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

May 8, 2020

Abstract

In this paper, we introduce a method of generating bootstrap samples with unknown patterns of cross sectional dependence which we call the cross sectional dependent wild bootstrap. This method is a cross sectional 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, 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 Donald Andrews, Xiaohong Chen, Yuichi Kitamura, and Matt Webb as well as seminar and conference participants at Queens University, Yale University and the University of York (UK), 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

Economics data are often characterized by dependence and heterogeneity and accounting for these features is important when performing statistical inference. The main goal of this paper is to propose a bootstrap method that is robust to cross sectional dependence and heterogeneity of unknown forms in the context of a linear regression model. Contrary to time series data, where observations are indexed by time, cross sectional observations are often 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 and is not as trivial to apply as the time series block bootstrap.

Our approach in this paper is based on a variation of the wild bootstrap and does not require resampling blocks of observations. Specifically, we consider a linear regression model and propose a residual-based wild bootstrap, where the external random variables used to perturb the residuals are cross sectionally dependent rather than i.i.d., as is usual in the regular wild bootstrap. In particular, the covariance between any two observations is equal to a kernel weight that depends on a distance measure between them. Shao (2010) proposed and studied this method under the name "dependent wild bootstrap" in the time series context, where the distance is given by the time lag between observations. Our extension to the spatial context is not trivial because spatial processes are not naturally ordered on the real line, as we already mentioned. Nevertheless, the method is as easy to apply as in the pure time series context provided we can measure the distance between any pair of observations.

Economic applications with spatial data often provide a set of "economic distance" measures which can be used for spatial robust inference (for instance, spatial heteroskedasticity and autocorrelation (HAC) covariance estimators have been proposed and studied by Conley (1999), Kelejian and Prucha (2007), and Kim and Sun (2011), to name a few). Here, we exploit the availability of these measures to generate bootstrap observations with cross sectional dependence. Different distance measures can be used depending on the application, provided that spatial dependence between two observations decays with the chosen distance. Examples that have been used in the literature include the trade volume or transportation costs between countries (as in Conley and Ligon (2002)) or the similarity of sector-level input and output structures (as in Chen and Conley (2001) and Conley and Dupor (2003)).

We study the first order asymptotic validity of our "cross sectional dependent wild bootstrap" under a set of regularity assumptions that include a linear array representation for the score vector. 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. A special case of this model is the popular spatial autoregressive (SAR) process, where each of the n observations is modelled as a linear transformation of n common i.i.d. innovations, with possibly different weights. 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 imposing stationarity conditions, which are often very restrictive in the spatial context where each unit might have a different number of neighbors. It also avoids having to index observations in a Euclidean space, as required with mixing conditions. Compared to Shao (2010), who assumes a stationary mixing time series, our assumptions allow for heterogenous spatial dependence in higher dimensions than one, but we rule out nonlinear forms of dependence.

One way to generate the cross sectionally dependent external random variables is to multiply a vector of i.i.d. auxiliary random variables by the eigendecomposition of the bootstrap kernel matrix. This matrix contains weights given by the 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, the validity of the cross sectional dependent wild bootstrap requires this matrix to be positive semi-definite. 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 higher dimension than one. A recent paper by Kojevnikov (2019) also considers a modification of the dependent wild bootstrap kernel function that satisfies this condition 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.

We provide a theoretical justification for bootstrap hypothesis tests based on studentized statistics. Studentization requires 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 under the null hypothesis and our simulation results show that the same is true in our context.

An alternative approach to the bootstrap is to use a HAC covariance estimator and rely on 'fixed-b' asymptotic critical values, where the bandwidth parameter is modelled as a fixed proportion of the sample size. This approach was considered by Bester el al. (2016) in the spatial context. 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 bootstrap. We follow Bester et al. (2008) and compute the fixed-b critical values by an i.i.d. bootstrap (see Gonçalves and Vogelsang (2011) for a theoretical justification of this bootstrap approximation in the time series context). The simulations show that our bootstrap method (which accounts for spatial dependence) outperforms the fixed-b i.i.d. bootstrap based approach.

Although we do not explicitly write our model in panel data form, it is important to note that our results can be readily applicable to the panel data setting. When a location structure is available, it suffices to include time as one of the attributes that define the observations' locations. In this case, it is natural to define the kernel function as the product of two kernels, one in the time series dimension and another in the spatial dimension. This approach was considered by Conley (1999) and Kim and Sun (2013) in the context of spatial HAC inference without the bootstrap.

It is also important to note that 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 recently studied by Djogbenou et al. (2019).

Multiple metrics are also easily allowed in our setup. This includes metrics approximating multiway clustering. Our example application in Section 6 illustrates using our methods with firm-level data where correlations across firms arise from both overlap in their local markets and similarity in their technologies.

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 cross sectional 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 illustrates the finite sample performance of the method in comparison to alternative asymptotic-based methods. In Section 6, 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 cross sectional or space-time dependence

We consider the following cross sectional 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

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

The following high level assumption is sufficient to derive the asymptotic distribution of $\sqrt{n}\left(\hat{\beta} - \beta\right)$.

Assumption $\mathcal{A}$

(i) As $n \longrightarrow \infty$,

$$J_n^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i \xrightarrow{d} N\left(0, I_p\right),$$

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

(ii) There is a nonsingular matrix $Q = p\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} x_i x_i'$.

It follows from Assumption $\mathcal{A}$ that

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

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. Under Assumption $\mathcal{A}$, 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). 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.

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 $s_i \in \mathbb{R}^\tau$, 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. We require $d_{ij} \geq 0$, $d_{ii} = 0$, $d_{ij} = d_{ji}$, but not the triangular inequality $d_{ij} \leq d_{ik} + d_{kj}$.

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). Without this location structure, e.g. with non-Euclidean travel costs and time lags jointly describing proximity between panel data observations, distance between any pair of observations can be viewed as depending on two metrics: a spatial/cross sectional metric and time lag. 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 will assume that there is a single distance measure in the remainder of this paper.

The existing spatial HAC literature also allows for the presence of measurement errors in d_{ij} (see e.g. Conley (1999), Conley and Molinari (2007), Kelejian and Prucha (2007) and Kim and Sun (2011)). For simplicity, we ignore this feature when showing the bootstrap validity, but we conjecture that the results follow under similar conditions on the measurement error as those used in this literature (which include in particular bounded errors).

3 Bootstrap inference

3.1 The bootstrap method

The bootstrap method we propose is a residual-based bootstrap, where the bootstrap residuals are obtained by a dependent wild bootstrap that accounts for cross sectional or space-time dependence. Hence, our method can be viewed as a generalization of Shao (2010)'s "dependent wild bootstrap" from the time series to the cross sectional context where dependence can be indexed in higher dimensions than one. 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 constraints 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 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 some choice of external random variables.

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 knowledge of the distance d_{ij} between observations, we generate $\{\eta_i : i = 1, \ldots, n\}$ such that their covariance structure is given by

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

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

$$C_n^* \equiv Var^*\left(\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right)\right) = \hat{Q}_n^{-1} J_n^* \hat{Q}_n^{-1},$$

where $\hat{Q}_n = n^{-1}\sum_{i=1}^{n} x_i x_i'$ and

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

given that $Cov^*\left(\eta_i, \eta_j\right) = K^*\left(\frac{d_{ij}}{d_n^*}\right)$ by (5). This shows that the cross sectional dependent wild bootstrap algorithm induces a bootstrap covariance matrix J_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 distance to i is less than d_n^*. More specifically, let

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

Then, $\mathcal{B}_{i,n}^*$ is the set of pseudo-neighbors that unit i has based on d_n^*, $\ell_{i,n}^*$ is the size of $\mathcal{B}_{i,n}^*$ and ℓ_n^* is its average.

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

Assumption $\mathcal{K}$ (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) $\frac{1}{\ell_n^*}\sup_i \sum_{j \notin \mathcal{B}_{i,n}^*}\left|K^*\left(\frac{d_{ij}}{d_n^*}\right)\right| = O(1)$. (iii) The matrix $\mathbb{K}_n^* = \left[K^*\left(d_{ij}/d_n^*\right)\right]_{i,j=1}^n$ is positive semi-definite for all n.

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. Condition $\mathcal{K}$ (ii) is automatically satisfied by truncated kernels for which $K^*(z) = 0$ for $|z| \geq 1$, but allows for kernels that do not truncate provided the tails of K^* decay sufficiently fast. For instance, with locations indexed on the line (such as a time series), 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. For higher dimensional spaces, it is harder to provide a sufficient condition that only involves the kernel function. However, if we map locations into a two-dimensional lattice and let d_{ij} be the Euclidean distance, a sufficient condition for part (ii) is that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|K^*\left(\sqrt{x^2+y^2}\right)\right| dx dy < \infty$ 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 $\mathbb{K}_n^* = \left[K^*\left(d_{ij}/d_n^*\right)\right]_{i,j=1}^n$ to be 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 be positive semi-definite. 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:

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

where $\nu \geq (\tau + 1)/2$. 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:

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

An example kernel in this class is

$$K^*(x) = c_0 \exp(-c_1 x^{2\gamma}), \tag{10}$$

with curvature parameter $\gamma \in (0, 1]$, and scaling parameters c_0, and c_1. Given that $K^*(0) = 1$, c_0 must be equal to 1, which amounts to the Gaussian kernel. In the simulations below, we also choose $\gamma = c_1 = 1$. 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; below we discuss one strategy to overcome this issue.

Under Assumption $\mathcal{K}$, $\mathbb{K}_n^*$ 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 $\left(\Phi_n' \Phi_n = I_n\right)$. 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 is as follows. Letting $L_n = \Phi_n \Lambda_n^{1/2}$, we set

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

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 and m 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 $L_n = diag\left(L_{g,n}\right)$ 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 $\eta = \left(\eta_1', \ldots, \eta_G'\right)'$, where for each g cluster, $\eta_g = \left(\varepsilon_g, \ldots, \varepsilon_g\right)'$ 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 $\mathcal{B}_1$

(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 $a = 1, \ldots, p$.

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) 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). 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 $\{V_i\}$.

We can view (12) as the cross sectional analogue of a time series linear process, where $r_{il}^{(a)}$ is of the form $r_{i-l}^{(a)}$ (in a time series setting, we would write $V_i^{(a)} = \sum_{j=0}^{\infty} r_j^{(a)} e_{i-j} = \sum_{l=-\infty}^{t} r_{i-l}^{(a)} e_l$). We can also think of (12) as an extension of some popular SAR models, where $V_i^{(a)}$ is of the form $V_i^{(a)} = \sum_{l=1}^{n} r_{il}^{(a)} e_l$, i.e. $V_i^{(a)}$ is a linear function of n innovations e_l, $l = 1, \ldots, n$ rather than of an infinite set of innovations.

[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.

To interpret the absolute summability conditions in part (iii) of Assumption $\mathcal{B}_1$, it is useful to think of $r_{il}^{(a)}$ as the response of $V_i^{(a)}$ to a unit shock in e_l. The absolute summability condition over i then implies that the aggregate absolute response $\sum_{i=1}^n \left| r_{il}^{(a)} \right|$ to a unit shock in e_l is bounded. This condition is automatically satisfied if only a finite number of units responds to a shock in e_l. The absolute summability condition of $r_{il}^{(a)}$ over l ensures that the response of V_i to a common set of shocks at all locations (i.e. $\Delta e_l = 1$ for all l) is finite, in absolute value. The absolute summability of the coefficients $\left\{ r_{il}^{(a)} \right\}$ across i and l ensures that the sequence $\{V_i\}$ is weakly dependent, which is needed for $\sqrt{n}$-convergence of the OLS estimator.

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 J_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 $\mathcal{B}_2$ There exists a constant $C_{q_0^*} < \infty$ such that

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left\| E\left(V_i V_j' \right) \right\| d_{ij}^{q_0^*} < C_{q_0^*},$$

for all n, where $\|\cdot\|$ denotes the Euclidean norm of a matrix.

Given our Assumption $\mathcal{B}_1$, we can rewrite Assumption $\mathcal{B}_2$ as a function of the coefficients $\left\{ r_{il}^{(a)} \right\}$ as

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left| \sum_{l=1}^\infty r_{il}^{(a)} r_{jl}^{(b)} \right| d_{ij}^{q_0^*} < C_{q_0^*},$$

for $a, b = 1, \ldots, p$, and all n.

Assumption $\mathcal{B}_2$ requires the degree of cross sectional dependence between V_i and V_j to decrease as a function of the 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: $\sum_{j=-\infty}^{+\infty} \left\| E\left(V_t V_{t+j}' \right) \right\| |j|^{q_0^*} < \infty$ (see Andrews, 1991, eq. (3.4)).

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

Assumption $\mathcal{B}_3$ For all i, $\ell_{i,n}^* \leq c\ell_n^*$ for some constant $c > 0$.

Assumption $\mathcal{B}_3$ 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 $i \in \left[d_n^*, n-d_n^*+1\right]$ if d_n^* is an integer) and $\ell_n^* > d_n$, and this implies that $\ell_{in}^* \leq 2\ell_n^*$ for all i. Thus, Assumption $\mathcal{B}_3$ is automatically satisfied in this case. We can also see that ℓ_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 ℓ_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 ℓ_n^* is proportional to $d_n^{*\tau}$, where τ is the dimension of the space. When $\tau = 1$, this implies that ℓ_n^* is proportional to d_n^*, as discussed above, whereas for $\tau = 2$ we obtain $\ell_n^* = \alpha d_n^{*2}$ for some bounded constant α. Since ℓ_n^* (and d_n^*) plays the role of the bandwidth parameter in the usual time series HAC literature, we will let $\ell_n^* \to \infty$ 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 ℓ_n^* (and hence a larger d_n^*) reduces the asymptotic bias but increases the variance of the bootstrap variance estimator at the rate $O\left(\ell_n^*/n\right)$.

The final assumption we need is a moment condition on the regressors. Note that Assumption $\mathcal{B}_1$ implies that a similar moment condition holds for the scores V_i, i.e. $E\left\|V_i\right\|^4 \leq M$.

Assumption $\mathcal{B}_4$ There exists a constant $M < \infty$ such that $E\left\|x_i\right\|^4 \leq M$.

Theorem 3.1 *Suppose Assumptions $\mathcal{A}$, $\mathcal{K}$ and $\mathcal{B}_1$-$\mathcal{B}_4$ hold. If $d_n^*, \ell_n^* \to \infty$ such that $\ell_n^*/n^{1/2} = o(1)$ 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 $J_n^* - J_n \to^P 0$ 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 ℓ_n^* is $\ell_n^*/\sqrt{n} \to 0$, which is stronger than the rate $\ell_n^*/n \to 0$ used for showing the consistency of bootstrap (spatial HAC) variance estimator J_n^* (cf. Theorem 1 of Kim and Sun, 2011, and the proof of Lemma

A.1). The stronger rate condition on ℓ_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 ℓ_n^* 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 $\ell_n^*/n^{1/2} \to 0$ as $n \to \infty$ to verify the Lyapunov's condition. Note that despite the 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

$$\begin{aligned} n^{-1}\sum_{k=1}^n \omega_k^2 &= V'\left(n^{-1}\sum_{k=1}^n \lambda_k\phi_k\phi_k'\right)V = n^{-1}V'\Phi_n\Lambda_n\Phi_n'V \\ &= n^{-1}V'\mathbb{K}_n^*V = n^{-1}\sum_{i,j=1}^n K^*\left(\frac{d_{ij}}{d_n^*}\right)V_iV_j' = J_{0,n}^*, \end{aligned}$$

a spatial HAC variance estimator. Although this estimator is infeasible as it is based on $V_i = x_iu_i$, we show in Lemma A.1 (ii) that the difference between this estimator and its feasible version J_n^* is asymptotically negligible under our assumptions.

As explained above, the relationship between ℓ_n^* and d_n^* depends in general on the dimension τ of the space indexing the locations. If we assume that ℓ_n^* is proportional to $d_n^{*\tau}$, as in Kim and Sun (2011), our rate condition on ℓ_n^* is equivalent to $d_n^*/n^{1/(2\tau)} \to 0$. The higher the space dimension, the slower d_n^* is allowed to grow as a function of n. An important question is the optimal choice of ℓ_n^* and d_n^*. Kim and Sun (2011) discuss the optimal bandwidth choice of these bandwidth parameters from the viewpoint of minimizing the MSE of the spatial HAC estimator. Their Corollary 1 implies that the optimal choice of d_n^* under this criterion is to select $d_{n,opt}^*$ proportionally to $n^{1/\left(2q_0^*+\tau\right)}$, where q_0^* is the Parzen's exponent of the bootstrap kernel. For $d_{n,opt}^*$ to satisfy the condition $d_{n,opt}^*/n^{1/(2\tau)} \to 0$, we must require $q_0^* > \frac{\tau}{2}$. In the one-dimensional case, any bootstrap kernel with $q_0^* > \frac{1}{2}$ satisfies this condition, but this imposes a more stringent restriction as τ increases. For $\tau = 2$, for example, this condition requires the choice of a kernel such that $q_0^* > 1$. Thus, in this case, the optimal MSE bandwidth for kernels with $q_0^* = 1$ such as Kelejian and Prucha (2007) kernel in (8) (which produces a positive semi-definite $\mathbb{K}_n^*$ matrix with Euclidean distance) is not compatible with the rate condition in Theorem 3.1. A slower rate of d_n^* is needed for bootstrap consistency for these kernels. On the other hand, the Gaussian kernel, which is another example that produces a positive definite $\mathbb{K}_n^*$ matrix, meets this condition when $\tau = 2$ because $q_0^* = 2$.

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

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

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

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

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^*. The possibility of using different choices for these quantities is important because the bootstrap data generating process requires stronger assumptions on $K^*(\cdot)$ (and d_n^*). In particular, as we discussed before, we need to ensure that the induced bootstrap covariance matrix $\mathbb{K}_n^*$ is positive semi-definite. This requirement is not necessary for establishing the consistency of $\hat{J}_n$ although it converges to a positive semi-definite matrix.

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 let $\ddot{\beta}$ denote 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

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

where

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

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 some 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:

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

and make the following assumption.

Assumption $\mathcal{W}$ (i) $K : \mathbb{R} \to [-1, 1]$ satisfies $K(0) = 1$, and $K(z) = K(-z)$ for all $z \in \mathbb{R}$. (ii) $\frac{1}{\ell_n} \sup_i \sum_{j \notin B_{i,n}} K\left(\frac{d_{ij}}{d_n}\right) = O(1)$. (iii) Let q_0 denote the Parzen characteristic exponent of $K(z)$. There exists a constant $C_{q_0} < \infty$ such that

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\| E\left(V_i V_j' \right) \right\| d_{ij}^{q_0} < C_{q_0},$$

for all n. (iv) For all i, $\ell_{i,n} \leq c \ell_n$ for some constant $c > 0$.

Assumption $\mathcal{W}$ imposes conditions on K and d_n which are similar to those imposed on K^* and d_n^* by Assumptions $\mathcal{K}$ (i) and (ii), and $\mathcal{B}_2 - \mathcal{B}_3$. The main difference is that we do not require the kernel function K to be positive definite. Hence, the spatial HAC variance estimator $\hat{J}_n$ used to studentize the Wald test statistic is not guaranteed to be positive semi-definite in finite samples. In practice, it is possible to induce a positive semi-definite variance estimator by replacing the negative eigenvalues of $\hat{J}_n$ by zero, as suggested by Politis (2011) (see also McMurry and Politis, 2010). We will suggest a similar eigenvalue modification to solve the problem of a non positive semi-definite bootstrap kernel matrix $\mathbb{K}_n^*$ in the next section.

Theorem 4.1 *Suppose Assumptions $\mathcal{A}$, $\mathcal{K}$, $\mathcal{W}$ and $\mathcal{B}_1$-$\mathcal{B}_4$ hold. If $E^* |v_i|^4 < M$ and $d_n, \ell_n \to \infty$ and $d_n^*, \ell_n^* \to \infty$ as $n \to \infty$ such that $\ell_n / n = o(1)$ and $\ell_n^* / n^{1/2} = o(1)$, then, under H_0, as $n \to \infty$,*

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

Assumption $\mathcal{W}$ (along with our remaining assumptions) is used to show the consistency of $\hat{J}_n^*$ for $J_n^* = Var^*\left(n^{-1/2} \sum_{i=1}^{n} x_i u_i^* \right)$. See Lemma A.3. Since J_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 $\mathcal{W}_n^*$ mimics the null distribution of $\mathcal{W}_n$ 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 $\mathcal{W}_n$. 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 $C_n = Q^{-1} J_n Q^{-1}$, 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 $\mathcal{W}_n$ 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.

5 Monte Carlo Simulations

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

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

where $\alpha = \beta = \gamma = 1$. The observations lie on a square lattice and are governed by:

$$u_j = \sum_{|a| \leq 2} \theta^{|a|} \psi_{j+a}; \quad x_j = \sum_{|a| \leq 2} \theta^{|a|} u_{1,j+a} \quad z_j = \sum_{|a| \leq 2} \theta^{|a|} u_{2,j+a}$$

where $\left(\psi_j, u_{1j}, u_{2j}\right)'$ are drawn independently across locations from an i.i.d. $N(0, I_3)$ distribution and $|a| = \max(d_1, d_2)$, where d_i is the absolute distance on the lattice. In other words, observations i and j are correlated if the maximum of the horizontal and vertical distance between them is at most 2. In that case, the weight given to each of the innovations ψ_i, u_{1i} and u_{2i} is given by θ raised to that maximum distance so it can take only 3 values, either 1 (for $j = i$), θ (for observations that lie on a square of length 2 centered at location j on the lattice) and θ^2 for observations that lie on a square of length 4 centered at j. Thus, 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:

$$T_n = \frac{\sqrt{n}\left(\hat{\beta} - 1\right)}{s\left(\hat{\beta}\right)}$$

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

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

using the Gaussian kernel, the data-based optimal bandwidth choice in Kim and Sun (2011), and Euclidean distance. This statistic is the square root of the statistic $\mathcal{W}_n$ defined in (14). We also considered the use of the maximum distance instead of Euclidean, but since the results were very similar, we decided not to report them to save space.

We compare this t statistic to 3 different critical values. The first one is the critical value from the standard normal distribution as suggested by (1). The second critical value is obtained from the fixed-b asymptotics in Bester et al. (2016) that assumes that the bandwidth d_n is a fixed proportion of the sample size. However, since this distribution is complicated, it was implemented using the i.i.d. pairs bootstrap as suggested in the working paper version (Bester et al., 2008). Finally, the last critical value is obtained using our cross sectional dependent wild bootstrap 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 Rademacher 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. We have also computed rejection rates with Parzen and Kelejian and Prucha (2007) kernels. The Parzen kernel does not guarantee positive semi-definiteness of $\mathbb{K}_n^*$ which makes it unsuitable as a bootstrap variance. We adjust it by replacing its negative eigenvalues by 0 and define:

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

where the i^{th} element on the diagonal is $\lambda_i^+ = max(\lambda_i, 0)$. On the other hand, the Kelejian and Prucha (2007) kernel has $q_0^* = 1$, which means that the MSE-optimal bandwidth expands too quickly to satisfy our conditions. We therefore implement it with the same bandwidth as with the Parzen kernel. As the two sets of results were nearly identical to those with the Gaussian kernel with MSE-optimal bandwidth choice, we only report those with the Gaussian.

Figure 1 reports the rejection rates for the three sets of critical values under the null hypothesis of $\beta = 1$. Each panel presents results for different sample sizes and includes three lines: the red line with diamonds corresponds to the rejection rate using the critical value from the standard normal distribution, the pink line with squares gives rejection rates using the Bester et al. (2016) (BCHV) distribution that assumes that the bandwidth d_n is a fixed proportion of the sample size, and the blue line with circles reports rejection rates for our cross sectional dependent wild bootstrap.

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 47.9% instead of 5%. This is reduced to 17.1% for $n = 400$.

The BCHV critical values perform much better and reduce the size distortions considerably. Again, for $\theta = .5$, the rejection rate is 33.4% for $n = 25$ and 10.3% for $n = 400$. Finally, the cross sectional dependent wild bootstrap (CSDW) performs similarly to BCHV when $n = 25$ but gives rejection rates closer to the nominal level for the two larger values of n. For $\theta = .5$ and $n = 400$, its rejection rate is 7.0%.

We conclude from these experiments that the cross sectional wild dependent bootstrap removes a large fraction of the size distortions associated with the use of the normal asymptotic critical values, and for values of n that are large enough (in our experiments $n \geq 100$), it outperforms the fixed-b asymptotics. Its superiority is also more pronounced with stronger spatial dependence (larger values of θ). Finally, our results seem robust to the choice of kernel and distance measure.

6 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.[2] 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.[3]

Specifically, we examine the relationship between manufacturing firms' sales growth and the level of expo-

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

[3] 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.

sure to import activity in their local markets. We estimate that relation using a cross-section of firms and the following specification:

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

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.[4] 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, capital intensity measured as the ratio of the book value of tangible assets to firm's total payroll, and the inverse of production worker share measured 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 cross sectional dependence will be present in this cross section of firms due to two main factors. 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 cross sectional 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.[5] 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.

[4] 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.

[5] We ignore elevation and the curvature of the Earth.

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

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

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 d_{ij} is equal to physical distance, providing at least some sense of units.

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

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

The choice of bandwidth or tuning parameter is clearly important. We present results in Table 1 using an ad hoc methodology for choosing this parameter. To get a ballpark range for d_n and d_n^*, we computed autocovariance functions for residuals as a function of d_{ij} nonparametrically using a local average, following Conley and Dupor (2001) and Conley and Topa (2002). For the range of distances with adequate precision, estimated autocovariances are decreasing and significantly positive until about 500 units and then borderline significant until about 1500 units. We chose 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 comparison: classical OLS, heteroskedasticity consistent (labelled White), clustered at the ECR-level, clustered at the 3-digit industry-level, and a Spatial HAC estimator (Conley 1999) using the same kernel and bandwidth as our bootstrap.[6]

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, the cross sectional dependent wild bootstrap CIs are the widest but for the coefficient for final goods, Cls using clustering on ECR are substantially larger than the rest.

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

[6]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.

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	Cluster Industry	Spatial HAC	Spatial Wild Bootstrap
Import Pen. Intermediate	4.342	1.666	1.799	1.833	1.696	1.838	2.039
Import Pen. Final	-3.033	1.584	1.454	2.164	1.344	1.557	1.647
Import Pen. Capital	-.134	1.307	1.027	1.271	1.199	1.013	1.041
Log real sales	.013	0.017	0.016	0.014	0.031	0.021	0.024
Age	-.003	0.004	0.004	0.003	0.004	0.004	0.005
Capital/Labor	.028	0.021	0.030	0.032	0.037	0.028	0.032
Total/Production Payroll	.00004	0.0005	0.0002	0.0002	0.0003	0.0002	0.0001
Constant	-.265	0.191	0.197	0.192	0.455	0.280	0.313

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 percentile-t confidence interval using 2,000 Bootstrap repetitions.

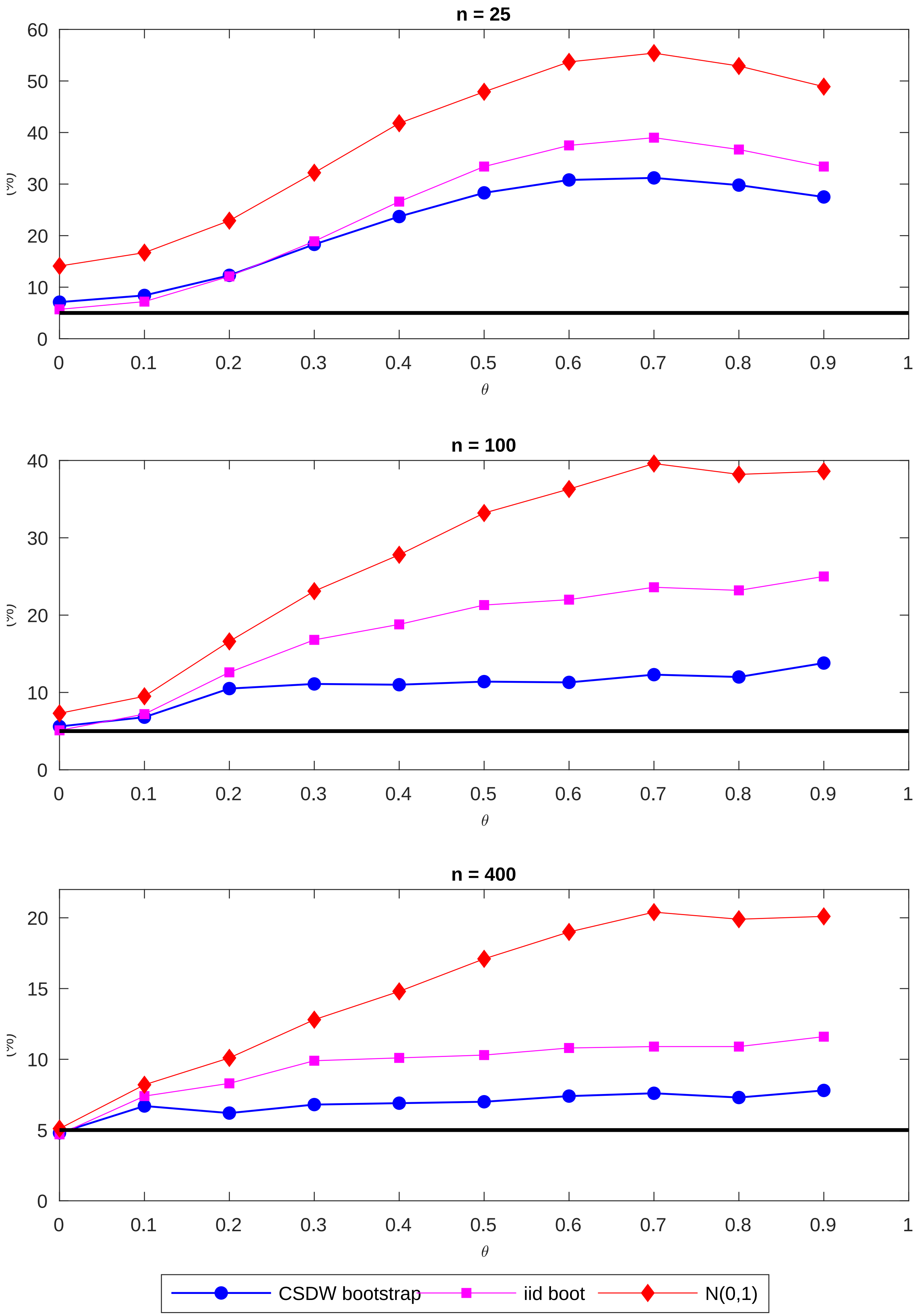
and sometimes smaller than White CIs across the three coefficients of interest. Both the spatial HAC and cross sectional dependent wild bootstrap attempt to allow for both types of correlation.

Finally, there is evidence that using the cross sectional 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 10% to 14%, possibly reflecting the greater robustness of the bootstrap intervals to finite sample deviations from the normal distribution.

7 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.

Figure 1. Size of 5% tests, Euclidean distance, Gaussian kernel



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 $(\Omega, \mathcal{F}, P)$). 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 $\delta > 0$, $P^*\left(\left|T_n^*\right| > \delta\right) = o_P(1)$. We write $T_n^* = O_{P^*}(1)$, in prob-P, when for all $\delta > 0$ there exists $M_\delta < \infty$ such that $\lim_{n\to\infty} P\left[P^*\left(\left|T_n^*\right| > M_\delta\right) > \delta\right] = 0$. By Markov's inequality, this follows if $E^*\left|T_n^*\right|^q = O_P(1)$ 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. $E^*\left(f\left(T_n^*\right)\right) \to^P E\left(f(D)\right)$ for all bounded and uniformly continuous functions f.

A.1 Auxiliary lemmas

Define

$$J_{0n}^* = Var^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i\eta_i\right)$$

and note that J_{0n}^* differs from $J_n^* = Var^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{V}_i\eta_i\right)$ by replacing $\hat{V}_i$ with V_i.

The following lemma establishes the consistency of J_{0n}^* and J_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 $\mathcal{W}_n^*$.

Lemma A.1 *Suppose that the conditions of Theorem 3.1 hold. Then (i)* $J_{0n}^* - J_n \to^P 0$ *and (ii)* $J_n^* - J_n \to^P 0$.

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

Lemma A.2 *Suppose Assumption* $\mathcal{K}$ *holds. Under Assumption* $\mathcal{B}_1$*, for any pair* (i,j),

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

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* $\mathbb{K}_n^* = \left(K^*\left(\frac{d_{ij}}{d_n^*}\right)\right)_{i,j=1,\ldots,n}$.

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

$$J_{0n}^* - J_n = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n K^*\left(\frac{d_{ij}}{d_n^*}\right)\left[V_iV_j - E\left(V_iV_j\right)\right] + \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\left(K^*\left(\frac{d_{ij}}{d_n^*}\right) - 1\right)E\left(V_iV_j\right) \equiv b_1 + b_2.$$

For b_1, since $E(b_1) = 0$, it suffices to prove that $Var(b_1) = o(1)$. We have

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

Adding and subtracting appropriately in (16), we can bound $Var(b_1)$ by

$$Var(b_1) \leq b_{11} + b_{12} + b_{13},$$

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

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

We can show that b_{12} and b_{13} are both of order $O\left(\frac{\ell_n^*}{n}\right)$ whereas $b_{11} = O\left(\frac{1}{n}\right)$. Thus, $Var(b_1) = o(1)$ under our assumptions provided $\frac{\ell_n^*}{n} = o(1)$. 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 $C_1 = O(1)$ in the proof of Theorem 3.1, so we omit the details here). We can write

$$\begin{aligned} b_{12} &= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1\in\mathcal{B}_{i_1,n}^*}\sum_{i_2=1}^{n}\sum_{j_2\in\mathcal{B}_{i_2,n}^*} K^*\left(\frac{d_{i_1 j_1}}{d_n^*}\right)K^*\left(\frac{d_{i_2 j_2}}{d_n^*}\right)\left|E\left(V_{i_1}V_{i_2}\right)E\left(V_{j_1}V_{j_2}\right)\right| \\ &+\frac{2}{n^2}\sum_{i_1=1}^{n}\sum_{j_1\notin\mathcal{B}_{i_1,n}^*}\sum_{i_2=1}^{n}\sum_{j_2\in\mathcal{B}_{i_2,n}^*} K^*\left(\frac{d_{i_1 j_1}}{d_n^*}\right)K^*\left(\frac{d_{i_2 j_2}}{d_n^*}\right)\left|E\left(V_{i_1}V_{i_2}\right)E\left(V_{j_1}V_{j_2}\right)\right| \\ &+\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1\notin\mathcal{B}_{i_1,n}^*}\sum_{i_2=1}^{n}\sum_{j_2\notin\mathcal{B}_{i_2,n}^*} K^*\left(\frac{d_{i_1 j_1}}{d_n^*}\right)K^*\left(\frac{d_{i_2 j_2}}{d_n^*}\right)\left|E\left(V_{i_1}V_{i_2}\right)E\left(V_{j_1}V_{j_2}\right)\right| \equiv b_{12}^{(1)} + b_{12}^{(2)} + b_{12}^{(3)}. \end{aligned}$$

We have that

$$\begin{aligned} b_{12}^{(1)} &\leq \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\left|E\left(V_{i_1}V_{i_2}\right)\right|\sum_{j_1\in\mathcal{B}_{i_1,n}^*}\sum_{j_2\in\mathcal{B}_{i_2,n}^*}\left|E\left(V_{j_1}V_{j_2}\right)\right| \\ &\leq \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\left|E\left(V_{i_1}V_{i_2}\right)\right|\sum_{j_1\in\mathcal{B}_{i_1,n}^*}\sum_{j_2=1}^{n}\left|E\left(V_{j_1}V_{j_2}\right)\right| \\ &\leq \underbrace{\left(\sup_c\sum_{j_2=1}^{n}\left|E\left(V_cV_{j_2}\right)\right|\right)}_{\leq\Delta}\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\left|E\left(V_{i_1}V_{i_2}\right)\right|\underbrace{\sum_{j_1\in\mathcal{B}_{i_1,n}^*}1}_{=\ell_{i_1,n}^*} \\ &\leq \Delta\frac{1}{n}\frac{1}{n}\sum_{i_1=1}^{n}\underbrace{\left(\sup_{i_1}\sum_{i_2=1}^{n}\left|E\left(V_{i_1}V_{i_2}\right)\right|\right)}_{\leq\Delta}\ell_{i_1,n}^* \leq \Delta^2\frac{1}{n}\left(\frac{1}{n}\sum_{i_1=1}^{n}\ell_{i_1,n}^*\right) = \Delta^2\frac{\ell_n^*}{n} = O\left(\frac{\ell_n^*}{n}\right), \end{aligned}$$

where we have used the fact that $\sup_i \sum_{j=1}^n \left|E\left(V_i V_j\right)\right| \leq \Delta$ under Assumption $\mathcal{B}_1$. Similarly,

$$\begin{aligned} b_{12}^{(2)} &\leq \frac{2}{n^2} \sum_{i_1=1}^n \sum_{j_1 \notin \mathcal{B}_{i_1,n}^*} \left|K^*\left(\frac{d_{i_1 j_1}}{d_n^*}\right)\right| \underbrace{\left(\sup_c \sum_{i_2=1}^n \left|E\left(V_c V_{i_2}\right)\right|\right)}_{\leq \Delta} \underbrace{\left(\sup_c \sum_{j_2=1}^n \left|E\left(V_c V_{j_2}\right)\right|\right)}_{\leq \Delta} \\ &\leq 2\Delta^2 \frac{\ell_n^*}{n} \left(\frac{1}{\ell_n^*} \sup_c \sum_{j_1 \notin \mathcal{B}_{c,n}^*} \left|K^*\left(\frac{d_{c j_1}}{d_n^*}\right)\right|\right) = O\left(\frac{\ell_n^*}{n}\right). \end{aligned}$$

Since a similar argument implies that $b_{12}^{(3)} = O\left(\frac{\ell_n^*}{n}\right)$, we conclude that $b_{12} = O\left(\frac{\ell_n^*}{n}\right)$.

For b_2, note that

$$\begin{aligned} |b_2| &= \left|\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left(K^*\left(\frac{d_{ij}}{d_n^*}\right) - 1\right) E\left(V_i V_j\right)\right| \\ &\leq \frac{1}{(d_n^*)^{q_0^*}} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left\|E\left(V_i V_j\right)\right\| d_{ij}^{q_0^*} \underbrace{\left|\frac{K^*\left(\frac{d_{ij}}{d_n^*}\right) - 1}{\left(\frac{d_{ij}}{d_n^*}\right)^{q_0^*}}\right|}_{\to K_{q_0^*}^* \text{ as } d_n^* \to \infty} \\ &= \frac{1}{(d_n^*)^{q_0^*}} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left\|E\left(V_i V_j\right)\right\| d_{ij}^{q_0^*} \left(K_{q_0^*}^* + o(1)\right) \\ &\leq \frac{1}{(d_n^*)^{q_0^*}} C_{q_0^*} K_{q_0^*}^* + o(1) = o(1) \text{ as } d_n^* \to \infty, \end{aligned}$$

where we use Assumption $\mathcal{B}_2$ to bound $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left\|E\left(V_i V_j\right)\right\| d_{ij}^{q_0^*}$. Hence, $b_2 = o(1)$, completing the proof of part (i).

For part (ii), given (i) it suffices to show that $J_n^* - J_{0n}^* = o_P(1)$. Since $x_i \hat{u}_i = x_i \left[u_i + x_i\left(\beta - \hat{\beta}\right)\right]$, we can write

$$J_n^* - J_{0n}^* = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n K^*\left(\frac{d_{ij}}{d_n^*}\right) x_i x_j \left(\hat{u}_i \hat{u}_j - u_i u_j\right) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n K^*\left(\frac{d_{ij}}{d_n^*}\right) x_i x_j \left[x_i x_j \left(\beta - \hat{\beta}\right)^2 + 2 x_j u_i \left(\beta - \hat{\beta}\right)\right] \equiv c_1 + c_2.$$

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

$$c_1 = O_P\left(\frac{1}{n}\right) \left(\frac{1}{n} \sum_{i=1}^n \sum_{j \in \mathcal{B}_{i,n}^*} K^*\left(\frac{d_{ij}}{d_n^*}\right) x_i^2 x_j^2 + \frac{1}{n} \sum_{i=1}^n \sum_{j \notin \mathcal{B}_{i,n}^*} K^*\left(\frac{d_{ij}}{d_n^*}\right) x_i^2 x_j^2\right) = O_P\left(\frac{\ell_n^*}{n}\right),$$

because by Markov's inequality, as $\Delta \to \infty$,

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n \sum_{j \in \mathcal{B}_{i,n}^*} K^*\left(\frac{d_{ij}}{d_n^*}\right) x_i^2 x_j^2\right| > \Delta \ell_n^*\right) \leq \frac{1}{\Delta} \frac{1}{\ell_n^* n} \sum_{i=1}^n \sum_{j \in \mathcal{B}_{i,n}^*} E\left(x_i^2 x_j^2\right) \leq \frac{1}{\Delta} \frac{1}{\ell_n^* n} \sum_{i=1}^n \sum_{j \in \mathcal{B}_{i,n}^*} M \to 0,$$

and

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{j\notin\mathcal{B}_{i,n}^{*}}K^{*}\left(\frac{d_{ij}}{d_{n}^{*}}\right)x_{i}^{2}x_{j}^{2}\right|>\Delta\ell_{n}^{*}\right)\leq\frac{1}{\Delta}\frac{1}{\ell_{n}^{*}n}\sum_{i=1}^{n}\sum_{j\notin\mathcal{B}_{i,n}^{*}}\left|K^{*}\left(\frac{d_{ij}}{d_{n}^{*}}\right)\right|E\left(x_{i}^{2}x_{j}^{2}\right)\leq\frac{M}{\Delta}\left(\frac{1}{\ell_{n}^{*}}\sup_{c}\sum_{j\notin\mathcal{B}_{c,n}^{*}}\left|K^{*}\left(\frac{d_{cj}}{d_{n}^{*}}\right)\right|\right)\to 0,$$

given Assumption $\mathcal{K}$(ii). For c_2,

$$\begin{aligned}c_{2} &= \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K^{*}\left(\frac{d_{ij}}{d_{n}^{*}}\right)x_{i}u_{i}x_{j}^{2}\left(\beta-\hat{\beta}\right)=O_{P}\left(\frac{1}{\sqrt{n}}\right)\left(\frac{1}{n}\sum_{j=1}^{n}x_{j}^{2}\sum_{i\in\mathcal{B}_{j,n}^{*}}K^{*}\left(\frac{d_{ij}}{d_{n}^{*}}\right)x_{i}u_{i}+\frac{1}{n}\sum_{j=1}^{n}x_{j}^{2}\sum_{i\notin\mathcal{B}_{j,n}^{*}}K^{*}\left(\frac{d_{ij}}{d_{n}^{*}}\right)x_{i}u_{i}\right)\\ &\equiv c_{21}+c_{22}.\end{aligned}$$

We have

$$\begin{aligned}c_{21} &\leq O_{P}\left(\sqrt{\frac{\ell_{n}^{*}}{n}}\right)\frac{1}{n}\sum_{j=1}^{n}x_{j}^{2}\left|\frac{1}{\sqrt{\ell_{n}^{*}}}\sum_{i\in\mathcal{B}_{j,n}^{*}}K^{*}\left(\frac{d_{ij}}{d_{n}^{*}}\right)x_{i}u_{i}\right|\\ &\leq O_{P}\left(\sqrt{\frac{\ell_{n}^{*}}{n}}\right)\left(\frac{1}{n}\sum_{j=1}^{n}x_{j}^{4}\right)^{1/2}\left(\frac{1}{n}\sum_{j=1}^{n}\left|\frac{1}{\sqrt{\ell_{n}}}\sum_{i\in\mathcal{B}_{j,n}^{*}}K^{*}\left(\frac{d_{ij}}{d_{n}^{*}}\right)x_{i}u_{i}\right|^{2}\right)^{1/2}=O_{P}\left(\sqrt{\frac{\ell_{n}^{*}}{n}}\right),\end{aligned}$$

because

$$\begin{aligned}&P\left(\frac{1}{n}\sum_{j=1}^{n}\left|\frac{1}{\sqrt{\ell_{n}^{*}}}\sum_{i\in\mathcal{B}_{j,n}^{*}}K^{*}\left(\frac{d_{ij}}{d_{n}^{*}}\right)x_{i}u_{i}\right|^{2}>\Delta\right)\leq\frac{1}{\Delta}\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\ell_{n}^{*}}\sum_{i_{1}\in\mathcal{B}_{j,n}^{*}}\sum_{i_{2}\in\mathcal{B}_{j,n}^{*}}K^{*}\left(\frac{d_{i_{1}j}}{d_{n}^{*}}\right)K^{*}\left(\frac{d_{i_{2}j}}{d_{n}^{*}}\right)E\left[V_{i_{1}}V_{i_{2}}\right]\\ \leq\ &\frac{1}{\Delta}\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\ell_{n}^{*}}\sum_{i_{1}\in\mathcal{B}_{j,n}^{*}}\sum_{i_{2}\in\mathcal{B}_{j,n}^{*}}\left|E\left[V_{i_{1}}V_{i_{2}}\right]\right|\to 0\end{aligned}$$

as Δ grows. For c_{22},

$$c_{22}\leq O_{P}(1)\left(\frac{1}{n}\sum_{j=1}^{n}x_{j}^{4}\right)^{1/2}\left(\frac{1}{n}\sum_{j=1}^{n}\left|\frac{1}{\sqrt{n}}\sum_{i\notin\mathcal{B}_{j,n}^{*}}K^{*}\left(\frac{d_{ij}}{d_{n}^{*}}\right)x_{i}u_{i}\right|^{2}\right)^{1/2}=o_{P}\left(\frac{\ell_{n}^{*}}{n}\right)$$

because

$$\begin{aligned}&P\left(\frac{1}{n}\sum_{j=1}^{n}\left|\frac{1}{\sqrt{n}}\sum_{i\notin\mathcal{B}_{j,n}^{*}}K^{*}\left(\frac{d_{ij}}{d_{n}^{*}}\right)x_{i}u_{i}\right|^{2}>\Delta\right)\leq\frac{1}{\Delta}\frac{1}{n}\sum_{j=1}^{n}\frac{1}{n}\sum_{i_{1}\notin\mathcal{B}_{j,n}^{*}}\sum_{i_{2}\notin\mathcal{B}_{j,n}^{*}}\left|K^{*}\left(\frac{d_{i_{1}j}}{d_{n}^{*}}\right)\right|\left|K^{*}\left(\frac{d_{i_{2}j}}{d_{n}^{*}}\right)\right|\left|E\left(V_{i_{1}}V_{i_{2}}\right)\right|\\ \leq\ &\frac{1}{\Delta}\frac{\ell_{n}^{*}}{n}\left(\frac{1}{\ell_{n}^{*}}\sup_{c}\sum_{i_{1}\notin\mathcal{B}_{c,n}^{*}}\left|K^{*}\left(\frac{d_{i_{1}c}}{d_{n}^{*}}\right)\right|\right)\underbrace{\left(\sup_{c}\sum_{i_{2}=1}^{n}\left|E\left(V_{c}V_{i_{2}}\right)\right|\right)}_{\leq M}\to 0\end{aligned}$$

as $n\to\infty$. Therefore, $J_{n}^{*}-J_{0n}^{*}=O_{P}\left(\frac{\ell_{n}^{*}}{n}\right)+O_{P}\left(\sqrt{\frac{\ell_{n}^{*}}{n}}\right)=o_{P}(1)$ under the rate condition on ℓ_{n}^{*}, which concludes the proof.

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

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|E\left(V_iV_j'\right)\phi_{ik}\phi_{jk}\right\| = \sum_{a=1}^{n}\sum_{b=1}^{n}\left\|E\left(V_aV_b'\right)\phi_k^{(a)}\phi_k^{(b)}\right\| = \underbrace{\sum_{a=1}^{n}\left\|E\left(V_aV_a'\right)\left(\phi_k^{(a)}\right)^2\right\|}_{(1):\text{ diagonal part}} + 2\underbrace{\sum_{a=1}^{n-1}\sum_{b=a+1}^{n}\left\|E\left(V_aV_b'\right)\phi_k^{(a)}\phi_k^{(b)}\right\|}_{(2):\text{ upper triangular part}}.$$

It follows that

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

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

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

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

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

assuming that the term in parenthesis is bounded. This last condition is slightly stronger than the usual weak dependence $\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|E\left(V_iV_j'\right)\right\| < \infty$, but it holds under Assumption $\mathcal{B}_1$.

A.2 Proof of main results in the paper

Proof of Theorem 3.1. Let $C_n = Q^{-1}J_nQ^{-1}$ 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

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

since under Assumption $\mathcal{A}$, $\hat{Q}_n \to^P Q$. Thus, it suffices to show that

$$J_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_iu_i^* \to^{d^*} N\left(0, I_p\right), \text{ in prob-}P, \tag{18}$$

to conclude that

$$C_n^{-1/2}\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right) \to^{d^*} N\left(0, I_p\right), \text{ in prob-}P. \tag{19}$$

Given that $C_n^{-1/2}\sqrt{n}\left(\hat{\beta}-\beta\right)\to^d N\left(0,I_p\right)$ under Assumption $\mathcal{A}$, (19) implies the result by Polya's Theorem and the continuity of the normal distribution. Using the definition of u_i^*, (18) follows if

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i\left(\hat{u}_i-u_i\right)\eta_i\to^{p^*}0,\ \text{in prob-}P,\ \text{and} \tag{20}$$

$$J_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_iu_i\eta_i\to^{d^*}N\left(0,I_p\right),\ \text{in prob-}P, \tag{21}$$

as $n\to\infty$. For (20), we note that

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

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

$$\begin{aligned}E^*\left|a_1\right|^2 &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}tr\left(x_ix_i'x_jx_j'\right)E^*\left(\eta_i\eta_j\right)=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|tr\left(x_ix_i'x_jx_j'\right)\right|K^*\left(\frac{d_{ij}}{d_n^*}\right) \\ &\leq \frac{\ell_n^*}{n}\frac{1}{\ell_n^*n}\sum_{i=1}^{n}\sum_{j\in\mathcal{B}_{i,n}^*}\left|tr\left(x_ix_i'x_jx_j'\right)\right|+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\notin\mathcal{B}_{i,n}^*}\left|tr\left(x_ix_i'x_jx_j'\right)\right|K^*\left(\frac{d_{ij}}{d_n^*}\right)=O_p\left(\frac{\ell_n^*}{n}\right)\to 0,\end{aligned} \tag{22}$$

as $\frac{\ell_n^*}{n}\to 0$, given our assumptions on ℓ_n^*. For the first term in (22), note in particular that by Markov's inequality, for some $M>0$,

$$P\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in\mathcal{B}_{i,n}^*}\left|tr\left(x_ix_i'x_jx_j'\right)\right|>M\ell_n^*\right)\leq\frac{1}{M}\frac{1}{\ell_n^*}\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in\mathcal{B}_{i,n}^*}\underbrace{E\left|tr\left(x_ix_i'x_jx_j'\right)\right|}_{\leq\Delta\text{ by Assumption }\mathcal{B}_4}\leq\frac{1}{M}\frac{1}{\ell_n^*}\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in\mathcal{B}_{i,n}^*}\Delta=\Delta\frac{1}{M}\frac{1}{\ell_n^*}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\ell_{i,n}^*\right)}_{=\ell_n^*\text{ by definition}},$$

which can be made arbitrarily small for M sufficiently large. For the second term,

$$P\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j\notin\mathcal{B}_{i,n}^*}\left|tr\left(x_ix_i'x_jx_j'\right)\right|K^*\left(\frac{d_{ij}}{d_n^*}\right)>M\ell_n^*\right)\leq\frac{1}{M}\frac{1}{\ell_n^*}\frac{\Delta}{n}\sum_{i=1}^{n}\sum_{j\notin\mathcal{B}_{i,n}^*}\left|K^*\left(\frac{d_{ij}}{d_n^*}\right)\right|\leq\frac{\Delta}{M}\left(\frac{1}{\ell_n^*}\sup_c\sum_{j\notin\mathcal{B}_{c,n}^*}\left|K^*\left(\frac{d_{cj}}{d_n^*}\right)\right|\right),$$

which goes to zero as Δ grows for any $M>0$ under Assumption $\mathcal{K}$(ii).

Next, we prove (21). Given Assumption $\mathcal{K}$, $\mathbb{K}_n^*$ 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 ~i.i.d.$(0,1)$, which implies that Z_k^* is an independent heterogeneous array. We will show that $n^{-1/2}\sum_{k=1}^{n} Z_k^* \to^{d^*} N\left(0, I_p\right)$, 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

$$J_{0n}^* = Var^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i\eta_i\right) = Var^*\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left(V'\lambda_k^{1/2}\phi_k\right)v_k\right) = V'\left(\frac{1}{n}\sum_{k=1}^{n}\lambda_k\phi_k\phi_k'\right)V.$$

By Lemma A.1, $J_{0n}^* - J_n \to^P 0$, 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 $\upsilon > 0$,

$$\frac{1}{n^{1+\upsilon/2}}\sum_{k=1}^{n} E^*\left\|Z_k^*\right\|^{2+\upsilon} \to^P 0. \tag{23}$$

Note that

$$\begin{aligned}
&\frac{1}{n^{1+\upsilon/2}}\sum_{k=1}^{n} E^*\left\|Z_k^*\right\|^{2+\upsilon} = \frac{1}{n^{1+\upsilon/2}}\sum_{k=1}^{n} E^*\left\|J_n^{-1/2}\left(\lambda_k^{1/2}V'\phi_k\right)v_k\right\|^{2+\upsilon} \\
&\leq \left\|J_n^{-1/2}\right\|^{2+\upsilon}\frac{1}{n^{1+\upsilon/2}}\left(\sup_a \lambda_a^{1+\upsilon/2}\right)\sum_{k=1}^{n} E^*\left\|\sum_{i=1}^{n} V_i\phi_{ik}v_k\right\|^{2+\upsilon} \\
&= \left\|J_n^{-1/2}\right\|^{2+\upsilon}\frac{1}{n^{1+\upsilon/2}}\left(\sup_a \lambda_a^{1+\upsilon/2}\right)\sum_{k=1}^{n} E^*\left(\left\|\left[\begin{array}{c}\sum_{i=1}^{n} V_i^{(1)}\phi_{ik}v_k \\ \vdots \\ \sum_{i=1}^{n} V_i^{(p)}\phi_{ik}v_k\end{array}\right]\right\|^2\right)^{1+\upsilon/2} \\
&= \left\|J_n^{-1/2}\right\|^{2+\upsilon}\frac{1}{n^{1+\upsilon/2}}\left(\sup_a \lambda_a^{1+\upsilon/2}\right)\sum_{k=1}^{n} E^*\left(\sum_{m=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n} V_i^{(m)}V_j^{(m)}\phi_{ik}\phi_{jk}v_k^2\right)^{1+\upsilon/2} \\
&= \left\|J_n^{-1/2}\right\|^{2+\upsilon}\frac{1}{n^{\upsilon/2}}\left(\max_a \lambda_a^{1+\upsilon/2}\right)\frac{1}{n}\sum_{k=1}^{n}\left(\sum_{m=1}^{p}\sum_{i=1}^{n}\sum_{j=1}^{n} V_i^{(m)}V_j^{(m)}\phi_{ik}\phi_{jk}\right)^{1+\upsilon/2} E^*\left|v_k\right|^{2+\upsilon}.
\end{aligned} \tag{24}$$

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

$$\frac{1}{n}\left(\max_a \lambda_a^2\right) = o_P(1), \tag{25}$$

$$\frac{1}{n}\sum_{k=1}^n\left(\sum_{m=1}^p\sum_{i=1}^n\sum_{j=1}^n V_i^{(m)}V_j^{(m)}\phi_{ik}\phi_{jk}\right)^2 = O_P(1), \tag{26}$$

and noting that $E^*|v_k|^4 < M$ by assumption. To prove (25), 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 $a = 1,\ldots,n$,

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

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

$$\lambda_a\left|\phi_{i_a a}\right| \le \sum_{j=1}^n\left|K^*\left(\frac{d_{i_a j}}{d_n^*}\right)\right|\left|\phi_{ja}\right| \Longleftrightarrow \lambda_a \le \sum_{j=1}^n\left|K^*\left(\frac{d_{i_a j}}{d_n^*}\right)\right|\frac{\left|\phi_{ja}\right|}{\left|\phi_{i_a a}\right|} \le \sum_{j=1}^n\left|K^*\left(\frac{d_{i_a j}}{d_n^*}\right)\right|.$$

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

$$\sup_a \lambda_a \le \sup_a\sum_{j=1}^n\left|K^*\left(\frac{d_{aj}}{d_n^*}\right)\right|.$$

Assumptions $\mathcal{K}$ and $\mathcal{B}_3$ imply

$$\sup_a\sum_{j=1}^n\left|K^*\left(\frac{d_{aj}}{d_n^*}\right)\right| \le \underbrace{\sup_a\sum_{j\in B_{a,n}}1}_{=\sup_a \ell_{a,n}^* < c\ell_n^*} + \sup_a\sum_{j\notin B_{a,n}}\left|K^*\left(\frac{d_{i_a j}}{d_n^*}\right)\right|,$$

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

$$\begin{aligned}\frac{1}{n}\sup_a\lambda_a^2 &\le \frac{1}{n}\sup_a\left(\sum_{j=1}^n\left|K^*\left(\frac{d_{aj}}{d_n^*}\right)\right|\right)^2 \le \frac{1}{n}\left(c\ell_n^* + \sup_a\sum_{j\notin B_{a,n}}\left|K^*\left(\frac{d_{i_a j}}{d_n^*}\right)\right|\right)^2 \\ &\le 2\frac{\left(c\ell_n^*\right)^2}{n} + 2\frac{\left(\ell_n^*\right)^2}{n}\left(\frac{1}{\ell_n^*}\sup_a\sum_{j\notin B_{a,n}}\left|K^*\left(\frac{d_{i_a j}}{d_n^*}\right)\right|\right)^2 \le O\left(\left(\frac{\ell_n^*}{n^{1/2}}\right)^2\right) = o(1)\end{aligned}$$

under the rate condition in Theorem 3.1.

Next we prove (26). We will focus on the special case where $p = 1$ for simplicity, and show that

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

which suffices to prove (26) given Markov's inequality. Letting $\tilde{V}_{ik} = V_i\phi_{ik}$, we have that

$$\frac{1}{n}\sum_{k=1}^n E\left|\sum_{i=1}^n V_i\phi_{ik}\right|^4 = \frac{1}{n}\sum_{k=1}^n E\left|\sum_{i=1}^n\tilde{V}_{ik}\right|^4 = \frac{1}{n}\sum_{k=1}^n\sum_{i_1=1}^n\sum_{i_2=1}^n\sum_{i_3=1}^n\sum_{i_4=1}^n E\left(\tilde{V}_{i_1k}\tilde{V}_{i_2k}\tilde{V}_{i_3k}\tilde{V}_{i_4k}\right) \equiv C_1 + C_2 + C_3 + C_4,$$

where, adding and subtracting appropriately,

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

We will now show that each of the terms C_1 through C_4 is $O(1)$ 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 $\mathcal{B}_1$). We will rely on this assumption as well as on the orthonormality of the eigenvectors $\phi_k = (\phi_{ik} : i = 1,\dots,n)$ 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

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

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

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

This implies that

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

Hence,

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

Similarly,

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

and

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

Putting everything together yields

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

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

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

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 $\sup_{1 \leq k \leq n} \left|\phi_{ik}^2\right| \leq 1$. Thus

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

implying that

$$\begin{aligned} C_1 &\leq \Delta \frac{1}{n} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \sum_{i_4=1}^{n} \sum_{l=1}^{\infty} \left| r_{i_1 l} r_{i_2 l} r_{i_3 l} r_{i_4 l} \right| = \Delta \frac{1}{n} \sum_{i_1=1}^{n} \sum_{l=1}^{\infty} \left| r_{i_1 l} \right| \left(\sum_{i_2=1}^{n} \left| r_{i_2 l} \right| \right) \left(\sum_{i_3=1}^{n} \left| r_{i_3 l} \right| \right) \left(\sum_{i_4=1}^{n} \left| r_{i_4 l} \right| \right) \\ &\leq \Delta \frac{1}{n} \sum_{i_1=1}^{n} \sum_{l=1}^{\infty} \left| r_{i_1 l} \right| \underbrace{\left(\sum_{i_2=1}^{\infty} \left| r_{i_2 l} \right| \right)}_{\leq M \text{ by Assumption } \mathcal{B}_1} \left(\sum_{i_3=1}^{\infty} \left| r_{i_3 l} \right| \right) \left(\sum_{i_4=1}^{\infty} \left| r_{i_4 l} \right| \right) \leq \Delta M^3 \frac{1}{n} \sum_{i_1=1}^{n} \underbrace{\sum_{l=1}^{\infty} \left| r_{i_1 l} \right|}_{\leq M} \leq Const. \end{aligned}$$

which proves that $C_1 = O(1)$. To complete the proof, note that each of C_2, C_3 and C_4 can be shown to be $O(1)$ by applying Lemma A.2.

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

Lemma A.3 *Suppose Assumptions $\mathcal{A}$, $\mathcal{K}$, $\mathcal{W}$ and $\mathcal{B}_1$-$\mathcal{B}_4$ hold. If $E^* |v_i|^4 < M$ and $d_n, \ell_n \to \infty$ and $d_n^*, \ell_n^* \to \infty$ as $n \to \infty$ such that $\ell_n / n = o(1)$ and $\ell_n^* / n^{1/2} = o(1)$, then (i) $\hat{J}_n^* - J_n^* \to^{P^*} 0$, in prob-P when unrestricted residuals are used, and (ii) $\hat{J}_n^* - J_n^* \to^{P^*} 0$, in prob-P, when restricted residuals are used, and H_0 is true.*

Proof of Lemma A.3. 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

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

where

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

and

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

First, note that $A_4 = o_P(1)$ by Lemma A.1. 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

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

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

$$E|A_{31}| \leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\underbrace{E\left(x_i^2x_j^2\right)}_{\leq M} \leq M\frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in\mathcal{B}_{i,n}}1\right)+M\frac{\ell_n}{n}\left(\frac{1}{\ell_n}\sup_c\sum_{j\notin\mathcal{B}_{c,n}}\left|K\left(\frac{d_{cj}}{d_n}\right)\right|\right) = O\left(\frac{\ell_n}{n}\right),$$

given in particular Assumption $\mathcal{K}$(ii). Thus, by Markov's inequality, $A_{31} = O_P\left(\frac{\ell_n}{n}\right) = o_P(1)$. For A_2,

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

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

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

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

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

where $e_1 = O_{P^*}(1)$ in probability. For this result, it suffices to show that $E\left(E^*\left(|e_1|^2\right)\right) = O(1)$. But

$$\begin{aligned}E^*\left(|e_1|^2\right) &= \frac{1}{n}\sum_{j=1}^{n}\frac{1}{\ell_n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}K\left(\frac{d_{i_1j}}{d_n}\right)K\left(\frac{d_{i_2j}}{d_n}\right)x_{i_1}^2x_{i_2}^2E^*\left(\eta_{i_1}\eta_{i_2}\right) \\ &= \frac{1}{n}\sum_{j=1}^{n}\frac{1}{\ell_n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}K\left(\frac{d_{i_1j}}{d_n}\right)K\left(\frac{d_{i_2j}}{d_n}\right)x_{i_1}^2x_{i_2}^2K^*\left(\frac{d_{i_1i_2}}{d_n^*}\right) \\ &\leq \frac{1}{n}\sum_{j=1}^{n}\frac{1}{\ell_n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}K\left(\frac{d_{i_1j}}{d_n}\right)K\left(\frac{d_{i_2j}}{d_n}\right)x_{i_1}^2x_{i_2}^2,\end{aligned}$$

implying that

$$E\left(E^*\left|e_1^2\right|\right) \leq \frac{1}{n}\sum_{j=1}^{n}\frac{1}{\ell_n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}K\left(\frac{d_{i_1j}}{d_n}\right)K\left(\frac{d_{i_2j}}{d_n}\right)\underbrace{E\left(x_{i_1}^2x_{i_2}^2\right)}_{\leq M} \leq \frac{M}{n}\sum_{j=1}^{n}\left(\frac{1}{\ell_n}\sum_{i=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)\right)^2 = O(1)$$

given in particular Assumption $\mathcal{W}(iii)$. Therefore, $A_{21}^{(1)} = O_{P^*}(\ell_n/n) = o_{P^*}(1)$, in probability. A similar argument implies that $A_{21}^{(2)} = O_{P^*}\left(\sqrt{\frac{\ell_n}{n}}\right) = o_{P^*}(1)$, in probability. Thus, to end the proof, we show that $A_1 =$

$o_{P^*}(1)$ in probability. We can write

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

where $A_{12}=-\left(J_n^*-J_{0n}^*\right)=o_P(1)$, as shown in the proof of Lemma A.1. 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

$$\begin{aligned}\left|A_{11}^{(1)}\right| &\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|K\left(\frac{d_{ij}}{d_n}\right)\right|\left|x_i^2x_j^2\eta_i\eta_j\right|\left(\beta-\hat{\beta}\right)^2\\&\leq O_P\left(\frac{\ell_n}{n}\right)\frac{1}{n}\sum_{i=1}^{n}\left|x_i^2\eta_i\right|\left(\frac{1}{\ell_n}\sum_{j\in\mathcal{B}_{i,n}}\left|x_j^2\eta_j\right|\right)+O_P\left(\frac{\ell_n}{n}\right)\frac{1}{n}\sum_{i=1}^{n}\left|x_i^2\eta_i\right|\left(\frac{1}{\ell_n}\sup_c\sum_{j\notin\mathcal{B}_{c,n}}\left|K\left(\frac{d_{cj}}{d_n}\right)\right|\left|x_j^2\eta_j\right|\right)\equiv h_1+h_2.\end{aligned}$$

For h_1,

$$h_1\leq O_P\left(\frac{\ell_n}{n}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\left|x_i^2\eta_i\right|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\ell_n}\sum_{j\in\mathcal{B}_{i,n}}\left|x_j^2\eta_j\right|\right)^2\right)^{1/2},$$

where each of the square root terms can be shown to be $O_{P^*}(1)$, in probability. By a similar argument, we can show that $h_2=O_{P^*}\left(\frac{\ell_n}{n}\right)$. Thus, $A_{11}^{(1)}=O_{P^*}\left(\frac{\ell_n}{n}\right)$ in prob-P. For $A_{11}^{(2)}$,

$$\left|A_{11}^{(2)}\right|=\left|\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)x_i^2x_ju_j\left(\beta-\hat{\beta}\right)\eta_i\eta_j\right|=O_P\left(\sqrt{\frac{\ell_n}{n}}\right)(m_1+m_2),$$

where

$$m_1\equiv\left|\frac{1}{n}\sum_{i=1}^{n}x_i^2\eta_i\frac{1}{\sqrt{\ell_n}}\sum_{j\in\mathcal{B}_{i,n}}K\left(\frac{d_{ij}}{d_n}\right)x_ju_j\eta_j\right| \text{ and } m_2=\left|\frac{1}{n}\sum_{i=1}^{n}x_i^2\eta_i\frac{1}{\sqrt{\ell_n}}\sum_{j\notin\mathcal{B}_{i,n}}K\left(\frac{d_{ij}}{d_n}\right)x_ju_j\eta_j\right|.$$

Using arguments similar to those used before, we can show that $m_1=m_2=O_{P^*}(1)$, in probability. Thus, $A_{11}^{(2)}=O_{P^*}\left(\sqrt{\frac{\ell_n}{n}}\right)$, concluding the proof that $A_{11}=o_{P^*}(1)$, in prob-P. Finally, we show that $A_{13}=o_{P^*}(1)$. We have

$$\begin{aligned}A_{13} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)V_iV_j\left(\eta_i\eta_j-1\right)\\&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)V_iV_j\left(\eta_i\eta_j-K^*\left(\frac{d_{ij}}{d_n^*}\right)\right)+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{d_{ij}}{d_n}\right)V_iV_j\left(K^*\left(\frac{d_{ij}}{d_n^*}\right)-1\right)\\&= A_{13}^{(1)}+A_{13}^{(2)},\end{aligned}\tag{27}$$

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

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

$$= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}K\left(\frac{d_{i_1j_1}}{d_n}\right)K\left(\frac{d_{i_2j_2}}{d_n}\right)E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right)\left(E^*\left(\eta_{i_1}\eta_{j_1}\eta_{i_2}\eta_{j_2}\right) - K^*\left(\frac{d_{i_1j_1}}{d_n^*}\right)K^*\left(\frac{d_{i_2j_2}}{d_n^*}\right)\right).$$

Let L_{ik} denote the (i,k)-th element of L_n such that $\mathbb{K}_n^* = L_nL_n'$. 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),

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

Similarly, it follows that

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

Thus,

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

Given this decomposition, it follows that

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

where

$$B_{11} = \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n} K\left(\frac{d_{i_1j_1}}{d_n}\right)K\left(\frac{d_{i_2j_2}}{d_n}\right)E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right)\left(\sum_{k=1}^{n}L_{i_1k}L_{j_1k}L_{i_2k}L_{j_2k}\left(E^*\left(v_k^4\right)-3\right)\right)$$

$$B_{12} = \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n} K\left(\frac{d_{i_1j_1}}{d_n}\right)K\left(\frac{d_{i_2j_2}}{d_n}\right)K^*\left(\frac{d_{i_1i_2}}{d_n^*}\right)K^*\left(\frac{d_{j_1j_2}}{d_n^*}\right)E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right)$$

$$B_{13} = \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n} K\left(\frac{d_{i_1j_1}}{d_n}\right)K\left(\frac{d_{i_2j_2}}{d_n}\right)K^*\left(\frac{d_{i_1j_2}}{d_n^*}\right)K^*\left(\frac{d_{j_1i_2}}{d_n^*}\right)E\left(V_{i_1}V_{j_1}V_{i_2}V_{j_2}\right).$$

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

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

To bound $B_{11}^{(1)}$, recall that $L_{ik}=\sqrt{\lambda_k}\phi_{ik}$ where $\max_m \lambda_m^2 = O\left(\left(\ell_n^*\right)^2\right)$. Then,

$$\left|\sum_{k=1}^{n}L_{i_1k}L_{j_1k}L_{i_2k}L_{j_2k}\right| = \left|\sum_{k=1}^{n}\lambda_k^2\phi_{i_1k}\phi_{j_1k}\phi_{i_2k}\phi_{j_2k}\right| \leq \left(\max_m \lambda_m^2\right)\max_i\sum_{k=1}^{n}\phi_{ik}^4 = O\left(\left(\ell_n^*\right)^2\right),$$

since $\max_i \sum_{k=1}^{n}\phi_{ik}^4 \leq 1$. Moreover, by Assumption $\mathcal{B}_1$,

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

by using an argument similar to the one used to study the term C_1 in the proof of Theorem 3.1. This implies that

$$\begin{aligned}
B_{11}^{(1)} &\leq O\left(\left(\ell_n^*\right)^2\right)\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}\left|\sum_{l=1}^{\infty}r_{i_1l}r_{j_1l}r_{i_2l}r_{j_2l}\left(E\left(e_l^4\right)-3\right)\right| \\
&\leq O\left(\left(\ell_n^*\right)^2\right)\underbrace{\frac{1}{n^2}\sum_{i_1=1}^{n}\left(\sum_{l=1}^{\infty}\left|r_{i_1l}\right|\sum_{j_1=1}^{n}\left|r_{j_1l}\right|\sum_{i_2=1}^{n}\left|r_{i_2l}\right|\sum_{j_2=1}^{n}\left|r_{j_2l}\right|\right)}_{=O(1/n)} = O\left(\frac{\left(\ell_n^*\right)^2}{n}\right).
\end{aligned}$$

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

$$\begin{aligned}
B_{11}^{(2)} &= M\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}\left|E\left(V_{i_1}V_{j_1}\right)E\left(V_{i_2}V_{j_2}\right)\right|\left|\sum_{k=1}^{n}L_{i_1k}L_{j_1k}L_{i_2k}L_{j_2k}\right| \\
&= M\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}\left|E\left(V_{i_1}V_{j_1}\right)E\left(V_{i_2}V_{j_2}\right)\right|\left|\sum_{k=1}^{n}\lambda_k^2\phi_{i_1k}\phi_{j_1k}\phi_{i_2k}\phi_{j_2k}\right| \\
&\leq M\left(\max_m \lambda_m^2\right)\frac{1}{n^2}\sum_{k=1}^{n}\underbrace{\left(\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\left|E\left(V_{i_1}V_{j_1}\right)\phi_{i_1k}\phi_{j_1k}\right|\right)}_{\leq M \text{ by Lemma A.2}}\left(\sum_{i_2=1}^{n}\sum_{j_2=1}^{n}\left|E\left(V_{i_2}V_{j_2}\right)\phi_{i_2k}\phi_{j_2k}\right|\right) \\
&\leq M\left(\max_m \lambda_m^2\right)\frac{1}{n^2}\sum_{k=1}^{n}M^2 = O\left(\frac{\left(\ell_n^*\right)^2}{n}\right).
\end{aligned}$$

Using the same procedure, we can show that $B_{11}^{(3)} = B_{11}^{(4)} = O\left(\left(\ell_n^*\right)^2/n\right)$. Hence, $B_{11} = O\left(\left(\ell_n^*\right)^2/n\right) = o(1)$ given that $\left(\ell_n^*\right)^2/n = o(1)$. Since we can also show that the terms B_{12} and B_{13} are $O(\ell_n/n) + O\left(\ell_n^*/n\right)$ 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 (27), note that

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

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

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

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

Proof of Theorem 4.1. It follows from Theorem 3.1 and Lemma A.3.

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

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

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

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

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

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

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

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

[17] 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.

[18] 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.

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

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

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

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

[23] 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.

[24] Kojevnikov, D. (2019), The Bootstrap for Network Dependent Processes, Vancouver School of Economics, University of British Columbia, Working paper.

[25] 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.

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

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

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

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

[30] 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.

[31] 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.

[32] 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.

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

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

[35] Robinson, P. (2011), Asymptotic Theory for Nonparametric Regression with Spatial Data, *Journal of Econometrics* 165, 5–19.

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

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

[38] 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.

[39] 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

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

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