$ when the null is true, irrespective of whether we use the restricted or unrestricted approach. This is sufficient to claim the first order asymptotic validity of the bootstrap critical values under the null hypothesis. When the null is not true, the bootstrap distribution of the bootstrap Wald test based on the unrestricted residuals still converges to the null limiting distribution of $\mathscr{W}_n$. This result follows because (i) the bootstrap distribution of $\sqrt{n}\left(\hat{\beta}^* - \hat{\beta} \right)$ converges to a normal distribution with mean zero and variance-covariance equal to $C_n = Q^{-1}J_nQ^{-1}$, and (ii) $\hat{J}_n^*$ is consistent for J_n, independently of the true value of β underlying the DGP. For the restricted approach, we can show that the same is true when the true value of β is equal to $\beta_0 + \delta/\sqrt{n}$. Hence, the restricted bootstrap Wald test mimics the null limiting distribution of $\mathscr{W}_n$ under a set of local alternatives. This ensures that the bootstrap Wald test achieves the same local power as the test based on asymptotic critical values.

Next, we provide a description of the steps involved in testing $H_0 : R\beta = r_0$ using our bootstrap method.

Algorithm: spatial dependent wild bootstrap for testing $H_0 : R\beta = r_0$

(i) Compute

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i, \;\; \hat{u}_i = y_i - x_i'\hat{\beta}, \;\; \text{and } \hat{V}_i = x_i \hat{u}_i, i = 1, \ldots, n.$$

(ii) Compute $\mathscr{W}_n = \sqrt{n}\left(R\hat{\beta} - r_0\right)' \left[R\hat{Q}_n^{-1}\hat{J}_n\hat{Q}_n^{-1}R'\right]^{-1} \sqrt{n}\left(R\hat{\beta} - r_0\right)$, where

$$\hat{Q}_n = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' \quad \text{and} \quad \hat{J}_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n}\right)\hat{V}_i\hat{V}_j',$$

for a given kernel K, a bandwidth d_n, and a set of distances $\left\{\tilde{d}_{ij}\right\}$. Section 6 provides a method for choosing d_n.

(iii) Compute the matrix $\mathbb{K}_n^* = \left[K^*\left(\tilde{d}_{ij}/d_n^*\right)\right]_{i,j=1}^{n}$ and its eigendecomposition $\mathbb{K}_n^* = \Phi_n\Lambda_n\Phi_n'$, where Λ_n is a diagonal matrix with the nonnegative eigenvalues of $\mathbb{K}_n^*$ and the columns of Φ_n are the associated orthonormal eigenvectors. Let $L_n = \Phi_n\Lambda_n^{1/2}$.

(iv) Generate an $n \times 1$ vector $v \sim (0, I_n)$. We recommend using a sequence of standard normal random variables, but another distribution could be used. Then generate a sequence of random variables $\{\eta_i : i = 1, \ldots, n\}$ by multiplying this vector v by L_n :

$$\eta = L_n v.$$

(v) Compute the restricted OLS estimator

$$\tilde{\beta} = \hat{\beta} - \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} R' \left[R\left(\sum_{i=1}^{n} x_i x_i'\right)R'\right]^{-1}\left(R\hat{\beta} - r_0\right)$$

and the restricted residuals $\tilde{u}_i = y_i - x_i'\tilde{\beta}$, $i = 1, \ldots, n$. Let

$$y_i^* = x_i'\tilde{\beta} + u_i^*, \text{ where } u_i^* = \tilde{u}_i\eta_i.$$

(vi) Compute $\mathscr{W}_n^* = \sqrt{n}\left(R\hat{\beta}^* - r_0\right)' \left[R\hat{Q}_n^{-1}\hat{J}_n^*\hat{Q}_n^{-1}R'\right]^{-1} \sqrt{n}\left(R\hat{\beta}^* - r_0\right)$ using

$$\hat{\beta}^* = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i^*,$$

bootstrap scores $\hat{V}_i^* = x_i\hat{u}_i^*$, and

$$\hat{J}_n^* = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n}\right)\hat{V}_i^*\hat{V}_j^{*\prime}.$$

(vii) Repeat steps (iv)-(vi) B times and compute the bootstrap p-value as:

$$\hat{p}_n = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left(\mathscr{W}_n^{*(b)} > \mathscr{W}_n\right),$$

where $\mathscr{W}_n^{*(b)}$ is the bootstrap Wald statistic in replication b and $\mathbf{1}(\cdot)$ is the indicator function. Reject the null hypothesis if $\hat{p}_n$ is less than the chosen significance level of the test.

Finally, we note that the spatial dependent wild bootstrap based on the modified non-negative kernel $\mathbb{M}_n^*$ discussed in Section 3.3 is not asymptotically valid when applied to studentized test statistics. In particular, suppose we consider testing $H_0 : \beta = \beta_0$ in a simple location model $y_i = \beta + u_i$ by relying on a t-test

$$t_n = \frac{\sqrt{n}\left(\hat{\beta} - \beta_0\right)}{se\left(\hat{\beta}\right)},$$

where $\hat{\beta} = \bar{y} = n^{-1}\sum_{i=1}^{n} y_i$ and $se\left(\hat{\beta}\right)$ is the square root of

$$\hat{J}_n = \frac{1}{n}\sum_{i,j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n}\right)(y_i - \bar{y})(y_i - \bar{y}) \equiv \frac{1}{n}\sum_{i,j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n}\right)\hat{V}_i\hat{V}_j,$$

with $\hat{V}_i = \hat{u}_i \equiv y_i - \hat{\beta}$. Now, consider the spatial dependent bootstrap based on the modified kernel $\mathbb{M}_n^*$. Specifically, let $y_i^* = \hat{\beta} + u_i^*$, with $u_i^* = \hat{u}_i\eta_i$, and $\eta = L_n v$, where $v \sim N(0, I_n)$ and $L_n = \Phi_n\Lambda_n$, as described in Section 3.3. The bootstrap analog of t_n is

$$t_n^* = \frac{\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right)}{se\left(\hat{\beta}^*\right)},$$

where $se\left(\hat{\beta}^*\right)$ is the square root of

$$\hat{J}_n^* = \frac{1}{n}\sum_{i,j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n}\right)(y_i^* - \bar{y}^*)(y_i^* - \bar{y}^*) \equiv \frac{1}{n}\sum_{i,j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n}\right)\hat{V}_i^*\hat{V}_j^*,$$

with $\hat{V}_i^* = \hat{u}_i^* \equiv y_i^* - \hat{\beta}^*$. We can show that $\hat{V}_i^* = u_i^* - \bar{u}^* = v_1(z_i - \bar{z})$, where we let $z_i \equiv \hat{u}_i\sqrt{\lambda_1}\phi_{i1}$. With this notation, we can write

$$t_n^* \equiv \frac{\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right)}{\sqrt{\hat{J}_n^*}} = \frac{v_1}{\sqrt{v_1^2}} Z_n = a_1 Z_n,$$

where $a_1 = \frac{v_1}{\sqrt{v_1^2}}$ is a discrete (Rademacher) random variable given by

$$a_1 = \begin{cases} 1 \text{ with prob. } 1/2 \\ -1 \text{ with prob. } 1/2, \end{cases}$$

depending on the sign of v_1, and $Z_n = \left(\frac{1}{n}\sum_{i,j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n}\right)(z_i - \bar{z})(z_j - \bar{z})\right)^{-1/2} n^{-1/2}\sum_{i=1}^{n} z_i$ is a function of the original sample. Thus, conditionally on the data, t_n^* has a discrete distribution function given by

$$\hat{F}_n(x) \equiv P^*(t_n^* \leq x) = P^*(a_1 Z_n \leq x) = \frac{1}{2}\mathbf{1}(Z_n \leq x) + \frac{1}{2}\mathbf{1}(-Z_n \leq x).$$

Since $\hat{F}_n(x)$ is discrete, we cannot expect it to converge to $\Phi(x)$, the limiting distribution function of t_n. Although the consistency of the bootstrap distribution is not necessary for the validity of a bootstrap p-value (as recently emphasized by Cavaliere and Georgiev, 2020), we can also show that the bootstrap p-value induced by this modified procedure has a discrete distribution and hence it cannot be uniformly distributed, as $n \to \infty$.

5 Extension to nonlinear models

In this section, we describe an extension of the score wild bootstrap of Kline and Santos (2012a) to the spatial context. This method is a fast resampling method which can be used for inference in nonlinear models estimated by asymptotically linear estimators such as the QMLE or the GMM estimator. The main idea of the score wild bootstrap is to perturb the score vector evaluated at the estimated parameter of interest by an external random variable η_i with mean zero and variance one. Kline and Santos (2012a) assume that η_i is i.i.d., as in the regular wild bootstrap. Instead, here we allow for spatial dependence of unknown form in the score contributions and generate η_i such that $Cov^*\left(\eta_i, \eta_j\right) = K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)$. For brevity, we only provide a description of the method, omitting the proof of asymptotic validity[3].

Let $\{Z_i : i = 1, \ldots, n\}$ denote a sample of observations on a random vector $Z \in \mathbb{R}^{d_Z}$ and let $\theta \in \Theta \subseteq \mathbb{R}^p$. For the linear model considered previously, $Z_i = (y_i, x_i')'$, but this decomposition does not need to hold in general. The parameter of interest is denoted by θ_0 and we assume that its estimator $\hat{\theta}$ admits the following asymptotic linear expansion:

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = A_n(\theta_0) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i(\theta_0) + o_P(1),$$

where $A_n(\theta)$ is a $p \times \ell$ matrix and $V_i(\theta) \equiv V(Z_i, \theta)$ is an $\ell \times 1$ vector, where $\ell \geq p$. If

$$A_n(\theta_0) \xrightarrow{P} A_0 \text{ and } \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i(\theta_0) \xrightarrow{d} N(0, J_0),$$

it follows that

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N(0, C_0),$$

where $C_0 \equiv A_0 J_0 A_0'$.

Two examples of $\hat{\theta}$ that fit into this framework are the QMLE and the GMM estimator. For QMLE,

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} q(Z_i, \theta),$$

where $q(\cdot, \theta)$ is a quasi-log-likelihood real valued function. Assuming that q is twice differentiable in θ, we can write

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = -H_n^{-1}(\theta_0) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i(\theta_0) + o_p(1),$$

[3] This could be established under conditions similar to those used by Kim and Sun (2011) (who studied the spatial GMM estimator) by applying the bootstrap uniform law of large numbers and the bootstrap central limit theorems derived in Gonçalves and White (2004). Providing primitive conditions on the data such that a linear array representation holds for the score vector is difficult and case-specific.

where

$$H_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 q(Z_i,\theta)}{\partial\theta\partial\theta'},$$

is the $p\times p$ Hessian matrix and $V_i(\theta_0) = \frac{\partial q(Z_i,\theta_0)}{\partial\theta}$ is the $p\times 1$ score vector for observation i. Thus, we have that

$$A_n(\theta_0) = -H_n^{-1}(\theta_0) \ \text{ and } \ V_i(\theta_0) = \frac{\partial q(Z_i,\theta_0)}{\partial\theta}.$$

A special case is the OLS estimator considered previously, where $V_i(\theta_0) = -2x_iu_i$ and $H_n(\theta) = 2n^{-1}\sum_{i=1}^{n}x_ix_i'$.

For GMM estimators, assume that

$$E[g(Z_i,\theta_0)] = 0,$$

where $g : Z\times\Theta \longrightarrow \mathbb{R}^{\ell}$ contains the ℓ moment conditions with $\ell \geq p$. The GMM estimator of θ_0 is defined as

$$\hat{\theta} = \arg\min_{\theta\in\Theta}\left(\frac{1}{n}\sum_{i=1}^{n}g(Z_i,\theta)\right)' W_n\left(\frac{1}{n}\sum_{i=1}^{n}g(Z_i,\theta)\right),$$

where W_n is a random positive definite symmetric matrix which converges in probability to $W > 0$. Assuming g is differentiable in θ, we define the Jacobian matrix of the moment conditions as

$$G_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial g(Z_i,\theta)}{\partial\theta'}.$$

Under standard regularity conditions that allow for spatially dependent observations (see e.g. Kim and Sun (2011)),

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) = -\left(G_n(\theta_0)' W_n G_n(\theta_0)\right)^{-1} G_n(\theta_0)' W_n \frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_i(\theta_0) + o_p(1),$$

implying that

$$A_n(\theta_0) = -\left(G_n(\theta_0)' W_n G_n(\theta_0)\right)^{-1} G_n(\theta_0)' W_n, \text{ and } V_i(\theta_0) = g_i(\theta_0).$$

Following Kline and Santos (2012a), we can approximate the distribution of $\sqrt{n}\left(\hat{\theta}-\theta_0\right)$ by relying on the bootstrap distribution of

$$v_n^* = A_n\left(\hat{\theta}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i^*,$$

where $V_i^* = V_i\left(\hat{\theta}\right)\eta_i$ and η_i are such that $Cov^*\left(\eta_i,\eta_j\right) = K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)$.

This method contains the spatial dependent wild bootstrap for linear regressions as a special case. In particular, note that for the linear model,

$$A_n\left(\hat{\theta}\right) = \left(n^{-1}\sum_{i=1}^{n}x_ix_i'\right)^{-1} = \hat{Q}_n^{-1}, \text{ and } \ V_i^* = V_i\left(\hat{\theta}\right)\eta_i = x_i\hat{u}_i\eta_i \equiv x_iu_i^*,$$

where $u_i^* = \hat{u}_i\eta_i$. Since $y_i^* = x_i'\hat{\beta} + u_i^*$, this implies that

$$v_n^* = \hat{Q}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_iu_i^* = \hat{Q}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i\left(y_i^* - x_i'\hat{\beta}\right) = \sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right),$$

which shows that v_n^* is equal to $\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right)$, as we considered before.

For GMM, the spatial score bootstrap simulates the critical values of

$$v_n^* = -\left(G_n\left(\hat{\theta}\right)' W_n G_n\left(\hat{\theta}\right)\right)^{-1} G_n\left(\hat{\theta}\right)' W_n \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_i\left(\hat{\theta}\right) \eta_i,$$

in order to approximate the distribution of $\sqrt{n}\left(\hat{\theta} - \theta_0\right)$. Since v_n^* does not depend on any bootstrap GMM estimator $\hat{\theta}^*$, this is a fast resampling method that does not require any optimization in the bootstrap world.

6 Monte Carlo Simulations

In this section, we consider a simulation experiment to document the properties of our proposed approach. Our design follows Lee and Robinson (2016) and Sun and Kim (2015). The data is generated as:

$$y_i = \alpha + \beta x_i + u_i$$

where $\alpha = 0$ and $\beta = 1$. The observations lie within a square of dimension $\sqrt{n} \times \sqrt{n}$ where n is the sample size. The locations $s_i = (s_{i_1}, s_{i_2})$ are drawn uniformly within that square once and for all for each design, i.e. we let $s_{i_1} \sim U[0, \sqrt{n}]$ independently of $s_{i_2} \sim U[0, \sqrt{n}]$. The distance between observations at locations s_i and s_j is Euclidean:

$$d_{ij} = \sqrt{(s_{i_1} - s_{j_1})^2 + (s_{i_2} - s_{j_2})^2}.$$

Regressor and errors have the same dependence structure. Each of $z_i = (x_i, u_i)'$ and z_j are drawn from a standard normal $N(0, I_2)$ distribution but with correlation $\theta^{d_{ij}}$ between them. The parameter θ controls the degree of dependence among observations with a higher value of θ leading to observations that are more highly correlated. In our experiments, we consider values of θ between 0 and .9 in increments of .1. We report results for three sample sizes: $n = 25$, 100, and 400 as a function of θ with 10,000 replications.

We consider rejection rates of the null hypothesis of $\beta = 1$ against a two-sided alternative at the 5% level using the t statistic:

$$t_n = \frac{\sqrt{n}\left(\hat{\beta} - 1\right)}{se\left(\hat{\beta}\right)},$$

where $se\left(\hat{\beta}\right)$ is the square root of the $(2,2)$ element of $\hat{Q}_n^{-1} \hat{J}_n \hat{Q}_n^{-1}$ with

$$\hat{J}_n = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n}\right) \hat{V}_i \hat{V}_j',$$

using the Gaussian kernel, the data-based choice of d_n discussed below, and $\tilde{d}_{ij} = d_{ij}$ is the Euclidean distance between the locations s_i and s_j. We will later consider the presence of measurement error in locations and the misspecification of the distance measure.

The choice of bandwidth is clearly important. Our suggested approach is to compute an estimate of $C(d_0^{(k)})$, the spatial analog of the autocovariance function of the residuals for a set of potential bandwidths $\{d_0^{(k)}, k = 1,\ldots,M\}$, arranged in increasing order of magnitude. We estimate $C(d_0^{(k)})$ nonparametrically using a local average estimator $\hat{C}(d_0^{(k)})$, following Conley and Dupor (2001) and Conley and Topa (2002). We view $\hat{C}(d_0^{(k)})$ as a test statistic for the null hypothesis of spatial independence at distance $d_0^{(k)}$, and construct an acceptance region via a bootstrap simulation imposing this null, based on bootstrap datasets with i.i.d. draws from the empirical distribution of residuals. The bootstrap analog of $\hat{C}(d_0^{(k)})$ is then constructed for each spatially independent bootstrap sample, with their distribution providing a reference distribution for the null of spatial independence. We identify the first distance in the ordered set $\{d_0^{(k)}, k = 1,\ldots,M\}$ for which the bootstrap test does not reject independence and select the bandwidth as the previous value of that distance.

The next algorithm details the steps involved.

Algorithm: Bandwidth selection

(i) Compute

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i, \quad \hat{u}_i = y_i - x_i'\hat{\beta}, \quad \text{and } \hat{V}_i = x_i\hat{u}_i, i = 1,\ldots,n.$$

(ii) Choose a set of potential bandwidths $d_0 = \left(d_0^{(1)},...,d_0^{(M)}\right)'$, ordered in increasing magnitude, and tolerance level ε. For each $k = 1,\ldots,M$, compute $\hat{C}(d_0^{(k)})$, the covariance between residuals at distance $d_0^{(k)}$, via a uniform kernel regression with tolerance of ε :

$$\hat{C}(d_0^{(k)}) = \frac{1}{\sum_{i,j} 1(|\tilde{d}_{ij} - d_0^{(k)})| < \varepsilon)} \sum_{i,j} 1(|\tilde{d}_{ij} - d_0^{(k)})| < \varepsilon)\hat{u}_i\hat{u}_j'.$$

(iii) Generate data using an i.i.d. bootstrap from the empirical distribution of residuals $\hat{u}_i$ for each location s_i. For each $k = 1,\ldots,M$, compute the bootstrap analog of $\hat{C}(d_0^{(k)})$,

$$\hat{C}^*(d_0^{(k)}) = \frac{1}{\sum_{i,j} 1(|\tilde{d}_{ij} - d_0^{(k)})| < \varepsilon)} \sum_{i,j} 1(|\tilde{d}_{ij} - d_0^{(k)})| < \varepsilon)\hat{u}_i^*\hat{u}_j^{*\prime}.$$

(iv) Repeat step (iii) B times and obtain a bootstrap confidence interval for $C\left(d_0^{(k)}\right)$ using the 2.5% and 97.5% quantiles of $\hat{C}^*(d_0^{(k)})$.

(v) We search for the first element of d_0 for which $\hat{C}\left(d_0^{(k)}\right)$ is within these bands and set the bandwidth as the previous element of d_0.

We implement this algorithm as follows. We set $d_0^{(k)} = c_k \times n^{1/6}$, for c_k between .5 and 4 in increments of .5, which gives $M = 8$. From Kim and Sun (2011), the chosen rate of expansion is optimal in the MSE-sense when locations lie in a two-dimensional space using the Gaussian kernel. We also set the tolerance level ε as $.1 \times n^{\frac{1}{6}}$. These values were arrived at with some experimentation.

We compare the test statistic to 3 different critical values. The first one is the critical value from the standard normal distribution. The second critical value is obtained from the i.i.d. bootstrap which approximates the fixed-b asymptotic distribution in Bester et al. (2016) that assumes that the bandwidth d_n is a fixed proportion of the sample size. As in the time series context (see e.g. Kiefer and Vogelsang (2005)), the fixed-b asymptotic distribution in the spatial context is nuisance parameter free, but is highly nonstandard. In contrast to the time series case, it is a complicated functional of Brownian sheets and it depends on the sampling region. Thus, for practical purposes, it is hard to implement the fixed-b asymptotic critical values without resorting to the i.i.d. bootstrap. Finally, the last critical value is obtained using our spatial dependent wild bootstrap (SDWB) using the restricted residuals to obtain the bootstrap draws. We implement it using the same Gaussian kernel and bandwidth used to compute the t statistic in the sample. We use independent standard normal random variables as external draws and $B = 399$ bootstrap samples. The use of the Gaussian kernel with Euclidean distance ensures that the matrix $\mathbb{K}_n^*$ is positive semi-definite.

Figure 1 reports the rejection rates for the three sets of critical values under the null hypothesis of $\beta = 1$. The first thing to note from Figure 1 is that, as expected, size distortions increase with higher dependence (higher value of θ). Second, the use of asymptotic normal critical values leads to sizable size distortions for small values of n. For example, for $n = 25$, the rejection rate for $\theta = .5$ is 25.1% instead of 5%. This is reduced to 13.7% for $n = 400$. The i.i.d. bootstrap critical values perform much better and reduce the size distortions considerably. Again, for $\theta = .5$, the rejection rate is 12.7% for $n = 25$ and 7.8% for $n = 400$. Finally, the spatial dependent wild bootstrap gives rejection rates closer to the nominal level, 10.9% for $n = 25$ and 6.5% for $n = 400$.

Figures 2 to 5 explore measurement error in locations as in Lee and Robinson (2016). Hence, location i becomes:

$$\tilde{s}_i = \left(s_{i_1} + \zeta_i^1, s_{i_2} + \zeta_i^2\right)$$

where $\zeta_i = \left(\zeta_i^1, \zeta_i^2\right)'$ is independently drawn from $N(0, \zeta I_2)$ with $\zeta = 2$ (Figure 2), $\zeta = 4$ (Figure 3) or $\zeta = 10$ (Figure 4). In other words, the data is generated as above with the correct locations s_i and using the Euclidean distance in the data-generating process. However, when implementing all test procedures, we use the incorrect random distance $\tilde{d}_{ij} = \sqrt{(\tilde{s}_{i_1} - \tilde{s}_{j_1})^2 + (\tilde{s}_{i_2} - \tilde{s}_{j_2})^2}$ arising from the mismeasured locations.

Comparing Figures 1 and 2, measurement error in locations worsens all methods. For $n = 25$, the i.i.d. bootstrap performs best, but our SDWB dominates it for $n = 100$ and 400 though the difference is much smaller than when the correct locations are used. Making measurement error larger reinforces these findings as shown in Figures 3 and 4. If measurement error is large enough, distances convey no information and are discarded.

The results with asymptotic theory converge towards the White heteroskedasticity-robust standard errors, and our method deteriorates analogously. The i.i.d. bootstrap is obviously more robust since it relies less on the existence of a distance measure (it still relies on it in the construction of the statistic). Nonetheless, our SDWB performs better for larger sample size and stronger dependence.

Finally, Figure 5 considers the case where the wrong distance measure is used, This experiment generates data as before by drawing locations uniformly in a square of dimension $\sqrt{n} \times \sqrt{n}$, but the distance between observations is the maximum distance:

$$d_{ij} = \max\left[\left|s_{i_1} - s_{j_1}\right|, \left|s_{i_2} - s_{j_2}\right|\right].$$

We suppose that the researcher believes that he is in the same context as for Figure 1 and uses the Euclidean distance between observations. This is a different type of misspecification than what was considered in Figures 2-4. Here, the locations are correctly observed, but the wrong metric between them is used. Because the Euclidean distance and Gaussian kernel are used in sample, the resulting $\mathbb{K}_n^*$ matrix is still positive semi-definite.

When comparing Figures 1 and 5, we see that the misspecification worsens results for all methods, but the effect is quite small, much less than a $N(0,2)$ measurement error in locations. For example, with $\theta = .5$, our bootstrap has a rejection rate of 11.8% for $n = 25$, 8.4% with $n = 100$ and 6.6% for $n = 400$ compared to 10.9%, 8.0, and 6.5% with no measurement error and 20.0%, 18.7%, and 7.9% with $N(0,2)$ measurement error in locations.

We conclude from these experiments that the spatial dependent wild bootstrap removes a large fraction of the size distortions associated with the use of the normal asymptotic critical values. Its superiority is especially pronounced with stronger spatial dependence (larger values of θ). Moreover, it outperforms the i.i.d. bootstrap except for cases with large misspecification combined with either small sample size or weak dependence.

7 Empirical Example

In this section we present an example application to illustrate our method. This application's goal is to understand how firms are affected by import behavior in their local markets. An extensive empirical literature has examined the role of import competition in the reallocation of manufacturing within and across industries, e.g. Bernard et al. (2006), Autor et al. (2014), Acemoglu et al. (2016). Recent work in this literature such as Utar (2017) and Sandoval (2020), has been concerned with the distinct effects of imports depending on their location in the supply chain. This motivates an empirical investigation of the impact of different types of imports upon firm outcomes. We examine a regression that provides stylized facts about the correlations between firms' growth and the level of importing activity in their local markets, distinguishing between three types of imports. These three categories are: final goods imports which may reflect competition facing domestic producers in the local market, intermediate goods imports which could reflect, e.g. the scale of operation by competitors or the supply of inputs in the market, and capital goods imports which may reflect varying access to technology and/or competitors' scale of operation.

We use Canadian firm-level data from the National Accounts Longitudinal Microdata File (NALMF), constructed by Statistics Canada, for the years 2003 and 2007. The NALMF contains all incorporated firms in Canada, and is mainly used to track GDP and employment of firms, and their locations. We use data from 2003 and 2007 and link wholesaler import data from Statistics Canada to the NALMF.[4] This provides data on firm-level imports that include their value, country of origin and product at the level of a ten-digit Harmonized System code. These import-linked data allow us to study how the import activity of Canadian wholesalers in intermediate, final, and capital good markets affect manufacturing firms' outcomes.[5]

Specifically, we examine the relationship between manufacturing firms' sales growth and the level of exposure to import activity in their local markets. We estimate that relation using a cross-section of firms and the following specification:

$$\text{Sales Growth}_i = \alpha + \theta X_i^{Final} + \gamma X_i^{Intermediate} + \delta X_i^{Capital} + Z_i' \psi + \varepsilon_i. \tag{15}$$

where i corresponds to a manufacturing firm. The dependent variable, Sales Growth$_i$, corresponds to the growth rate of real sales between 2003 and 2007. The local market of firm i is taken to be its Economic Region (ECR) among the 72 ECRs defined by Statistics Canada.[6] Importing activity variables X are defined at the ECR level and reflect 2003 activity. We define the parameter vector as $\beta = (\alpha, \theta, \gamma, \delta, \psi')'$.

X_i^{Final} is computed as a ratio. Its numerator is the value of all imports by wholesalers of final goods within firm $i's$ ECR. Its denominator is the total value of all imports and domestic sales of manufacturing firms in firm $i's$ ECR. The import measures $X_i^{Intermediate}$ and $X_i^{Capital}$ are defined analogously. See Sandoval (2020) for an extensive discussion of the merits of these particular measures of import activity.

Our conditioning information in Z_i includes 2003 data on firm age, the logarithm of real sales, and measures of capital and skill intensity. Following Bernard et al. (2006) we measure capital intensity as the natural log of a capital/labor ratio using the book value of tangible assets divided by firm's total payroll, and measure skill intensity as the ratio of the total payroll to the payments to production workers. We focus on a cross-section of manufacturing firms with more than 20 workers in 2003, yielding a sample of 6120 firms. Approximately 88% of Canadian manufacturing workers in 2003 worked in these sample firms.

We anticipate that spatial dependence will be present in this cross section of firms due to two main factors.

[4] See Sandoval (2020) for a detailed description of these linked data.

[5] We classify imports according to their end-use as intermediate, final or capital goods using the correspondence tables between the HS and Broad Economic Categories (BEC) classification.

[6] Approximately 96% of the firms in our sample have single locations; for the remaining firms, we take a firm's location as the location of its designated headquarters.

Firms that are close in terms of travel time will have relevant local markets that overlap. When firms' local markets overlap they will tend to face correlated shocks, e.g. labor supply shocks. We use physical distance between firms as our measure of the overlap between firms' local markets. Correlated unobservables could also easily arise due to similarities in firms' technology making them vulnerable to a common set of shocks or changes to their technology. We represent firms' technology via two characteristics that we use to generate a 'technology distance': their capital to labor ratio and the fraction of total payroll going to production workers.

We combine our two distance measures to implement the spatial dependent wild bootstrap. For physical distance, we use the coordinates of the centroids of the ECR in which firms are located as firm coordinates and use straight line distance as our measure of firms' physical distance.[7] Each firm's technology is summarized by a two-dimensional vector containing its capital/labor ratio and ratio of total payroll to production worker payroll, both in 2003. Technology distance is calculated as the Euclidean distance between firms' two-dimensional technology characteristics vectors.

We add both distance measures after scaling them so neither dominates. The combined distance for firms i and j , $\tilde{d}_{ij}$, is constructed by adding a scaled multiple of their technology distance to their physical distance:

$$\tilde{d}_{ij} = \text{physical distance} + \text{scale} \times (\text{technology distance}).$$

The technology scale factor is constructed so that the median of the scaled technology distance is equal to the median physical distance (560 km). Thus for two firms with identical measured technology, our combined distance $\tilde{d}_{ij}$ is equal to physical distance, providing at least some sense of units.

We use the same kernel and bandwidth for the spatial dependent wild bootstrap procedure and spatial HAC:

$$K^*(\tilde{d}_{ij}/d_n) = K(\tilde{d}_{ij}/d_n) = exp(-(\tilde{d}_{ij}/d_n)^2), \text{ where } d_n = d_n^*.$$

We present results in Table 1 using a methodology for choosing d_n similar to that described in Section 6. Specifically, we obtain $\hat{C}(d_0^{(k)})$ for a range of values for $d_0^{(k)}$. Figure 6 gives these estimates for an assortment of distances, along with upper and lower ends of a 90% acceptance region for spatial independence. $\hat{C}(d_0^{(k))})$ is normalized in Figure 6 by dividing by the sample variance of residuals. Although spatial covariances can be small relative to the variance of residuals, it is important to note that there can be a very large number of firms at the smaller distances from each other so their covariances' sum can still be substantial relative to their variance. Pointwise hypothesis tests for spatial independence can be done by simply comparing the covariance estimates (circles) to the acceptance region (between dashes). [8] Estimated autocovariances are generally decreasing and

[7]Centroids are calculated from Statistics Canada maps of 2016 ECR boundaries.

[8]It is important to remember that these acceptance regions are for the null hypothesis of independence rather than zero correlation. Rejections can occur due to differences in the sampling distribution under dependence vs independence, even if the covariance at the given distance were zero. We anticipate that with spatial dependence rejections at smaller distances will occur largely due to nonzero autocovariances. At larger distances we anticipate some (correct) rejections even though true covariance is zero at that distance due to the sampling variability being larger under dependence than independence.

Table 1: Inference for regression (14) predicting the growth of sales between 2003 and 2007

	$\hat{\beta}$	Half width of 95% confidence intervals							
		OLS	White	Cluster ECR	Clustered Wild Bootstrap ECR	Cluster Industry	Clustered Wild Bootstrap Industry	Spatial HAC	Spatial Wild Boot-strap
Import Pen. Intermediate	4.342	1.666	1.799	1.833	2.048	1.696	1.746	1.838	2.074
Import Pen. Final	-3.033	1.584	1.454	2.164	1.465	1.344	2.063	1.557	1.695
Import Pen. Capital	-0.130	1.307	1.027	1.271	1.119	1.199	1.061	1.013	1.021
Log real sales	0.013	0.017	0.016	0.014	0.010	0.031	0.019	0.021	0.023
Age	-0.003	0.004	0.004	0.003	0.003	0.004	0.004	0.004	0.005
Capital intensity	0.028	0.021	0.030	0.032	0.034	0.037	0.028	0.028	0.031
Skill intensity (100s)	0.004	0.051	0.023	0.023	0.012	0.025	0.011	0.023	0.013
Constant	-0.265	0.191	0.197	0.192	0.200	0.455	0.266	0.280	0.311

Notes: the import penetration variables are computed at the ECR-level and for the year 2003. The reminder of the regressors refer to firm-level data for the year 2003. For the Wild bootstrap we compute a symmetric t-percentile confidence interval using 2,000 Bootstrap repetitions.

independence is rejected until about 500-700 units and then 'borderline rejected' until about 1100 units. This motivated our choice of a bandwidth of 560 for our reported estimates; the implied weight K is greater than .14 for only 25% of the pairs of firms. We obtained qualitatively very similar results with bandwidth choices up to 1120; at this bandwidth K is greater than .14 for 53% of firm pairs and K is greater than .61 for 25% of pairs.

We also include in Table 1 confidence intervals using different methods for computing standard errors and critical values. Specifically, we compute asymptotic theory-based intervals that rely on different standard errors: classical OLS, heteroskedasticity consistent (labelled White), clustered at the ECR-level, clustered at the 3-digit industry-level, and spatial HAC. We also include three bootstrap-based intervals: cluster wild bootstrap at the ECR level, cluster wild bootstrap at the industry level and our spatial dependent wild bootstrap using the same kernel and bandwidth as in the spatial HAC.[9]

The results presented in Table 1 have several key features. The relative sizes of confidence intervals across methods differ across elements of β. For example, for the first element of β, the coefficient on intermediate goods imports, our spatial wild bootstrap CIs are the widest but for the coefficient for final goods imports, Cls using clustering on ECR are largest.

There is evidence of substantial dependence as a function of physical distance, but its impact on inference again varies across elements of β. This can be seen by comparing e.g. White CIs with ECR cluster for the capital imports coefficient where the length of CIs differ by 24%, but for the intermediate goods coefficient, these CIs are nearly the same length. There is also some evidence of correlations due to similar technology, which are partly reflected in CIs under Clustered by Industry. CIs using Industry clusters are sometimes larger and sometimes smaller than White CIs across the three coefficients of interest. Both the spatial HAC and spatial

[9] We do not compare to blocking/clustering methods allowing for general dependence structures as in Bester, Conley, and Hansen (2012) or Ibragimov and Mueller (2010) due to the difficulties in defining appropriate blocks/clusters when dependence is characterized by multiple metrics.

dependent wild bootstrap attempt to allow for both types of correlation.

Finally, there is evidence that using the spatial dependent wild bootstrap matters relative to a spatial HAC estimator using the same kernel. Across all parameters the difference in length of the CIs is typically about 10% to 20%, possibly reflecting the greater robustness of the bootstrap intervals to finite sample deviations from the normal distribution.

8 Conclusion

This paper has proposed a method for generating bootstrap data under spatial and space-time dependence of unknown form. It is implemented by multiplying a vector of external variables by the eigendecomposition of a bootstrap kernel. The wild bootstrap and wild cluster bootstrap are special cases of this approach and do not require the decomposition of a full $n \times n$ matrix, but our method can also be used to generate data with dependence patterns for which no alternative method exists. Simulation experiments suggest that there are gains from generating bootstrap samples that replicate the spatial patterns in the data.

Results by Zhang and Shao (2013) show that the Gaussian dependent bootstrap is second order correct for the Gaussian location model under fixed-b asymptotics. Extending these results to the regression model with non-Gaussian errors and spatial dependence is an open but challenging topic for future research (see Kline and Santos (2012b) for results for the standard wild bootstrap in the i.i.d. context).

Another interesting extension of our results would be to investigate the properties of the spatial dependent wild bootstrap when locations are randomly selected from a given population rather then being fixed, as we have assumed here. This setup has been recently considered by Müller and Watson (2021), who propose a new estimator of the spatial long run variance using a fixed number of principal components obtained from a "worst-case" benchmark parametric model for the covariance structure of the error term in a location model. A critical value is then constructed using this benchmark model under the additional assumption of Gaussianity so as to ensure that the size of the resulting test is asymptotically correct. Contrary to our setup, Müller and Watson (2021) assume locations to be randomly selected from a density g and show that allowing for a nonuniform density has implications for the conventional fixed-b limiting distributions. The choice of sampling design may also impact bootstrap validity (see e.g. Lahiri and Zhu (2006) for results on the spatial block bootstrap). Interestingly, Shao (2010) shows that the wild dependent bootstrap is asymptotically valid when the sampling design is stochastic with a potentially nonuniform spatial density g. It would be interesting to extend these results to our context, where spatial dependence is not restricted to be indexed on the line.

A Appendix

As usual in the bootstrap literature, we use P^* to denote the bootstrap probability measure, conditional on the original sample (defined on a given probability space $(\Omega, \mathscr{F}, P)$). For any bootstrap statistic T_n^*, we write $T_n^* = o_{P^*}(1)$, in prob-P, or $T_n^* \to^{P^*} 0$, in prob-P, when for any $\delta > 0$, $P^*(|T_n^*| > \delta) = o_P(1)$. We write $T_n^* = O_{P^*}(1)$, in prob-P, when for all $\delta > 0$ there exists $M_\delta < \infty$ such that $\lim_{n\to\infty} P[P^*(|T_n^*| > M_\delta) > \delta] = 0$. By Markov's inequality, this follows if $E^*|T_n^*|^q = O_P(1)$ for some $q > 0$. Finally, we write $T_n^* \to^{d^*} D$, in probability, if conditional on a sample with probability that converges to one, T_n^* weakly converges to the distribution D under P^*, i.e. $E^*(f(T_n^*)) \to^P E(f(D))$ for all bounded and uniformly continuous functions f.

A.1 Auxiliary lemmas

Define

$$J_{boot,n} = Var^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i \eta_i\right)$$

and note that $J_{boot,n}$ differs from $\hat{J}_{boot,n} = Var^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{V}_i \eta_i\right)$ by replacing $\hat{V}_i$ with V_i. Recall that $\tilde{d}_{ij} = d_{ij} + \xi_{ij}$, where d_{ij} is deterministic, and ξ_{ij} is a measurement error which is independent of $\{e_l\}$ and $\{x_i\}$. Let $\Psi_n = \{\xi_{ij}, i, j = 1, ..., n\}$.

Lemma A.1 *Under Assumptions 1(i) and (ii), we have*

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n E\left|K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right| = O\left(\frac{E\ell_n^*}{n}\right)$$

as $E\ell_n^*, d_n^* \to \infty$ *such that* $E\ell_n^*/n \to 0$.

Proof of Lemma A.1. Note that

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \left|K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right| = \frac{1}{n^2}\sum_{i=1}^n\sum_{j\in\mathscr{B}_{i,n}^*}^n \left|K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right| + \frac{1}{n^2}\sum_{i=1}^n\sum_{j\notin\mathscr{B}_{i,n}^*}^n \left|K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right|,$$

where $\mathscr{B}_{i,n}^*$ is a random set containing the neighbors of i using $\tilde{d}_{ij}$. Thus, we have that

$$\frac{1}{n^2}\sum_{i=1}^n E\sum_{j=1}^n \left|K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right| = \frac{1}{n^2}\sum_{i=1}^n E\left(\sum_{j\in\mathscr{B}_{i,n}^*}^n \left|K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right|\right) + \frac{1}{n^2}\sum_{i=1}^n E\left(\sum_{j\notin\mathscr{B}_{i,n}^*}^n \left|K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right|\right),$$

If K^* truncates, the second term is automatically zero, whereas the first term can be bounded by $n^{-2}E\sum_{i=1}^n\sum_{j\in\mathscr{B}_{i,n}^*}^n 1 = n^{-1}E\left(n^{-1}\sum_{i=1}^n \ell_{i,n}^*\right) = n^{-1}E\ell_n^*$. When K^* does not truncate, then we use Assumption 1(ii) to bound the second term. Specifically, we obtain that

$$\frac{1}{n^2}\sum_{i=1}^n E\sum_{j\notin\mathscr{B}_{i,n}^*}^n \left|K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right| \le \frac{E\ell_n^*}{n}\left(\frac{1}{E\ell_n^*}\sup_i E\sum_{j\notin\mathscr{B}_{i,n}^*}^n \left|K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right|\right) = O\left(\frac{E\ell_n^*}{n}\right).$$

The following lemma establishes the consistency of $J_{boot,n}$ and $\hat{J}_{boot,n}$ towards J_n. This is a key result for proving the asymptotic validity of the bootstrap distribution of $\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right)$ and the corresponding Wald test $\mathscr{W}_n^*$.

Lemma A.2 *Suppose that the conditions of Theorem 3.1 hold. Then (i) $J_{boot,n} - J_n \to^P 0$ and (ii) $\hat{J}_{boot,n} - J_n \to^P 0$.*

Our next result is an auxiliary result used to prove Theorem 3.1.

Lemma A.3 *Suppose Assumptions 1 and 2 hold. Then, for any pair (i,j), conditionally on $\Psi_n = \{\xi_{ij}\}$,*

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left|E\left(V_iV_j'|\Psi_n\right)\phi_{ik}\phi_{jk}\right| \leq M,$$

uniformly in $k = 1,\ldots,n$, where ϕ_{ik} is the i^{th} element of $\phi_k = (\phi_{1k},\ldots,\phi_{nk})'$, the k^{th} eigenvector of $\mathbb{K}_n^* = \left(K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right)_{i,j=1,\ldots,n}$, and where the constant M is independent of Ψ_n.

Proof of Lemma A.2. Part (i) Since $J_{boot,n} - J_n \to^P 0$ if and only if $\alpha' J_{boot,n}\alpha - \alpha' J_n \alpha \to^P 0$ for any $p \times 1$ vector α, we consider the case that $J_{boot,n}$ and J_n are scalars without loss of generality. Write

$$J_{boot,n} - J_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\left[V_iV_j - E\left(V_iV_j\right)\right] + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right) - 1\right)E\left(V_iV_j\right) \equiv b_1 + b_2.$$

For b_1, note that by the law of iterated expectations and the independence between $\{V_i\}$ and $\Psi_n \equiv \{\xi_{ij}\}$,

$$E\left(b_1\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left[K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)E\left(V_iV_j - E\left(V_iV_j\right)|\Psi_n\right)\right] = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left[K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\left(E\left(V_iV_j|\Psi_n\right) - E\left(V_iV_j\right)\right)\right] = 0$$

where $E\left(V_iV_j|\Psi_n\right) = E\left(V_iV_j\right)$ given Assumption 4 (i). Hence, it suffices to prove that $Var\left(b_1\right) = o(1)$. We have

$$\begin{aligned} Var\left(b_1\right) &= Var\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\left[V_iV_j - E\left(V_iV_j\right)\right]\right) \\ &= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n} E\left[K^*\left(\frac{\tilde{d}_{i_1j_1}}{d_n^*}\right)K^*\left(\frac{\tilde{d}_{i_2j_2}}{d_n^*}\right)\right]\left[E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right) - E\left(V_{i_1}V_{j_1}\right)E\left(V_{i_2}V_{j_2}\right)\right], \end{aligned} \tag{16}$$

using again the law of iterated expectations and the independence assumption between $\{V_i\}$ and Ψ_n. Adding and subtracting appropriately in (16), we can bound $Var\left(b_1\right)$ by

$$Var\left(b_1\right) \leq b_{11} + b_{12} + b_{13}, \text{ where}$$

$$b_{11} = \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}\left|E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right) - E\left(V_{i_1}V_{j_1}\right)E\left(V_{i_2}V_{j_2}\right) - E\left(V_{i_1}V_{i_2}\right)E\left(V_{j_1}V_{j_2}\right) - E\left(V_{i_1}V_{j_2}\right)E\left(V_{j_1}V_{i_2}\right)\right|$$

and

$$
\begin{aligned}
b_{12} &= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n} E\left[K^*\left(\frac{\tilde{d}_{i_1j_1}}{d_n^*}\right)K^*\left(\frac{\tilde{d}_{i_2j_2}}{d_n^*}\right)\right]\left|E\left(V_{i_1}V_{i_2}\right)E\left(V_{j_1}V_{j_2}\right)\right| \\
b_{13} &= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n} E\left[K^*\left(\frac{\tilde{d}_{i_1j_1}}{d_n^*}\right)K^*\left(\frac{\tilde{d}_{i_2j_2}}{d_n^*}\right)\right]\left|E\left(V_{i_1}V_{j_2}\right)E\left(V_{j_1}V_{i_2}\right)\right|.
\end{aligned}
$$

We can show that b_{12} and b_{13} are both of order $O\left(\frac{E\ell_n^*}{n}\right)$, whereas $b_{11} = O\left(\frac{1}{n}\right)$. Thus, $Var\left(b_1\right) = o\left(1\right)$ under our assumptions provided $\frac{E\ell_n^*}{n} = o\left(1\right)$. Next, we focus on the term b_{12} (the argument for b_{13} is the same and the proof that $b_{11} = O\left(\frac{1}{n}\right)$ follows by an argument similar to the one used to show that $C_1 = O\left(1\right)$ in the proof of Theorem 3.1, so we omit the details here). We can write

$$
\begin{aligned}
b_{12} &\leq \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n} E\left|K^*\left(\frac{\tilde{d}_{i_1j_1}}{d_n^*}\right)\right|\sum_{i_2=1}^{n}\left|E\left(V_{i_1}V_{i_2}\right)\right|\sum_{j_2=1}^{n}\left|E\left(V_{j_1}V_{j_2}\right)\right| \\
&\leq \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n} E\left|K^*\left(\frac{\tilde{d}_{i_1j_1}}{d_n^*}\right)\right|\underbrace{\left(\sup_c\sum_{i_2=1}^{n}\left|E\left(V_cV_{i_2}\right)\right|\right)}_{\leq\Delta}\underbrace{\left(\sup_c\sum_{j_2=1}^{n}\left|E\left(V_cV_{j_2}\right)\right|\right)}_{\leq\Delta} \\
&\leq \Delta^2\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n} E\left|K^*\left(\frac{\tilde{d}_{i_1j_1}}{d_n^*}\right)\right| = O\left(\frac{E\ell_n^*}{n}\right) = o\left(1\right),
\end{aligned}
$$

if $\frac{E\ell_n^*}{n} \to 0$ by Lemma A.1. Note that we have used the fact that $\sup_i\sum_{j=1}^{n}\left|E\left(V_iV_j\right)\right| \leq \Delta$ under Assumption 2.

For b_2, note that

$$
\left|b_2\right| \leq \frac{1}{\left(d_n^*\right)^{q_0^*}}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|E\left(V_iV_j\right)\right\|\tilde{d}_{ij}^{q_0^*}\left(K_{q_0^*}^* + o\left(1\right)\right) = O_P\left(\frac{1}{\left(d_n^*\right)^{q_0^*}}\right) \text{ as } d_n^* \to \infty,
$$

because

$$
P\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|E\left(V_iV_j\right)\right\|\tilde{d}_{ij}^{q_0^*} > \Delta\right) \leq \frac{1}{\Delta}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|E\left(V_iV_j\right)\right\|E\left(\tilde{d}_{ij}^{q_0^*}\right) \to 0 \text{ as } \Delta \to \infty
$$

given Assumption 4. Hence, $b_2 = o_P\left(1\right)$, completing the proof of part (i).

For part (ii), given (i) it suffices to show that $\hat{J}_{boot,n} - J_{boot,n} = o_P\left(1\right)$. Since $x_i\hat{u}_i = x_i\left[u_i + x_i\left(\beta - \hat{\beta}\right)\right]$, we can write

$$
\hat{J}_{boot,n} - J_{boot,n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)x_ix_j\left(\hat{u}_i\hat{u}_j - u_iu_j\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)x_ix_j\left[x_ix_j\left(\beta-\hat{\beta}\right)^2 + 2x_ju_i\left(\beta-\hat{\beta}\right)\right] \equiv c_1 + c_2.
$$

Because $\hat{\beta} - \beta = O_P\left(n^{-1/2}\right)$,

$$
c_1 = O_P\left(1\right)\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)x_i^2x_j^2\right) = O_P\left(\frac{E\ell_n^*}{n}\right),
$$

because by Markov's inequality

$$
P\left(\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)x_i^2x_j^2\right| > \Delta\right) \leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left|K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right|\underbrace{E\left(x_i^2x_j^2\right)}_{\leq M} \leq \frac{M}{\Delta}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left|K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right| = O\left(\frac{E\ell_n^*}{n}\right)
$$

under Lemma A.1. For c_2,

$$c_2 = \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right) x_i u_i x_j^2 \left(\beta - \hat{\beta}\right) = O_P(1)\left(\frac{1}{n\sqrt{n}}\sum_{j=1}^{n} x_j^2 \sum_{i=1}^{n} K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right) x_i u_i\right).$$

We have

$$\begin{aligned}
|c_2| &\leq O_P(1)\frac{1}{n}\sum_{j=1}^{n} x_j^2 \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right) x_i u_i\right| \\
&\leq O_P(1)\left(\frac{1}{n}\sum_{j=1}^{n} x_j^4\right)^{1/2}\left(\frac{1}{n}\sum_{j=1}^{n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right) x_i u_i\right|^2\right)^{1/2} = O_P\left(\sqrt{\frac{E\ell_n^*}{n}}\right),
\end{aligned}$$

because

$$\begin{aligned}
&P\left(\frac{1}{n}\sum_{j=1}^{n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right) x_i u_i\right|^2 > \Delta\right) \leq \frac{1}{\Delta}\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\left|K^*\left(\frac{\tilde{d}_{i_1 j}}{d_n^*}\right)K^*\left(\frac{\tilde{d}_{i_2 j}}{d_n^*}\right)\right|\left|E\left[V_{i_1}V_{i_2}\right]\right| \\
\leq\ & \frac{1}{\Delta}\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i_1=1}^{n} E\left|K^*\left(\frac{\tilde{d}_{i_1 j}}{d_n^*}\right)\right|\underbrace{\sup_c \sum_{i_2=1}^{n}\left|E\left[V_c V_{i_2}\right]\right|}_{\leq M} = O\left(\frac{E\ell_n^*}{n}\right),
\end{aligned}$$

by Lemma A.1. Therefore, $\hat{J}_{boot,n} - J_{boot,n} = o_P(1)$ under the rate condition on $E\ell_n^*$, which concludes the proof.

Proof of Lemma A.3. The proof uses the weak dependence of V_i and the fact that $\sum_{i=1}^{n}\phi_{ik}^2 = 1$ for any realization of $\mathbb{K}_n^*$. Let's rearrange the sequence of $\{\phi_{ik}, i = 1, \ldots, n\}$ as $\left\{\phi_k^{(a)}, a = 1, \ldots, n\right\}$ for each k in a way that $\left|\phi_k^{(a)}\right|$ is the a-th largest component among $\{|\phi_{ik}|, i = 1, \ldots, n\}$. That is, $\left|\phi_k^{(1)}\right| \geq \left|\phi_k^{(2)}\right| \geq \ldots \geq \left|\phi_k^{(n)}\right|$. Using this, conditionally on Ψ_n, we can rewrite

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|E\left(V_i V_j'\right)\phi_{ik}\phi_{jk}\right\| = \sum_{a=1}^{n}\sum_{b=1}^{n}\left\|E\left(V_a V_b'\right)\phi_k^{(a)}\phi_k^{(b)}\right\| = \underbrace{\sum_{a=1}^{n}\left\|E\left(V_a V_a'\right)\left(\phi_k^{(a)}\right)^2\right\|}_{(1)\text{: diagonal part}} + 2\underbrace{\sum_{a=1}^{n-1}\sum_{b=a+1}^{n}\left\|E\left(V_a V_b'\right)\phi_k^{(a)}\phi_k^{(b)}\right\|}_{(2)},$$

where we have used the independence of $\{V_i\}$ and Ψ_n.

It follows that

$$(1) = \sum_{a=1}^{n}\left\|E\left(V_a V_a'\right)\right\|\left(\phi_k^{(a)}\right)^2 \leq \sup_c \left\|E\left(V_c V_c'\right)\right\|\sum_{a=1}^{n}\left(\phi_k^{(a)}\right)^2 = \sup_c\left\|E\left(V_c V_c'\right)\right\| < \infty,$$

since $\sum_{a=1}^{n}\left(\phi_k^{(a)}\right)^2 = 1$. Similarly,

$$(2) = \sum_{a=1}^{n-1}\sum_{b=a+1}^{n}\left\|E\left(V_a V_b'\right)\right\|\left|\phi_k^{(a)}\right|\left|\phi_k^{(b)}\right| \leq \sum_{a=1}^{n-1}\left|\phi_k^{(a)}\right|\sum_{b=a+1}^{n}\left\|E\left(V_a V_b'\right)\right\|\left|\phi_k^{(b)}\right| \leq \sum_{a=1}^{n-1}\left|\phi_k^{(a)}\right|\sum_{b=a+1}^{n}\left\|E\left(V_a V_b'\right)\right\|\left|\phi_k^{(a)}\right|,$$

where the second inequality is due to $\left|\phi_k^{(a)}\right| \geq \left|\phi_k^{(b)}\right|$ with $a < b$. Then,

$$(2) \leq \sum_{a=1}^{n-1}\left(\phi_k^{(a)}\right)^2\sum_{b=a+1}^{n}\left\|E\left(V_a V_b'\right)\right\| \leq \sum_{a=1}^{n}\left(\phi_k^{(a)}\right)^2\left(\sup_c \sum_{b=1}^{n}\left\|E\left(V_c V_b'\right)\right\|\right) = \left(\sup_c \sum_{b=1}^{n}\left\|E\left(V_c V_b'\right)\right\|\right) < \infty.$$

assuming that the term in parenthesis is bounded. This last condition is slightly stronger than the usual weak dependence $\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|E\left(V_i V_j'\right)\right\| < \infty$, but it holds under Assumption 2.

A.2 Proof of main results in the paper

Proof of Theorem 3.1. Let $C_n = Q^{-1} J_n Q^{-1}$ and define its square root matrix as $C_n^{1/2} = Q^{-1} J_n^{1/2}$, where $J_n^{1/2}$ is such that $J_n^{1/2}\left(J_n^{1/2}\right)' = J_n$ and it exists by assumption. It follows that $C_n^{-1/2} = J_n^{-1/2} Q$ and

$$C_n^{-1/2}\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right) = C_n^{-1/2}\hat{Q}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i u_i^* = J_n^{-1/2} Q \hat{Q}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i u_i^* = J_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i u_i^* + o_{P^*}(1),$$

since under Assumption 6 (ii), $\hat{Q}_n \to^P Q$. Thus, it suffices to show that

$$J_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i u_i^* \to^{d^*} N(0, I_p), \text{ in prob-}P, \tag{17}$$

to conclude that

$$C_n^{-1/2}\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right) \to^{d^*} N(0, I_p), \text{ in prob-}P. \tag{18}$$

Given that $C_n^{-1/2}\sqrt{n}\left(\hat{\beta} - \beta\right) \to^d N(0, I_p)$ under our assumptions, (18) implies the result by Polya's Theorem and the continuity of the normal distribution. Using the definition of u_i^*, (17) follows if

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i(\hat{u}_i - u_i)\eta_i \to^{p^*} 0, \text{ in prob-}P, \text{ and} \tag{19}$$

$$J_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i u_i \eta_i \to^{d^*} N(0, I_p), \text{ in prob-}P, \tag{20}$$

as $n \to \infty$. For (19), we note that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i(\hat{u}_i - u_i)\eta_i = \underbrace{\frac{1}{n}\sum_{i=1}^n x_i x_i' \eta_i}_{=a_1}\underbrace{\sqrt{n}\left(\beta - \hat{\beta}\right)}_{=O_P(1)},$$

so it suffices to show that $a_1 = o_{P^*}(1)$ in prob-P. By Markov's inequality, this follows if $E^*|a_1|^2 = o_P(1)$. Routine calculations show that

$$E^*|a_1|^2 = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n tr\left(x_i x_i' x_j x_j'\right)E^*\left(\eta_i \eta_j\right) = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \left|tr\left(x_i x_i' x_j x_j'\right)\right| K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right) = O_P\left(\frac{E\ell_n^*}{n}\right) \to 0, \tag{21}$$

as $\frac{E\ell_n^*}{n} \to 0$, given Lemma A.1.

Next, we prove (20). Given Assumption 1, $\mathbb{K}_n^*$ is symmetric and positive semi-definite, which implies that $\mathbb{K}_n^* = \Phi_n \Lambda_n \Phi_n'$, where Λ_n is a diagonal matrix with the nonnegative eigenvalues of $\mathbb{K}_n^*$ and the columns of Φ_n are the associated orthonormal eigenvectors. Then, L_n can be written as

$$L_n = \Phi_n \Lambda_n^{1/2} = \left[\lambda_1^{1/2}\phi_1, \ldots, \lambda_n^{1/2}\phi_n\right],$$

implying that

$$\eta = \Phi_n \Lambda_n^{1/2} v = \left[\lambda_1^{1/2}\phi_1, \ldots, \lambda_n^{1/2}\phi_n\right] v,$$

where $v \sim$ i.i.d.$(0, I_n)$. Given that $V_i = x_i u_i$ is $p \times 1$, let

$$\underset{p\times n}{V'} = \left[\begin{array}{ccc} V_1 & \dots & V_n \end{array}\right].$$

It follows that

$$J_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i u_i \eta_i = J_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i \eta_i = J_n^{-1/2}\frac{1}{\sqrt{n}}V'\eta = \frac{1}{\sqrt{n}}\sum_{k=1}^n \left(J_n^{-1/2}\lambda_k^{1/2}V'\phi_k\right)v_k = \frac{1}{\sqrt{n}}\sum_{k=1}^n Z_k^*,$$

where by definition,

$$Z_k^* = J_n^{-1/2}\left(\lambda_k^{1/2}V'\phi_k\right)v_k.$$

Note that $\left(\lambda_k^{1/2}J_n^{-1/2}V'\phi_k\right)$ is a $p \times 1$ vector of constants conditional on the data and that $v_k \sim$i.i.d.$(0,1)$, which implies that Z_k^* is an independent heterogeneous array. We will show that $n^{-1/2}\sum_{k=1}^n Z_k^* \to^{d^*} N(0, I_p)$, in prob-P, by applying Lyapunov's CLT (see e.g. Proposition 2.27 of van der Vaart (1998)). First, note that conditionally on the data, $E^*\left(Z_k^*\right) = 0$ and

$$Var^*\left(\frac{1}{\sqrt{n}}\sum_{k=1}^n Z_k^*\right) = \frac{1}{n}\sum_{k=1}^n \lambda_k J_n^{-1/2}V'\phi_k\phi_k'V\left(J_n^{-1/2}\right) = J_n^{-1/2}V'\left(\frac{1}{n}\sum_{k=1}^n \lambda_k\phi_k\phi_k'\right)V\left(J_n^{-1/2}\right)' = J_n^{-1/2}J_{0n}^*\left(J_n^{-1/2}\right)',$$

where

$$J_{boot,n} = Var^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i\eta_i\right) = Var^*\left(\frac{1}{\sqrt{n}}\sum_{k=1}^n\left(V'\lambda_k^{1/2}\phi_k\right)v_k\right) = V'\left(\frac{1}{n}\sum_{k=1}^n\lambda_k\phi_k\phi_k'\right)V.$$

By Lemma A.2, $J_{boot,n} - J_n \to^P 0$, which then implies that

$$Var^*\left(\frac{1}{\sqrt{n}}\sum_{k=1}^n Z_k^*\right) \to^P I_p.$$

Hence, it remains to check Lyapunov's condition, which requires that for some $\upsilon > 0$,

$$\frac{1}{n^{1+\upsilon/2}}\sum_{k=1}^n E^*\|Z_k^*\|^{2+\upsilon} \to^P 0. \tag{22}$$

Note that

$$\begin{aligned}
&\frac{1}{n^{1+\upsilon/2}}\sum_{k=1}^n E^*\|Z_k^*\|^{2+\upsilon} = \frac{1}{n^{1+\upsilon/2}}\sum_{k=1}^n E^*\left\|J_n^{-1/2}\left(\lambda_k^{1/2}V'\phi_k\right)v_k\right\|^{2+\upsilon}\\
&\leq \left\|J_n^{-1/2}\right\|^{2+\upsilon}\frac{1}{n^{1+\upsilon/2}}\left(\sup_a\lambda_a^{1+\upsilon/2}\right)\sum_{k=1}^n E^*\left\|\sum_{i=1}^n V_i\phi_{ik}v_k\right\|^{2+\upsilon}\\
&= \left\|J_n^{-1/2}\right\|^{2+\upsilon}\frac{1}{n^{\upsilon/2}}\left(\max_a\lambda_a^{1+\upsilon/2}\right)\frac{1}{n}\sum_{k=1}^n\left(\sum_{m=1}^p\sum_{i=1}^n\sum_{j=1}^n V_i^{(m)}V_j^{(m)}\phi_{ik}\phi_{jk}\right)^{1+\upsilon/2}E^*|v_k|^{2+\upsilon}.
\end{aligned} \tag{23}$$

We will show that the Lyapunov condition holds for $\upsilon = 2$ by showing that

$$\frac{1}{n}\left(\max_a\lambda_a^2\right) = o_P(1), \tag{24}$$

$$\frac{1}{n}\sum_{k=1}^n\left(\sum_{m=1}^p\sum_{i=1}^n\sum_{j=1}^n V_i^{(m)}V_j^{(m)}\phi_{ik}\phi_{jk}\right)^2 = O_P(1), \tag{25}$$

and noting that $E^* |v_k|^4 < M$ by assumption. To prove (24), since $\mathbb{K}_n^* \phi_a = \lambda_a \phi_a$, for $a = 1, \ldots, n$, we have that for each $i = 1, \ldots, n$, and $a = 1, \ldots, n$,

$$\sum_{j=1}^{n} K^* \left(\frac{\tilde{d}_{ij}}{d_n^*} \right) \phi_{ja} = \lambda_a \phi_{ia}.$$

Let $i = i_a$ such that $\left|\phi_{i_a a}\right| = \max_i |\phi_{ia}|$. Then, for $a = 1, ..., n$, it follows that

$$\lambda_a \left|\phi_{i_a a}\right| \leq \sum_{j=1}^{n} \left| K^* \left(\frac{\tilde{d}_{i_a j}}{d_n^*} \right) \right| \left|\phi_{ja}\right| \iff \lambda_a \leq \sum_{j=1}^{n} \left| K^* \left(\frac{\tilde{d}_{i_a j}}{d_n^*} \right) \right| \frac{\left|\phi_{ja}\right|}{\left|\phi_{i_a a}\right|} \leq \sum_{j=1}^{n} \left| K^* \left(\frac{\tilde{d}_{i_a j}}{d_n^*} \right) \right|.$$

Thus, we can obtain an upper bound for $\{\lambda_a\}$ as follows:

$$\sup_a \lambda_a \leq \sup_a \sum_{j=1}^{n} \left| K^* \left(\frac{\tilde{d}_{aj}}{d_n^*} \right) \right|.$$

Assumptions 1 and 5 imply

$$\sup_a \sum_{j=1}^{n} \left| K^* \left(\frac{\tilde{d}_{aj}}{d_n^*} \right) \right| \leq \underbrace{\sup_a \sum_{j \in \mathscr{B}_{a,n}^*} 1}_{= \sup_a \ell_{a,n}^* < cE\ell_n^*} + \sup_a \sum_{j \notin \mathscr{B}_{a,n}^*} \left| K^* \left(\frac{\tilde{d}_{i_a j}}{d_n^*} \right) \right|,$$

and since $\lambda_a \geq 0$ for $a = 1, ..., n$,

$$\begin{aligned} \frac{1}{n} \sup_a \lambda_a^2 &\leq \frac{1}{n} \sup_a \left(\sum_{j=1}^{n} \left| K^* \left(\frac{\tilde{d}_{aj}}{d_n^*} \right) \right| \right)^2 \leq \frac{1}{n} \left(cE\ell_n^* + \sup_a \sum_{j \notin \mathscr{B}_{a,n}^*} \left| K^* \left(\frac{\tilde{d}_{i_a j}}{d_n^*} \right) \right| \right)^2 \\ &\leq 2 \frac{(cE\ell_n^*)^2}{n} + 2 \frac{(E\ell_n^*)^2}{n} \left(\frac{1}{E\ell_n^*} \sup_a \sum_{j \notin \mathscr{B}_{a,n}^*} \left| K^* \left(\frac{\tilde{d}_{i_a j}}{d_n^*} \right) \right| \right)^2 \leq O_P \left(\left(\frac{E\ell_n^*}{n^{1/2}} \right)^2 \right) = o_P(1) \end{aligned}$$

given Assumption 1(ii) and the fact that we let $\frac{E\ell_n^*}{n^{1/2}} = o(1)$.

Next we prove (25). We will focus on the special case where $p = 1$ for simplicity, and show that

$$\frac{1}{n} \sum_{k=1}^{n} E \left(\sum_{i=1}^{n} \sum_{j=1}^{n} V_i \phi_{ik} V_j \phi_{jk} \right)^2 = E \left[\frac{1}{n} \sum_{k=1}^{n} E \left(\left(\sum_{i=1}^{n} V_i \phi_{ik} \right)^4 |\Psi_n \right) \right] = O(1),$$

which suffices to prove (25) given Markov's inequality. In particular, we will argue conditionally on Ψ_n and show that the average of the conditional expectation is bounded by a constant. This is enough to prove that the unconditional expectation of the average is bounded. Letting $\tilde{V}_{ik} = V_i \phi_{ik}$, we have that

$$\frac{1}{n} \sum_{k=1}^{n} E \left(\left| \sum_{i=1}^{n} \tilde{V}_{ik} \right|^4 |\Psi_n \right) = \frac{1}{n} \sum_{k=1}^{n} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \sum_{i_4=1}^{n} E \left(\tilde{V}_{i_1 k} \tilde{V}_{i_2 k} \tilde{V}_{i_3 k} \tilde{V}_{i_4 k} |\Psi_n \right) \equiv C_1 + C_2 + C_3 + C_4,$$

where, adding and subtracting appropriately,

$$\begin{aligned}
C_1 &= \frac{1}{n}\sum_{k=1}^{n}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}\left\{\begin{array}{c} E\left(\tilde{V}_{i_1k}\tilde{V}_{i_2k}\tilde{V}_{i_3k}\tilde{V}_{i_4k}|\Psi_n\right) - E\left(\tilde{V}_{i_1k}\tilde{V}_{i_2k}|\Psi_n\right)E\left(\tilde{V}_{i_3k}\tilde{V}_{i_4k}|\Psi_n\right) \\ -E\left(\tilde{V}_{i_1k}\tilde{V}_{i_3k}|\Psi_n\right)E\left(\tilde{V}_{i_2k}\tilde{V}_{i_4k}|\Psi_n\right) - E\left(\tilde{V}_{i_1k}\tilde{V}_{i_4k}|\Psi_n\right)E\left(\tilde{V}_{i_2k}\tilde{V}_{i_3k}|\Psi_n\right)\end{array}\right\} \\
C_2 &= \frac{1}{n}\sum_{k=1}^{n}\left\{\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}E\left(\tilde{V}_{i_1k}\tilde{V}_{i_2k}|\Psi_n\right)\right\}\left\{\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}E\left(\tilde{V}_{i_3k}\tilde{V}_{i_4k}|\Psi_n\right)\right\} \\
C_3 &= \frac{1}{n}\sum_{k=1}^{n}\left\{\sum_{i_1=1}^{n}\sum_{i_3=1}^{n}E\left(\tilde{V}_{i_1k}\tilde{V}_{i_3k}|\Psi_n\right)\right\}\left\{\sum_{i_2=1}^{n}\sum_{i_4=1}^{n}E\left(\tilde{V}_{i_2k}\tilde{V}_{i_4k}|\Psi_n\right)\right\} \\
C_4 &= \frac{1}{n}\sum_{k=1}^{n}\left\{\sum_{i_1=1}^{n}\sum_{i_4=1}^{n}E\left(\tilde{V}_{i_1k}\tilde{V}_{i_4k}|\Psi_n\right)\right\}\left\{\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}E\left(\tilde{V}_{i_2k}\tilde{V}_{i_3k}|\Psi_n\right)\right\}.
\end{aligned}$$

We will now show that each of the terms C_1 through C_4 is bounded by a constant given our assumptions. Recall that $\tilde{V}_{ik} \equiv V_i\phi_{ik}$, where $V_i = \sum_{\ell=1}^{\infty} r_{i\ell}e_\ell$ given the linear array representation of V_i (Assumption 2). We will rely on this assumption as well as on the orthonormality of the eigenvectors $\phi_k = (\phi_{ik} : i = 1, \ldots, n)$ to prove the desired results. Write

$$\tilde{V}_{ik} = V_i\phi_{ik} = \sum_{l=1}^{\infty}(r_{il}\phi_{ik})e_l = \sum_{l=1}^{\infty}\tilde{r}_{il,k}e_l, \text{ where } \tilde{r}_{il,k} = r_{il}\phi_{ik}.$$

Using the fact that e_l are i.i.d.$(0,1)$, it follows that

$$\begin{aligned}
E\left(\tilde{V}_{i_1k}\tilde{V}_{i_2k}\tilde{V}_{i_3k}\tilde{V}_{i_4k}|\Psi_n\right) &= \sum_{l_1=1}^{\infty}\sum_{l_2=1}^{\infty}\sum_{l_3=1}^{\infty}\sum_{l_4=1}^{\infty}\tilde{r}_{i_1l_1,k}\tilde{r}_{i_2l_2,k}\tilde{r}_{i_3l_3,k}\tilde{r}_{i_4l_4,k}E\left(e_{l_1}e_{l_2}e_{l_3}e_{l_4}\right) \\
&= \sum_{l=1}^{\infty}\tilde{r}_{i_1l,k}\tilde{r}_{i_2l,k}\tilde{r}_{i_3l,k}\tilde{r}_{i_4l,k}E\left(e_l^4\right) + \sum_{l_1=1}^{\infty}\tilde{r}_{i_1l_1,k}\tilde{r}_{i_2l_1,k}\sum_{l_2=1,l_1\neq l_2}^{\infty}\tilde{r}_{i_3l_2,k}\tilde{r}_{i_4l_2,k} \\
&\quad + \sum_{l_1=1}^{\infty}\tilde{r}_{i_1l_1,k}\tilde{r}_{i_3l_1,k}\sum_{l_2=1,l_1\neq l_2}^{\infty}\tilde{r}_{i_2l_2,k}\tilde{r}_{i_4l_2,k} + \sum_{l_1=1}^{\infty}\tilde{r}_{i_1l_1,k}\tilde{r}_{i_4l_1,k}\sum_{l_2=1,l_1\neq l_2}^{\infty}\tilde{r}_{i_2l_2,k}\tilde{r}_{i_3l_2,k} \\
&\equiv d_1 + d_2 + d_3 + d_4.
\end{aligned}$$

Now, notice that for a given pair (i,j), e.g. (i_1,i_2), we have that

$$E\left(\tilde{V}_{i_1k}\tilde{V}_{i_2k}|\Psi_n\right) = E\left[\left(\sum_{l_1=1}^{\infty}\tilde{r}_{i_1l_1,k}e_{l_1}\right)\left(\sum_{l_2=1}^{\infty}\tilde{r}_{i_2l_2,k}e_{l_2}\right)\Bigg|\Psi_n\right] = \sum_{l_1=1}^{\infty}\sum_{l_2=1}^{\infty}\tilde{r}_{i_1l_1,k}\tilde{r}_{i_2l_2,k}\underbrace{E\left(e_{l_1}e_{l_2}\right)}_{=0 \text{ if } l_1\neq l_2 \text{ and } 1 \text{ if } l_1=l_2} = \sum_{l_1=1}^{\infty}\tilde{r}_{i_1l_1,k}\tilde{r}_{i_2l_1,k}.$$

This implies that

$$\begin{aligned}
E\left(\tilde{V}_{i_1k}\tilde{V}_{i_2k}|\Psi_n\right)E\left(\tilde{V}_{i_3k}\tilde{V}_{i_4k}|\Psi_n\right) &= \left(\sum_{l_1=1}^{\infty}\tilde{r}_{i_1l_1,k}\tilde{r}_{i_2l_1,k}\right)\left(\sum_{l_2=1,l_2\neq l_1}^{\infty}\tilde{r}_{i_3l_2,k}\tilde{r}_{i_4l_2,k} + \underbrace{\tilde{r}_{i_3l_1,k}\tilde{r}_{i_4l_1,k}}_{\text{when } l_1=l_2}\right) \\
&= \left(\sum_{l_1=1}^{\infty}\tilde{r}_{i_1l_1,k}\tilde{r}_{i_2l_1,k}\right)\left(\sum_{l_2=1,l_2\neq l_1}^{\infty}\tilde{r}_{i_3l_2,k}\tilde{r}_{i_4l_2,k}\right) + \left(\sum_{l_1=1}^{\infty}\tilde{r}_{i_1l_1,k}\tilde{r}_{i_2l_1,k}\tilde{r}_{i_3l_1,k}\tilde{r}_{i_4l_1,k}\right)
\end{aligned}$$

Hence,

$$d_2 \equiv \left(\sum_{l_1=1}^{\infty}\tilde{r}_{i_1l_1,k}\tilde{r}_{i_2l_1,k}\right)\left(\sum_{l_2=1,l_2\neq l_1}^{\infty}\tilde{r}_{i_3l_2,k}\tilde{r}_{i_4l_2,k}\right) = E\left(\tilde{V}_{i_1k}\tilde{V}_{i_2k}|\Psi_n\right)E\left(\tilde{V}_{i_3k}\tilde{V}_{i_4k}|\Psi_n\right) - \sum_{l_1=1}^{\infty}\tilde{r}_{i_1l_1,k}\tilde{r}_{i_2l_1,k}\tilde{r}_{i_3l_1,k}\tilde{r}_{i_4l_1,k}.$$

Similarly,

$$d_3 \equiv \left(\sum_{l_1=1}^{\infty} \tilde{r}_{i_1 l_1,k}\tilde{r}_{i_3 l_1,k}\right)\left(\sum_{l_2=1,l_2\neq l_1}^{\infty} \tilde{r}_{i_2 l_2,k}\tilde{r}_{i_4 l_2,k}\right) = E\left(\tilde{V}_{i_1k}\tilde{V}_{i_3k}|\Psi_n\right)E\left(\tilde{V}_{i_2k}\tilde{V}_{i_4k}|\Psi_n\right) - \sum_{l_1=1}^{\infty}\tilde{r}_{i_1 l_1,k}\tilde{r}_{i_2 l_1,k}\tilde{r}_{i_3 l_1,k}\tilde{r}_{i_4 l_1,k}$$

and

$$d_4 \equiv \left(\sum_{l_1=1}^{\infty} \tilde{r}_{i_1 l_1,k}\tilde{r}_{i_2 l_1,k}\right)\left(\sum_{l_2=1,l_2\neq l_1}^{\infty} \tilde{r}_{i_3 l_2,k}\tilde{r}_{i_4 l_2,k}\right) = E\left(\tilde{V}_{i_1k}\tilde{V}_{i_4k}|\Psi_n\right)E\left(\tilde{V}_{i_2k}\tilde{V}_{i_3k}|\Psi_n\right) - \sum_{l_1=1}^{\infty}\tilde{r}_{i_1 l_1,k}\tilde{r}_{i_2 l_1,k}\tilde{r}_{i_3 l_1,k}\tilde{r}_{i_4 l_1,k}.$$

Putting everything together yields

$$\begin{aligned}
&E\left(\tilde{V}_{i_1k}\tilde{V}_{i_2k}\tilde{V}_{i_3k}\tilde{V}_{i_4k}|\Psi_n\right) - E\left(\tilde{V}_{i_1k}\tilde{V}_{i_2k}|\Psi_n\right)E\left(\tilde{V}_{i_3k}\tilde{V}_{i_4k}|\Psi_n\right)\\
&- E\left(\tilde{V}_{i_1k}\tilde{V}_{i_3k}|\Psi_n\right)E\left(\tilde{V}_{i_2k}\tilde{V}_{i_4k}|\Psi_n\right) - E\left(\tilde{V}_{i_1k}\tilde{V}_{i_4k}|\Psi_n\right)E\left(\tilde{V}_{i_2k}\tilde{V}_{i_3k}|\Psi_n\right)\\
&= \sum_{l=1}^{\infty}\tilde{r}_{i_1 l,k}\tilde{r}_{i_2 l,k}\tilde{r}_{i_3 l,k}\tilde{r}_{i_4 l,k}E\left(e_l^4\right) - 3\sum_{l=1}^{\infty}\tilde{r}_{i_1 l,k}\tilde{r}_{i_2 l,k}\tilde{r}_{i_3 l,k}\tilde{r}_{i_4 l,k} = \sum_{l=1}^{\infty}\tilde{r}_{i_1 l,k}\tilde{r}_{i_2 l,k}\tilde{r}_{i_3 l,k}\tilde{r}_{i_4 l,k}\underbrace{\left(E\left(e_l^4\right)-3\right)}_{=\kappa_4},
\end{aligned}$$

which then implies that

$$C_1 = \kappa_4\frac{1}{n}\sum_{k=1}^{n}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}\sum_{l=1}^{\infty}\tilde{r}_{i_1 l,k}\tilde{r}_{i_2 l,k}\tilde{r}_{i_3 l,k}\tilde{r}_{i_4 l,k}.$$

To bound this term, note that $\kappa_4 < \Delta$ under our assumptions and therefore

$$\begin{aligned}
C_1 &\le \Delta\frac{1}{n}\sum_{k=1}^{n}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}\sum_{l=1}^{\infty}\left|\tilde{r}_{i_1 l,k}\tilde{r}_{i_2 l,k}\tilde{r}_{i_3 l,k}\tilde{r}_{i_4 l,k}\right|\\
&= \Delta\frac{1}{n}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}\sum_{l=1}^{\infty}\left|r_{i_1 l}r_{i_2 l}r_{i_3 l}r_{i_4 l}\right|\sum_{k=1}^{n}\left|\phi_{i_1k}\phi_{i_2k}\phi_{i_3k}\phi_{i_4k}\right|\\
&\le \Delta\frac{1}{n}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}\sum_{l=1}^{\infty}\left|r_{i_1 l}r_{i_2 l}r_{i_3 l}r_{i_4 l}\right|\left(\sum_{k=1}^{n}\left(\phi_{i_1k}\phi_{i_2k}\right)^2\right)^{1/2}\left(\sum_{k=1}^{n}\left(\phi_{i_3k}\phi_{i_4k}\right)^2\right)^{1/2}\\
&\le \Delta\frac{1}{n}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}\sum_{l=1}^{\infty}\left|r_{i_1 l}r_{i_2 l}r_{i_3 l}r_{i_4 l}\right|\left(\sum_{k=1}^{n}\phi_{i_1k}^4\right)^{1/4}\left(\sum_{k=1}^{n}\phi_{i_2k}^4\right)^{1/4}\left(\sum_{k=1}^{n}\phi_{i_3k}^4\right)^{1/4}\left(\sum_{k=1}^{n}\phi_{i_4k}^4\right)^{1/4}\\
&\le \Delta\frac{1}{n}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}\sum_{l=1}^{\infty}\left|r_{i_1 l}r_{i_2 l}r_{i_3 l}r_{i_4 l}\right|\left[\sup_i\left(\sum_{k=1}^{n}\phi_{ik}^4\right)\right]
\end{aligned}$$

We know that Φ_n is such that $\Phi_n'\Phi_n = \Phi_n\Phi_n' = I_n$, which implies that for each i, $\sum_{k=1}^{n}\phi_{ik}^2 = 1$. Write

$$\sum_{k=1}^{n}\phi_{ik}^4 = \sum_{k=1}^{n}\phi_{ik}^2\phi_{ik}^2.$$

Because $\sum_{k=1}^{n}\phi_{ik}^2 = 1$ for each i, it must be the case that $\sup_{1\le k\le n}\left|\phi_{ik}^2\right| \le 1$. Thus

$$\sum_{k=1}^{n}\phi_{ik}^4 = \sum_{k=1}^{n}\phi_{ik}^2\phi_{ik}^2 \le \sum_{k=1}^{n}\left|\phi_{ik}^2\right|\sup_k\left|\phi_{ik}^2\right| \le \sum_{k=1}^{n}\left|\phi_{ik}^2\right| = 1,$$

implying that

$$\begin{aligned} C_1 &\leq \Delta \frac{1}{n} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \sum_{i_4=1}^{n} \sum_{l=1}^{\infty} |r_{i_1 l} r_{i_2 l} r_{i_3 l} r_{i_4 l}| = \Delta \frac{1}{n} \sum_{i_1=1}^{n} \sum_{l=1}^{\infty} |r_{i_1 l}| \left(\sum_{i_2=1}^{n} |r_{i_2 l}| \right) \left(\sum_{i_3=1}^{n} |r_{i_3 l}| \right) \left(\sum_{i_4=1}^{n} |r_{i_4 l}| \right) \\ &\leq \Delta \frac{1}{n} \sum_{i_1=1}^{n} \sum_{l=1}^{\infty} |r_{i_1 l}| \underbrace{\left(\sum_{i_2=1}^{\infty} |r_{i_2 l}| \right)}_{\leq M \text{ by Assumption 2}} \left(\sum_{i_3=1}^{\infty} |r_{i_3 l}| \right) \left(\sum_{i_4=1}^{\infty} |r_{i_4 l}| \right) \leq \Delta M^3 \frac{1}{n} \sum_{i_1=1}^{n} \underbrace{\sum_{l=1}^{\infty} |r_{i_1 l}|}_{\leq M} \leq Const. \end{aligned}$$

To show that C_2, C_3, C_4 are also bounded by some constant that does not depend on Ψ_n, we apply Lemma A.3.

To prove Theorem 4.1, we rely on the following lemma.

Lemma A.4 *Suppose Assumptions 1-7 hold. If* $E^* |v_i|^4 < M$ *and* $d_n, E\ell_n \to \infty$ *and* $d_n^*, E\ell_n^* \to \infty$ *as* $n \to \infty$ *such that* $E\ell_n/n = o(1)$ *and* $E\ell_n^*/n^{1/2} = o(1)$*, then (i)* $\hat{J}_n^* - \hat{J}_{boot,n} \to^{P^*} 0$*, in prob-P when unrestricted residuals are used, and (ii)* $\hat{J}_n^* - \hat{J}_{boot,n} \to^{P^*} 0$*, in prob-P, when restricted residuals are used, and* H_0 *is true.*

Proof of Lemma A.4. We focus on the proof of (i) since (ii) follows by similar arguments because $\tilde{\beta} - \beta$ is $\sqrt{n}$-convergent under H_0. Without loss of generality, we take $p = 1$. Let

$$\hat{u}_i^* = y_i^* - x_i' \hat{\beta}^* = y_i^* - x_i' \hat{\beta} + x_i' \left(\hat{\beta} - \hat{\beta}^* \right) = u_i^* + x_i' \left(\hat{\beta} - \hat{\beta}^* \right),$$

where $u_i^* = \hat{u}_i \eta_i$. It follows that

$$\hat{J}_n^* - \hat{J}_{boot,n} = \left(\hat{J}_n^* - \hat{J}_n \right) + \left(\hat{J}_n - \hat{J}_{boot,n} \right) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n} \right) x_i x_j \left(\hat{u}_i^* \hat{u}_j^* - \hat{u}_i \hat{u}_j \right) + \left(\hat{J}_n - \hat{J}_{boot,n} \right) \equiv A_1 + A_2 + A_3 + A_4,$$

where

$$A_1 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n} \right) x_i x_j \left[u_i^* u_j^* - \hat{u}_i \hat{u}_j \right]$$

and

$$A_2 = 2 \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n} \right) x_i x_j^2 u_i^* \left(\hat{\beta} - \hat{\beta}^* \right), \; A_3 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n} \right) x_i^2 x_j^2 \left(\hat{\beta} - \hat{\beta}^* \right)^2, \text{ and } A_4 = \hat{J}_n - J_n^*.$$

First, note that $A_4 = o_P(1)$ by Lemma A.2. Next, we show that A_2 and A_3 are $o_{P^*}(1)$, in probability. For these terms, we can use the fact that $\sqrt{n}\left(\hat{\beta} - \hat{\beta}^* \right) = O_{P^*}(1)$. Starting with A_3, we can write

$$A_3 = \underbrace{\left[\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{\tilde{d}_{ij}}{d_n} \right) x_i^2 x_j^2 \right]}_{= A_{31} = o_P(1)} \underbrace{\left(\sqrt{n} \left(\hat{\beta} - \hat{\beta}^* \right) \right)^2}_{= O_{P^*}(1)} = o_{P^*}(1),$$

in probability, since we can show that $A_{31} = o_P(1)$. Indeed,

$$E|A_{31}| \leq \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left| K\left(\frac{\tilde{d}_{ij}}{d_n} \right) \right| \underbrace{E\left(x_i^2 x_j^2 \right)}_{\leq M} \leq M \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left| K\left(\frac{\tilde{d}_{ij}}{d_n} \right) \right| = O\left(\frac{E\ell_n}{n} \right),$$

by Lemma A.1. Thus, by Markov's inequality, $A_{31} = O_P\left(\frac{E\ell_n}{n}\right) = o_P(1)$. For A_2,

$$A_2 = \underbrace{\left[\frac{1}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{\tilde{d}_{ij}}{d_n}\right)x_i x_j^2 u_i^*\right]}_{A_{21}=o_{P^*}(1)} \times \underbrace{\sqrt{n}\left(\hat{\beta}-\hat{\beta}^*\right)}_{=O_{P^*}(1)},$$

since we can show that the term in square brackets is $o_{P^*}(1)$. To see this, note that

$$A_{21} = \frac{1}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{\tilde{d}_{ij}}{d_n}\right)x_i\left(\hat{u}_i-u_i\right)\eta_i x_j^2 + \frac{1}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{\tilde{d}_{ij}}{d_n}\right)x_i u_i \eta_i x_j^2 \equiv A_{21}^{(1)} + A_{21}^{(2)}.$$

Starting with $A_{21}^{(1)}$, note that

$$\begin{aligned}
\left|A_{21}^{(1)}\right| &= \left|\frac{1}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{\tilde{d}_{ij}}{d_n}\right)x_i^2\left(\beta-\hat{\beta}\right)\eta_i x_j^2\right| \le O_P(1)\frac{1}{n}\sum_{j=1}^{n}\left|x_j^2\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{\tilde{d}_{ij}}{d_n}\right)x_i^2\eta_i\right| \\
&\le O_P(1)\underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}x_j^4\right)^{1/2}}_{=O_P(1)}\underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{\tilde{d}_{ij}}{d_n}\right)x_i^2\eta_i\right)^2\right)^{1/2}}_{=e_1=O_{P^*}(E\ell_n/n)\text{ in prob.}},
\end{aligned}$$

where $e_1 = O_{P^*}(E\ell_n/n)$ in probability. For this result, it suffices to show that $E\left(E^*\left(|e_1|^2\right)\right) = O(E\ell_n/n)$. But

$$\begin{aligned}
E^*\left(|e_1|^2\right) &= \frac{1}{n}\sum_{j=1}^{n}\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}K\left(\frac{\tilde{d}_{i_1 j}}{d_n}\right)K\left(\frac{\tilde{d}_{i_2 j}}{d_n}\right)x_{i_1}^2 x_{i_2}^2 E^*\left(\eta_{i_1}\eta_{i_2}\right) \\
&= \frac{1}{n}\sum_{j=1}^{n}\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\left(\frac{\tilde{d}_{i_1 j}}{d_n}\right)K\left(\frac{\tilde{d}_{i_2 j}}{d_n}\right)x_{i_1}^2 x_{i_2}^2 K^*\left(\frac{\tilde{d}_{i_1 i_2}}{d_n^*}\right) \\
&\le \frac{1}{n}\sum_{j=1}^{n}\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\left|K\left(\frac{\tilde{d}_{i_1 j}}{d_n}\right)\right|\left|K\left(\frac{\tilde{d}_{i_2 j}}{d_n}\right)\right|x_{i_1}^2 x_{i_2}^2,
\end{aligned}$$

implying that

$$\begin{aligned}
E\left(E^*\left|e_1^2\right|\right) &\le \frac{1}{n}\sum_{j=1}^{n}\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}E\left|K\left(\frac{\tilde{d}_{i_1 j}}{d_n}\right)\right|\underbrace{\left|K\left(\frac{\tilde{d}_{i_2 j}}{d_n}\right)\right|}_{\le 1}\underbrace{E\left(x_{i_1}^2 x_{i_2}^2\right)}_{\le M} \\
&\le \frac{M}{n^2}\sum_{j=1}^{n}\sum_{i_1=1}^{n}E\left|K\left(\frac{\tilde{d}_{i_1 j}}{d_n}\right)\right| = O\left(\frac{E\ell_n}{n}\right)
\end{aligned}$$

as shown in Lemma A.1. Therefore, $A_{21}^{(1)} = O_{P^*}(E\ell_n/n) = o_{P^*}(1)$, in probability. A similar argument implies that $A_{21}^{(2)} = O_{P^*}\left(\sqrt{\frac{E\ell_n}{n}}\right) = o_{P^*}(1)$, in probability. Thus, to end the proof, we show that $A_1 = o_{P^*}(1)$ in probability. We can write

$$\begin{aligned}
A_1 = &\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)x_i x_j\left[x_i x_j\left(\beta-\hat{\beta}\right)^2 + 2x_i u_j\left(\beta-\hat{\beta}\right)\right]\eta_i\eta_j \\
&-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)x_i x_j\left[x_i x_j\left(\beta-\hat{\beta}\right)^2 + 2x_i u_j\left(\beta-\hat{\beta}\right)\right] \\
&+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)x_i x_j u_i u_j\left(\eta_i\eta_j - 1\right) \equiv A_{11} + A_{12} + A_{13},
\end{aligned}$$

where $A_{12} = -\left(J_n^* - J_{0n}^*\right) = o_P(1)$, as shown in the proof of Lemma A.2. Thus, it suffices to show that A_{11} and A_{13} are $o_{P^*}(1)$, in probability.

We can decompose A_{11} as $A_{11} = A_{11}^{(1)} + A_{11}^{(2)}$, where

$$
\begin{aligned}
\left|A_{11}^{(1)}\right| &\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|K\left(\frac{d_{ij}}{d_n}\right)\right|\left|x_i^2 x_j^2 \eta_i \eta_j\right|\left(\beta - \hat{\beta}\right)^2 \\
&= O_P(1)\frac{1}{n}\sum_{i=1}^{n}\left|x_i^2 \eta_i\right|\left(\frac{1}{n}\sum_{j=1}^{n}\left|K\left(\frac{\tilde{d}_{ij}}{d_n}\right)\right|\left|x_j^2 \eta_j\right|\right) \\
&\leq O_P(1)\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\left|x_i^2 \eta_i\right|^2\right)^{1/2}}_{=O_{P^*}(1)}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j=1}^{n}\left|K\left(\frac{\tilde{d}_{ij}}{d_n}\right)\right|\left|x_j^2 \eta_j\right|\right)^2\right)^{1/2}}_{=e_1=o_{P^*}(1) \text{ in prob}}.
\end{aligned}
$$

where $E\left(E^*\left(e_1\right)\right) = o(1)$, as shown before. Thus, $A_{11}^{(1)} = O_{P^*}\left(\frac{E\ell_n}{n}\right)$ in prob-P. Using arguments similar to those used before, we can show that $A_{11}^{(2)} = o_{P^*}(1)$, in probability, concluding the proof that $A_{11} = o_{P^*}(1)$, in prob-P. Finally, we show that $A_{13} = o_{P^*}(1)$. We have

$$
\begin{aligned}
A_{13} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{\tilde{d}_{ij}}{d_n}\right)V_i V_j\left(\eta_i \eta_j - 1\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{\tilde{d}_{ij}}{d_n}\right)V_i V_j\left(\eta_i \eta_j - K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{\tilde{d}_{ij}}{d_n}\right)V_i V_j\left(K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right) - 1\right) \\
&= A_{13}^{(1)} + A_{13}^{(2)},
\end{aligned} \tag{26}
$$

For $A_{13}^{(1)}$, we prove that $Var^*\left(A_{13}^{(1)}\right) = o_P(1)$ since $E^*\left(A_{13}^{(1)}\right) = 0$. By Markov's inequality, it suffices to show that $E\left(Var^*\left(A_{13}^{(1)}\right)\right) = o(1)$. We have that

$$
E\left(Var^*\left(A_{13}^{(1)}\right)\right) = E\left\{E^*\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{\tilde{d}_{ij}}{d_n}\right)V_i V_j\left(\eta_i \eta_j - K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right)\right)^2\right\}
$$

$$
= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right)E\left[K\left(\frac{\tilde{d}_{i_1 j_1}}{d_n}\right)K\left(\frac{\tilde{d}_{i_2 j_2}}{d_n}\right)\left(E^*\left(\eta_{i_1}\eta_{j_1}\eta_{i_2}\eta_{j_2}\right) - K^*\left(\frac{\tilde{d}_{i_1 j_1}}{d_n^*}\right)K^*\left(\frac{\tilde{d}_{i_2 j_2}}{d_n^*}\right)\right)\right].
$$

Let L_{ik} denote the (i,k)-th element of L_n such that $\mathbb{K}_n^* = L_n L_n'$. In particular, letting $L_n = \Phi_n \Lambda_n^{1/2}$ implies that

$$
\eta_i = \sum_{k=1}^{n} L_{ik} v_k = \sum_{k=1}^{n}\underbrace{\left(\sqrt{\lambda_k}\phi_{ik}\right)}_{=L_{ik}} v_k,
$$

where v_k is i.i.d.$(0,1)$. This decomposition implies that for any pair (i,j),

$$
E^*\left(\eta_i \eta_j\right) = K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right) = \sum_{k_1=1}^{n}\sum_{k_2=1}^{n} L_{ik_1} L_{jk_2} E^*\left(v_{k_1} v_{k_2}\right) = \sum_{k=1}^{n} L_{ik} L_{jk}.
$$

Similarly, it follows that

$$\begin{aligned}
E^*\left(\eta_{i_1}\eta_{j_1}\eta_{i_2}\eta_{j_2}\right) &= \sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\sum_{k_3=1}^{n}\sum_{k_4=1}^{n} L_{i_1k_1}L_{j_1k_2}L_{i_2k_3}L_{j_2k_4}E^*\left(v_{k_1}v_{k_2}v_{k_3}v_{k_4}\right) \\
&= \sum_{k=1}^{n} L_{i_1k}L_{j_1k}L_{i_2k}L_{j_2k}\left(E^*\left(v_k^4\right)-3\right) \\
&\quad + \sum_{k_1=1}^{n} L_{i_1k_1}L_{j_1k_1}\sum_{k_3=1}^{n} L_{i_2k_3}L_{j_2k_3} + \sum_{k_1=1}^{n} L_{i_1k_1}L_{i_2k_1}\sum_{k_2=1}^{n} L_{j_1k_2}L_{j_2k_2} + \sum_{k_1=1}^{n} L_{i_1k_1}L_{j_2k_1}\sum_{k_2=1}^{n} L_{j_1k_2}L_{i_2k_2} \\
&= \sum_{k=1}^{n} L_{i_1k}L_{j_1k}L_{i_2k}L_{j_2k}\left(E^*\left(v_k^4\right)-3\right) \\
&\quad + K^*\left(\frac{\tilde{d}_{i_1j_1}}{d_n^*}\right)K^*\left(\frac{\tilde{d}_{i_2j_2}}{d_n^*}\right) + K^*\left(\frac{\tilde{d}_{i_1i_2}}{d_n^*}\right)K^*\left(\frac{\tilde{d}_{j_1j_2}}{d_n^*}\right) + K^*\left(\frac{\tilde{d}_{i_1j_2}}{d_n^*}\right)K^*\left(\frac{\tilde{d}_{j_1i_2}}{d_n^*}\right).
\end{aligned}$$

Thus,

$$\begin{aligned}
E^*\left(\eta_{i_1}\eta_{j_1}\eta_{i_2}\eta_{j_2}\right) - K^*\left(\frac{\tilde{d}_{i_1j_1}}{d_n^*}\right)K^*\left(\frac{\tilde{d}_{i_2j_2}}{d_n^*}\right) &= \sum_{k=1}^{n} L_{i_1k}L_{j_1k}L_{i_2k}L_{j_2k}\left(E^*\left(v_k^4\right)-3\right) \\
&\quad + K^*\left(\frac{\tilde{d}_{i_1i_2}}{d_n^*}\right)K^*\left(\frac{\tilde{d}_{j_1j_2}}{d_n^*}\right) + K^*\left(\frac{\tilde{d}_{i_1j_2}}{d_n^*}\right)K^*\left(\frac{\tilde{d}_{j_1i_2}}{d_n^*}\right).
\end{aligned}$$

Given this decomposition, it follows that

$$E\left(Var^*\left(A_{13}^{(1)}\right)\right) = B_{11} + B_{12} + B_{13},$$

where

$$\begin{aligned}
B_{11} &= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n} E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right)E\left[K\left(\frac{\tilde{d}_{i_1j_1}}{d_n}\right)K\left(\frac{\tilde{d}_{i_2j_2}}{d_n}\right)\left(\sum_{k=1}^{n} L_{i_1k}L_{j_1k}L_{i_2k}L_{j_2k}\left(E^*\left(v_k^4\right)-3\right)\right)\right] \\
B_{12} &= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n} E\left[K\left(\frac{\tilde{d}_{i_1j_1}}{d_n}\right)K\left(\frac{\tilde{d}_{i_2j_2}}{d_n}\right)K^*\left(\frac{\tilde{d}_{i_1i_2}}{d_n^*}\right)K^*\left(\frac{\tilde{d}_{j_1j_2}}{d_n^*}\right)\right]E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right) \\
B_{13} &= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n} E\left[K\left(\frac{\tilde{d}_{i_1j_1}}{d_n}\right)K\left(\frac{\tilde{d}_{i_2j_2}}{d_n}\right)K^*\left(\frac{\tilde{d}_{i_1j_2}}{d_n^*}\right)K^*\left(\frac{\tilde{d}_{j_1i_2}}{d_n^*}\right)\right]E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right).
\end{aligned}$$

Since $\left(E^*\left(v_k^4\right)-3\right) \leq M$ by assumption, by adding and subtracting appropriately, we can bound B_{11} by

$$\begin{aligned}
B_{11} &\leq M\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}\left|E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right)\right|E\left|\sum_{k=1}^{n} L_{i_1k}L_{j_1k}L_{i_2k}L_{j_2k}\right|. \\
&\leq M\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}\left(\begin{array}{c} E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right)-E\left(V_{i_1}V_{j_1}\right)E\left(V_{i_2}V_{j_2}\right) \\ -E\left(V_{i_1}V_{i_2}\right)E\left(V_{j_1}V_{j_2}\right)-E\left(V_{i_1}V_{j_2}\right)E\left(V_{j_1}V_{i_2}\right)\end{array}\right)E\left|\sum_{k=1}^{n} L_{i_1k}L_{j_1k}L_{i_2k}L_{j_2k}\right| \\
&\quad + \left(\left|E\left(V_{i_1}V_{j_1}\right)E\left(V_{i_2}V_{j_2}\right)\right| + \left|E\left(V_{i_1}V_{i_2}\right)E\left(V_{j_1}V_{j_2}\right)\right| + \left|E\left(V_{i_1}V_{j_2}\right)E\left(V_{j_1}V_{i_2}\right)\right|\right)E\left|\sum_{k=1}^{n} L_{i_1k}L_{j_1k}L_{i_2k}L_{j_2k}\right| \\
&= B_{11}^{(1)} + B_{11}^{(2)} + B_{11}^{(3)} + B_{11}^{(4)}.
\end{aligned}$$

To bound $B_{11}^{(1)}$, we rely on Assumption 2 to write

$$E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right)-E\left(V_{i_1}V_{j_1}\right)E\left(V_{i_2}V_{j_2}\right)-E\left(V_{i_1}V_{i_2}\right)E\left(V_{j_1}V_{j_2}\right)-E\left(V_{i_1}V_{j_2}\right)E\left(V_{j_1}V_{i_2}\right) = \sum_{l=1}^{\infty} r_{i_1l}r_{j_1l}r_{i_2l}r_{j_2l}\left(E\left(e_l^4\right)-3\right),$$

by using an argument similar to the one used to study the term C_1 in the proof of Theorem 3.1. Also, recall that $L_{ik} = \sqrt{\lambda_k}\phi_{ik}$. Then,

$$\left|\sum_{k=1}^{n} L_{i_1k}L_{j_1k}L_{i_2k}L_{j_2k}\right| = \left|\sum_{k=1}^{n} \lambda_k^2 \phi_{i_1k}\phi_{j_1k}\phi_{i_2k}\phi_{j_2k}\right|.$$

This implies that

$$\begin{aligned}
B_{11}^{(1)} &\leq M\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}\left|\sum_{l=1}^{\infty} r_{i_1l}r_{j_1l}r_{i_2l}r_{j_2l}\left(E\left(e_l^4\right)-3\right)\right| E\sum_{k=1}^{n}\left|\lambda_k^2\phi_{i_1k}\phi_{j_1k}\phi_{i_2k}\phi_{j_2k}\right| \\
&\leq M^2\frac{1}{n^2}\sum_{k=1}^{n} E\left[\lambda_k^2\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}\sum_{l=1}^{\infty}\left|r_{i_1l}r_{j_1l}r_{i_2l}r_{j_2l}\phi_{i_1k}\phi_{j_1k}\phi_{i_2k}\phi_{j_2k}\right|\right] \\
&\leq M^2\frac{1}{n\sqrt{n}}\sum_{k=1}^{n} E\left[\lambda_k^2\frac{1}{\sqrt{n}}\sum_{i_1=1}^{n}\left|\phi_{i_1k}\right|\sum_{l=1}^{\infty}\left|r_{i_1l}\right|\sum_{j_1=1}^{n}\left|r_{j_1l}\phi_{j_1k}\right|\sum_{i_2=1}^{n}\left|r_{i_2l}\phi_{i_2k}\right|\sum_{j_2=1}^{n}\left|r_{j_2l}\phi_{j_2k}\right|\right].
\end{aligned}$$

Note that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i_1=1}^{n}\left|\phi_{i_1k}\right|\sum_{l=1}^{\infty}\left|r_{i_1l}\right|\sum_{j_1=1}^{n}\left|r_{j_1l}\right|\left|\phi_{j_1k}\right|\sum_{i_2=1}^{n}\left|r_{i_2l}\right|\left|\phi_{i_2k}\right|\sum_{j_2=1}^{n}\left|r_{j_2l}\right|\underbrace{\left|\phi_{j_2k}\right|}_{\leq 1} \\
\leq\ &\underbrace{\frac{1}{\sqrt{n}}\left(\sum_{i_1=1}^{n} 1^2\right)^{1/2}}_{=1}\underbrace{\left(\sum_{i_1=1}^{n}\phi_{i_1k}^2\right)^{1/2}}_{=1}\underbrace{\left(\sum_{l=1}^{\infty}\left|r_{i_1l}\right|\right)}_{\leq M}\left(\sum_{j_1=1}^{n}\left|r_{j_1l}\right|\right)\left(\sum_{i_2=1}^{n}\left|r_{i_2l}\right|\right)\left(\sum_{j_2=1}^{n}\left|r_{j_2l}\right|\right) \leq M^4,
\end{aligned}$$

which implies

$$B_{11}^{(1)} \leq M^6\frac{1}{n\sqrt{n}}E\left(\sum_{k=1}^{n}\lambda_k^2\right). \tag{27}$$

Since $\{\lambda_k\}$ are eigenvalues of $\mathbb{K}_n^*$, we have

$$\sum_{k=1}^{n}\lambda_k^2 = tr\left(\mathbb{K}_n^*\mathbb{K}_n^*\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right)^2. \tag{28}$$

Plugging (28) into (27) yields

$$B_{11}^{(1)} \leq M^6 E\left(\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right)^2\right) \leq O\left(\sqrt{n}\right)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left|K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right| = O\left(\frac{E\ell_n^*}{\sqrt{n}}\right).$$

For $B_{11}^{(2)}$, we have

$$
\begin{aligned}
B_{11}^{(2)} &= M\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}\left|E\left(V_{i_1}V_{j_1}\right)E\left(V_{i_2}V_{j_2}\right)\right|E\left|\sum_{k=1}^{n}L_{i_1k}L_{j_1k}L_{i_2k}L_{j_2k}\right| \\
&= M\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}\left|E\left(V_{i_1}V_{j_1}\right)E\left(V_{i_2}V_{j_2}\right)\right|E\left|\sum_{k=1}^{n}\lambda_k^2\phi_{i_1k}\phi_{j_1k}\phi_{i_2k}\phi_{j_2k}\right| \\
&\leq M\frac{1}{n^2}E\sum_{k=1}^{n}\lambda_k^2\underbrace{\left(\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\left|E\left(V_{i_1}V_{j_1}\right)\phi_{i_1k}\phi_{j_1k}\right|\right)}_{\leq M\text{ by Lemma A.3}}\left(\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}E\left(V_{i_2}V_{j_2}\right)\phi_{i_2k}\phi_{j_2k}\right) \\
&\leq M^3\frac{1}{n^2}E\left(\sum_{k=1}^{n}\lambda_k^2\right)=M^3\frac{1}{n^2}E\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\left(K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right)^2\right) \\
&\leq M^3\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left|K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right|=O_P\left(\frac{E\ell_n^*}{n}\right).
\end{aligned}
$$

Using the same procedure, we can show that $B_{11}^{(3)}=B_{11}^{(4)}=O\left(E\ell_n^*/n\right)$. Hence, $B_{11}=O\left(E\ell_n^*/\sqrt{n}\right)=o(1)$ given that $E\ell_n^*/\sqrt{n}=o(1)$. Since we can also show that the terms B_{12} and B_{13} are $o(1)$ by a similar argument, this concludes the proof that $A_{13}^{(1)}=o_{P^*}(1)$ in prob-*P*.

For $A_{13}^{(2)}$, the second term in (26), note that

$$
E\left(A_{13}^{(2)}\right)\leq\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|E\left(V_iV_j\right)\right|\left(1-K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)\right)=o(1),
$$

as d_n^* grows, as proved in the proof of Lemma A.2 (see term b_2 in particular). Hence, it is sufficient to show $Var\left(A_{13}^{(2)}\right)=o(1)$. We have

$$
\begin{aligned}
Var\left(A_{13}^{(2)}\right) &= Var\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{\tilde{d}_{ij}}{d_n}\right)V_iV_j\left(K^*\left(\frac{\tilde{d}_{ij}}{d_n^*}\right)-1\right)\right) \\
&= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}E\left[K\left(\frac{\tilde{d}_{i_1j_1}}{d_n}\right)K\left(\frac{\tilde{d}_{i_2j_2}}{d_n}\right)\left(K^*\left(\frac{\tilde{d}_{i_1j_1}}{d_n^*}\right)-1\right)\left(K^*\left(\frac{\tilde{d}_{i_2j_2}}{d_n^*}\right)-1\right)\right] \\
&\quad\times\left[E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right)-E\left(V_{i_1}V_{j_1}\right)E\left(V_{i_2}V_{j_2}\right)\right] \\
&\leq \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}E\left|K\left(\frac{\tilde{d}_{i_1j_1}}{d_n}\right)K\left(\frac{\tilde{d}_{i_2j_2}}{d_n}\right)\right|\left|E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right)-E\left(V_{i_1}V_{j_1}\right)E\left(V_{i_2}V_{j_2}\right)\right|=o(1),
\end{aligned}
$$

as showed above. Therefore, $A_{13}^{(2)}=o_P(1)$, completing the proof.

Proof of Theorem 3.2. The proof is in the text.

Proof of Theorem 4.1. It follows from Theorem 3.1 and Lemma A.4.

References

[1] Acemoglu, D., Autor D.H., Dorn, D., Hanson, G.H. and Price, B., (2016), Import competition and the great US employment sag of the 2000s, *Journal of Labor Economics*, 34, 141-198.

[2] Andrews, D. (1991), Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, 59, 817-858.

[3] Autor D.H., Dorn, D., Hanson, G.H. and Song, J., (2014), Trade adjustment: work-level evidence, *The Quarterly Journal of Economics*, 129(4), 1799-1860.

[4] Bartlett, M.S. (1955), *An Introduction to Stochastic Processes*. Cambridge University Press, Cambridge.

[5] Bernard, A.B., Jensen, B.J. and Schott, P.J., (2006), Survival of the best fit: Exposure to low-wage countries and the (uneven) growth of US manufacturing plants, *Journal of International Economics*, 68(1), 219-237.

[6] Bester, C.A., Conley, T.G., Hansen, C. B. and Vogelsang, T. J. (2016), Fixed-b Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators, *Econometric Theory*, 32, 154–186.

[7] Bester, C.A., Conley, T.G., Hansen, C. B. and Vogelsang, T. J. (2008), Fixed-b Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators, manuscript.

[8] Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2008), Bootstrap-based Improvements for Inference with Clustered Errors, *Review of Economics and Statistics*, 90, 414–427.

[9] Cavaliere, G. and Georgiev, I. (2020), Inference under random limit bootstrap measures, *Econometrica*, 88, 2547-2574.

[10] Chen, X. and Conley, T. (2001), A New Semiparametric Spatial Model for Panel Time Series, *Journal of Econometrics*, 105, 59–83.

[11] Conley, T.G. (1999), GMM Estimation with Cross-sectional Dependence, *Journal of Econometrics*, 92, 1–45.

[12] Conley, T. and Dupor, B. (2003), A Spatial Analysis of Sectoral Complementarity, *Journal of Political Economy*, 111, 311–352.

[13] Conley, T.G. and Ligon, E. (2002), Economic Distance and Cross-Country Spillovers, *Journal of Economic Growth*, 7(2), 157–188.

[14] Conley, T.G. and F. Molinari (2007), Spatial correlation robust inference with errors in location or distance, *Journal of Econometrics*, 140, 76–96.

[15] Conley, T.G. and G. Topa (2002), Socio-Economic Distance and Spatial Patterns in Unemployment, *Journal of Applied Econometrics*, 17, 303–327.

[16] Davidson, R. and MacKinnon, J. (1999), Bootstrap testing in nonlinear models, *International Economic Review*, 40, 1999, 487–508.

[17] Djogbenou, A., MacKinnon, J. and Nielsen, M. (2019). Asymptotic theory and wild bootstrap inference with clustered errors, *Journal of Econometrics*, 212, 393-412.

[18] Gneiting, T., (2002), Compactly Supported Correlation Functions, *Journal of Multivariate Analysis*, 83, 493–508.

[19] Gonçalves, S. and Vogelsang, T. J. (2011), Block bootstrap puzzles in HAC robust testing: the sophistication of the naive bootstrap, *Econometric Theory*, 27, 745–791.

[20] Hidalgo, J. and Schafgans, M. (2017), Inference and Testing Breaks in Large Dynamic Panels with Strong Cross-Sectional Dependence, *Journal of Econometrics*, 196, 259–274.

[21] Ibragimov, R. and U. Müller (2010), t-Statistic Based Correlation and Heterogeneity Robust Inference, *Journal of Business and Economics Statistics*, Vol.28, 453-468.

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

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

[24] Kim, M. S. and Sun, Y. (2011), Spatial Heteroskedasticity and Autocorrelation Consistent Estimation of Covariance Matrix, *Journal of Econometrics*, 160, 349–371.

[25] Kim, M. S. and Sun, Y. (2013), Heteroskedasticity and Spatiotemporal Dependence Robust Inference for Linear Panel Models with Fixed Effects, *Journal of Econometrics*, 177, 85–108.

[26] Kline, P. and A. Santos (2012a), A score based approach to wild bootstrap inference. *Journal of Econometric Methods*, 1, 23-41.

[27] Kline, P. and A. Santos (2012b), Higher order properties of the wild bootstrap under misspecification, Journal of Econometrics, 171, 54-70.

[28] Kojevnikov, D. (2021), The Bootstrap for Network Dependent Processes, Tilburg University, Working paper.

[29] Lahiri, S. N. and Zhu, J. (2006), Resampling Methods for Spatial Regression Models Under a Class of Stochastic Designs, *The Annals of Statistics*, 34, 1774–1813.

[30] Liang, K.-Y., and S. L. Zeger (1986), Longitudinal data analysis using generalized linear models, *Biometrika* 73:13–22.

[31] Lee, J. and Robinson, P. (2016), Series Estimation under Cross-sectional Dependence, *Journal of Econometrics*, 190, 1–17.

[32] Martellosio, F. (2012), The correlation structure of spatial autoregressions, *Econometric Theory* 28, 1373-91.

[33] McMurry, T. and D.N. Politis (2010), Banded and tapered estimates of autocovariance matrices and the linear process bootstrap, Journal of Time Series Analysis, 31, 471-482.

[34] Moulton, B. R. (1986), Random group effects and the precision of regression estimates, *Journal of Econometrics* 32, 385–397.

[35] Müller, U.K. and M.W. Watson (2021), Spatial correlation robust inference, manuscript.

[36] Nordman, D. J., Lahiri, S. N. and Fridley, B. L. (2007), Optimal Block Size for Variance Estimation by a Spatial Block Bootstrap Method, *Sankhya*, 69, 468–493.

[37] Phillips, P., Sun, Y. and Jin, S (2007), Consistent HAC Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation, *Journal of Statistical Planning and Inference*, 137, 985–1023.

[38] Pinkse, J. , Slade, M. E. and Brett, C. (2002), Spatial Price Competition: A Semiparametric Approach, *Econometrica*, 70, 1111–1153.

[39] Politis, D.N. (2011), Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices, *Econometric Theory*, 27,703-744.

[40] Robinson, P. (2011), Asymptotic Theory for Nonparametric Regression with Spatial Data, *Journal of Econometrics* 167, 521–542.

[41] Robinson, P. and S. Thawornkaiwong (2012), Statistical inference on regression with spatial dependence, *Journal of Econometrics* 165, 5–19.

[42] Sandoval, A. (2020), Import wholesalers and the effect of trade on Canadian producers, University of Western Ontario, Working paper, http://www.aldouwoecon.com

[43] Shao, X. (2010), The Dependent Wild Bootstrap, *Journal of the American Statistical Association*, 105, 218-235.

[44] Stein, M.L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, Springer.

[45] Sun, Y. and Kim, M. S. (2015), Asymptotic F Test in a GMM Framework with Cross Sectional Dependence, *Review of Economics and Statistics*, 97, 210–223.

[46] Utar H. (2017), Characteristics of international trade intermediaries and their location in the supply chain, In: Christensen B., Kowalczyk C. (eds) Globalization. Springer, Berlin, Heidelberg

[47] van der Vaart, A. (1998), *Asymptotic Statistics*. Cambridge University Press, Cambridge.

[48] Yaglom, A. M. (1987), *Correlation Theory of Stationary and Related Random Functions*, Springer, New York.

[49] Zhang, X. and X. Shao (2013), Fixed-smoothing asymptotics for time series, *Annals of Statistics*, 41, 1329-1349.

Figure 1. Rejection rate with no measurement error

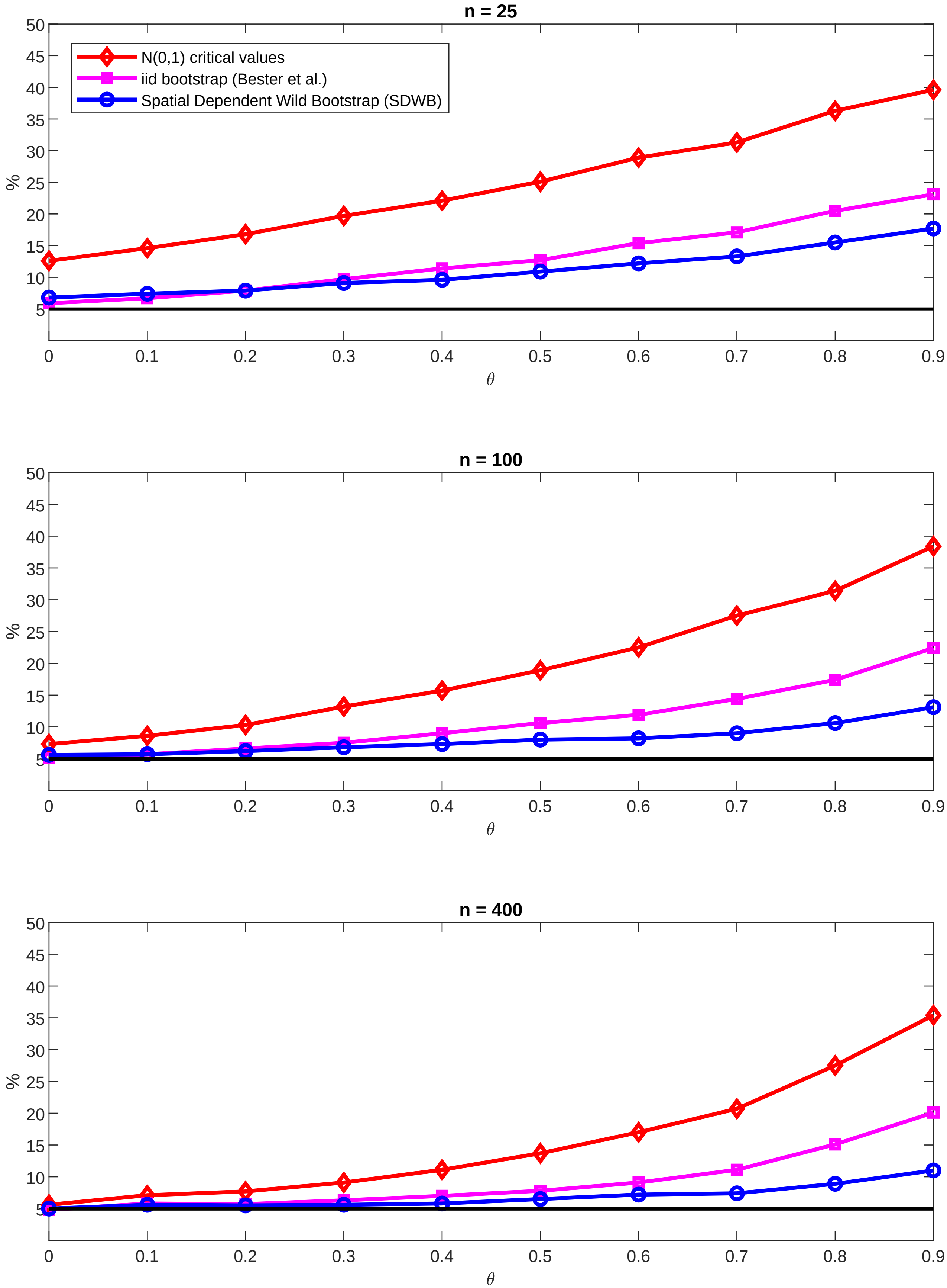


Figure 2. Rejection rate with N(0,2) error in locations

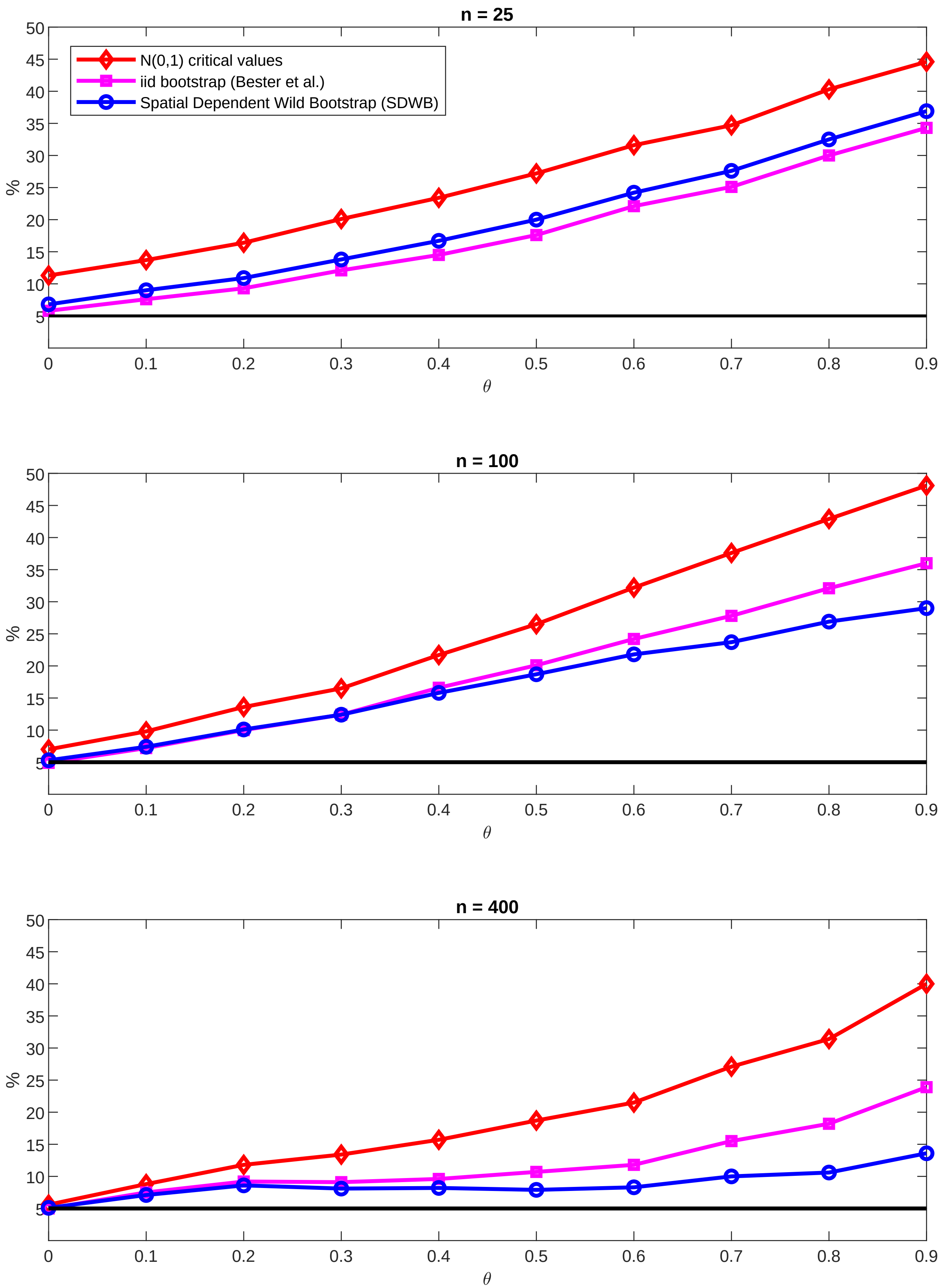


Figure 3. Rejection rate with N(0,4) error in locations

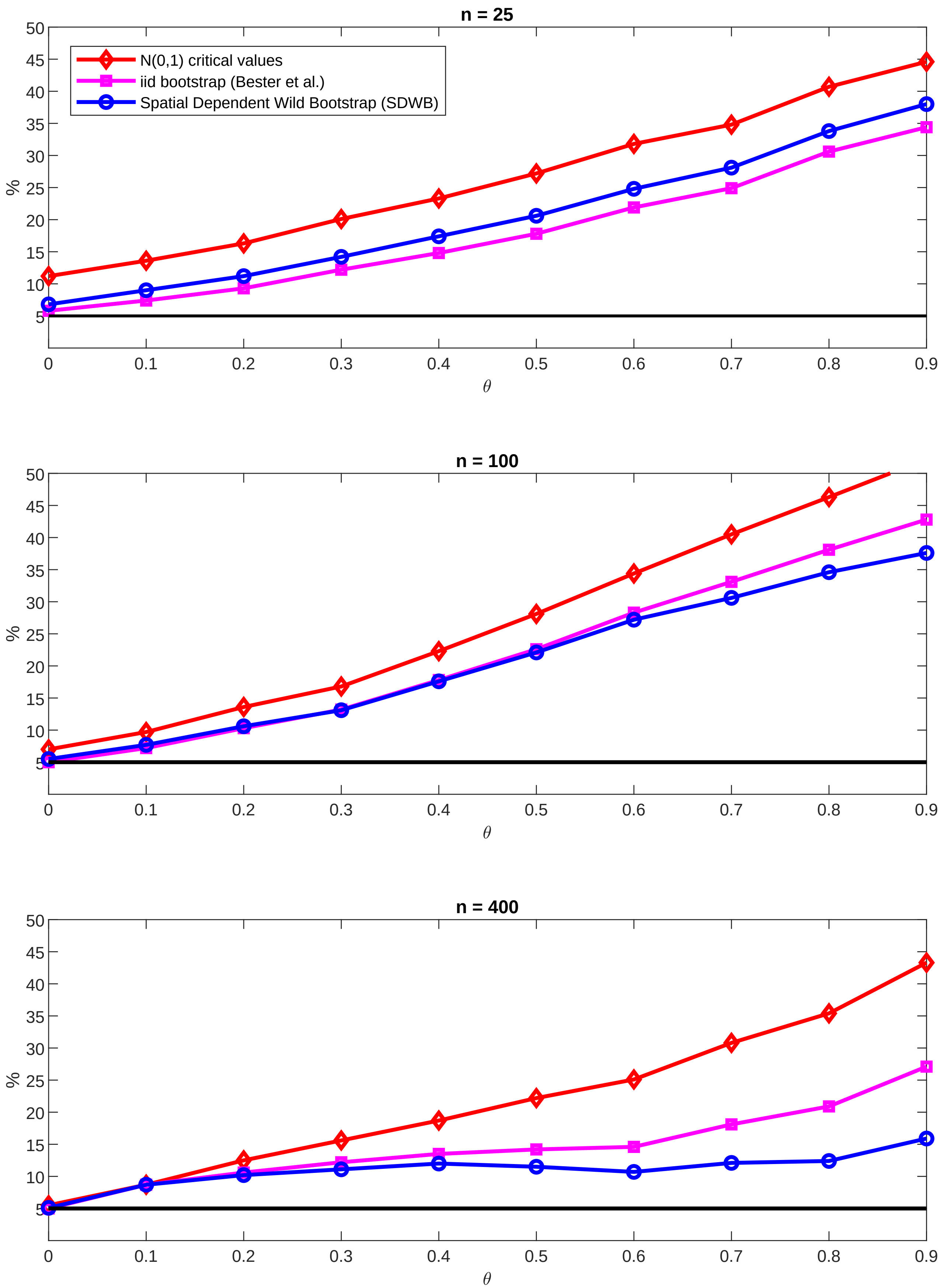


Figure 4. Rejection rate with N(0,10) error in locations

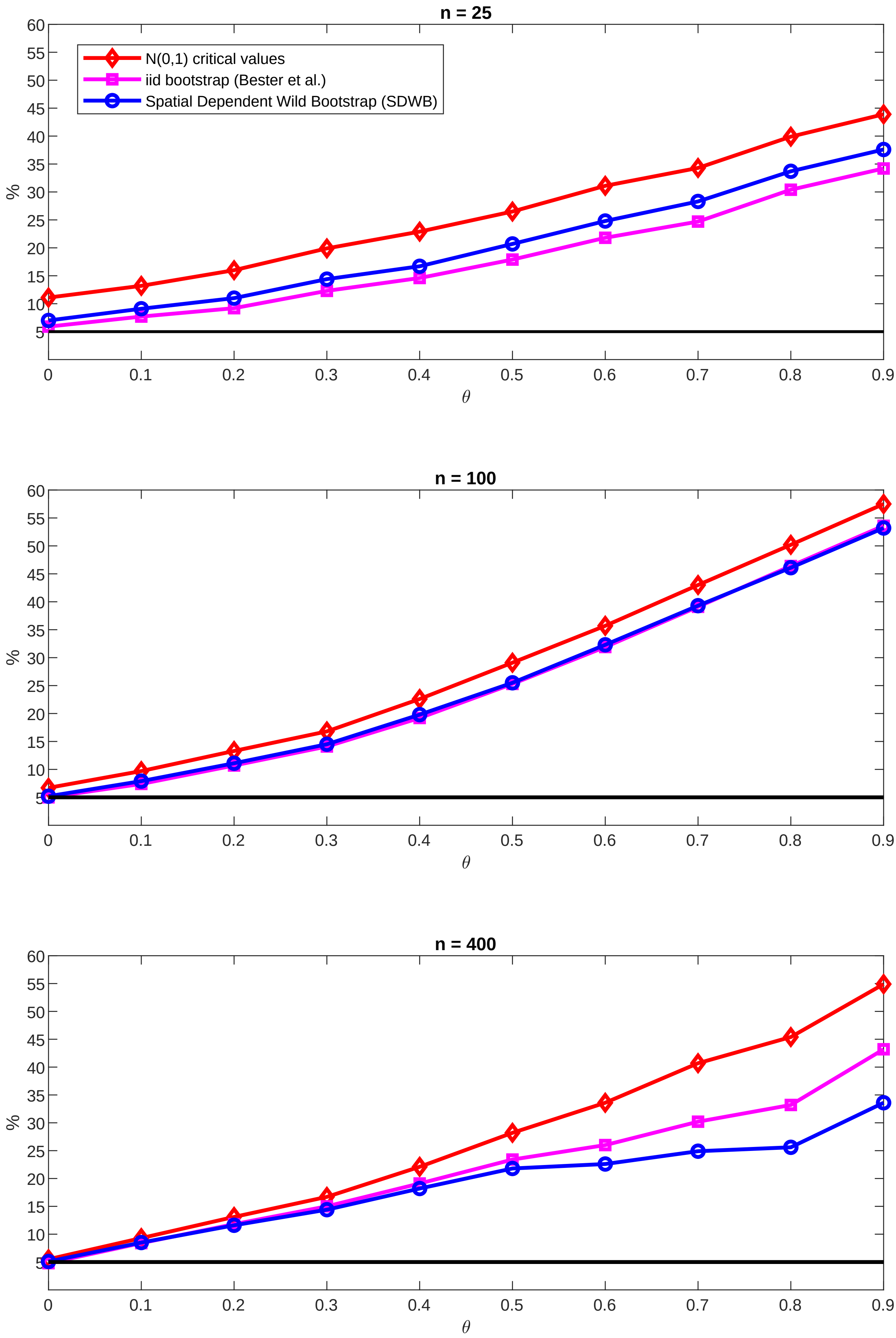
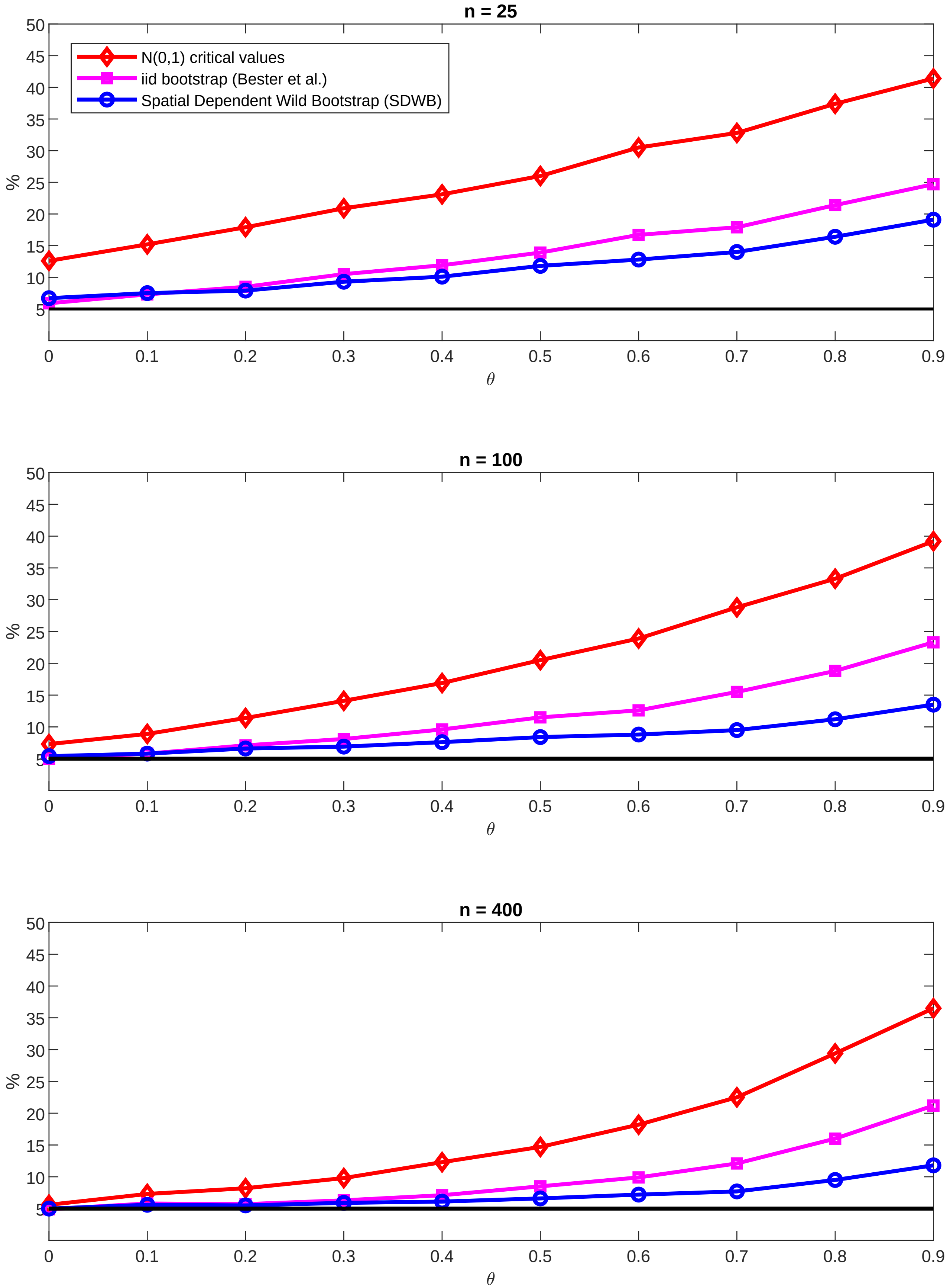


Figure 5. Rejection rate, max distance in DGP, Euclidean in sample



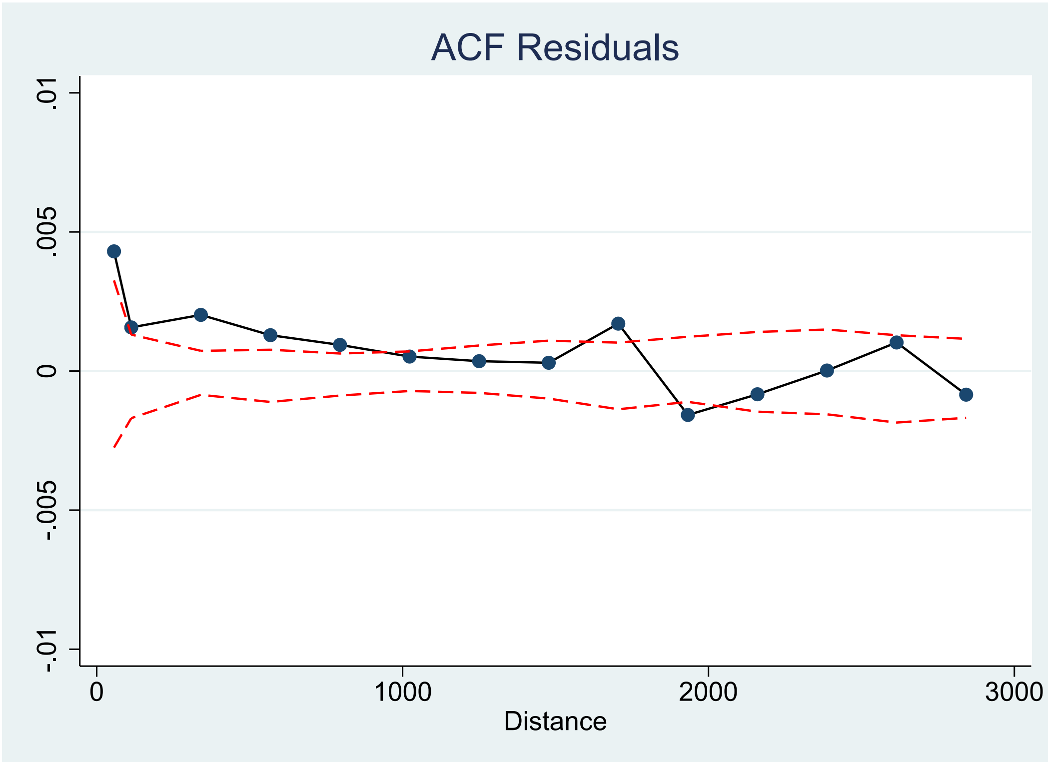


Figure 6. Circles are local average estimates of spatial covariances and dashed lines represent edges of a 90% acceptance region for the null hypothesis of spatial independence. Uniform kernel with tolerance $\delta = 57$ for smallest distance and $\delta = 113$ for all others. Covariances are normalized by dividing by sample variance of residuals.