Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 140 (2007) 131–154
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

HAC estimation in a spatial framework

Harry H. Kelejian*, Ingmar R. Prucha

Department of Economics, University of Maryland, College Park, MD 20742, USA

Received in revised form 1 February 2006; accepted 5 September 2006
Available online 19 December 2006

Abstract

We suggest a non-parametric heteroscedasticity and autocorrelation consistent (HAC) estimator of the variance–covariance (VC) matrix for a vector of sample moments within a spatial context. We demonstrate consistency under a set of assumptions that should be satisfied by a wide class of spatial models. We allow for more than one measure of distance, each of which may be measured with error. Monte Carlo results suggest that our estimator is reasonable in finite samples. We then consider a spatial model containing various complexities and demonstrate that our HAC estimator can be applied in the context of that model.

JEL classification: C12; C14; C21

Keywords: Heteroscedasticity and autocorrelation consistent (HAC) estimator; Instrumental variable estimator; Spatial models

1. Introduction[1]

Spatial models are an important tool in economics, regional science and geography in analyzing a wide range of empirical issues.[2] Typically, these models focus on spatial interactions, which could be due to competition between cross sectional units, copy-cat policies, net work issues, spill-overs, externalities, regional issues, etc. Applications in the

*Corresponding author. Tel.: +1 301 405 3492; fax: +1 301 405 3542.
E-mail address: kelejian@econ.umd.edu (H.H. Kelejian).

[1]We thank Badi Baltagi and three anonymous referees for valuable suggestions and helpful comments. We are also indebted to John Benedetto, Immanuel Bomze, Roger Horn, and Benedikt Pötscher for helpful discussions. We also thank Irani Arraiz for excellent research assistance. Both authors gratefully acknowledge support from the National Science Foundation through Grant SES-0001780.

[2]Classic references on spatial models are Cliff and Ord (1973, 1981), Anselin (1988), and Cressie (1993).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2006.09.005

recent literature include, for example, the determinants of various forms of productivity, various categories of local public expenditures, vote seeking and tax setting behavior, population and employment growth, contagion problems, and the determinants of welfare expenditures.[3] To facilitate the empirical analysis of spatial issues the formal development of estimation methods for spatial models has received increasing attention in recent years.[4]

The purpose of this paper is two-fold: First we suggest, within a spatial context, a non-parametric heteroscedasticity and autocorrelation consistent (HAC) estimator of a variance–covariance (VC) matrix for a vector of sample moments of the form $n^{-1/2}H'u$, where H is a non-stochastic matrix, u is a vector of disturbances, and n is the sample size—i.e., a spatial HAC, (SHAC). The need to estimate the VC matrix of such a vector of sample moments arises frequently within the context of instrumental variable (IV) estimation. We demonstrate the consistency of our SHAC estimator under a set of relatively simple assumptions that cover, in particular, the important and widely used class of Cliff–Ord type models.

HAC estimators have been the focus of extensive research in the time series literature. A classic reference in that literature is Grenander and Rosenblatt (1957). Contributions to this research in the econometrics literature include, among others, Newey and West (1987), Gallant and White (1988), Andrews (1991), Andrews and Monahan (1992), Pötscher and Prucha (1997) and de Jong and Davidson (2000).

In the statistics literature Priestley (1964) made early contributions towards an extension of HAC estimation for spatial processes within the context of estimating spectral densities of stationary random fields (with the index an element of $\mathbf{Z}^2$). The theoretical econometrics literature relating to HAC estimators for spatially dependent data is relatively sparse. To the best of our knowledge, the first contributions to the theoretical econometrics literature are Conley (1996, 1999). However, the approach we take in this paper differs from that of Conley in important ways. Conley assumes that the underlying data generating process is represented by continuous-index random field (with the index an element of a metric space), and explicitly models sampling from this process. He assumes that the data generating process is spatially stationary and spatially alpha mixing. Our setup is different and aims, among other things, to accommodate spatial processes that are generated by Cliff–Ord type models. Those models do not explicitly index observations in terms of elements of a metric space (although they can accommodate such interpretations) and generate the observations as the solution of a simultaneous equation system. Spatial dependences are modeled in terms of a so-called spatial weights matrix. Even if the underlying innovations are i.i.d., this will in general result in a spatial process that is non-stationary simply if the respective units have different numbers of neighbors, as is frequently the case in applications.[5] Our dependence assumptions are stated in terms of

[3]Some applications along these lines are, e.g., Audretsch and Feldmann (1996), Bell and Bockstael (2000), Bernat (1996), Besley and Case (1995), Bollinger and Ihlanfeldt (1997), Buettner (1999), Case (1991), Case et al. (1993), Dowd and LeSage (1997), Holtz-Eakin (1994), Kelejian and Robinson (2000, 1997, 1993), Pinkse et al. (2002), Pulvino (1998), Rey and Boarnet (2004), Shroder (1995), and Vigil (1998).

[4]Recent theoretical contributions include Baltagi and Li (2004, 2001a,b), Baltagi et al. (2003), Conley (1999), Kelejian and Prucha (2004, 2002, 2001, 1999, 1998, 1997), Kelejian et al. (2004), Lee (2004, 2003, 2002, 2001a,b), LeSage (2000, 1997), Pace and Barry (1997), Pinkse and Slade (1998), Pinkse et al. (2002), and Rey and Boarnet (2004).

[5]This is consistent with the view of, e.g., Fuentes (2002a,b) who states that spatial processes are often ''non-stationary, in the sense that the spatial structure depends on location''. Of course, there are also many situations where stationarity is appropriate and our setup allows for a wide set of stationary processes.

simple conditions on a decomposition of the VC matrix, which accommodates non-stationary and (unconditionally) heteroscedastic processes. Another distinguishing feature is that our setup and proofs allows for a triangular array structure of the data. The reason for this is that technically this structure arises due the presence of spatial lags, which are often the focus of attention in Cliff–Ord type models.

Pinkse et al. (2002) consider a specific Cliff–Ord type spatial model and demonstrate, within the context of that model, the consistency of a SHAC estimator. Their consistency result is given under a set of high level assumptions, which seem substantially more complex than those maintained in this paper. Driscoll and Kraay (1998) also provide results regarding the consistent non-parametric estimation of a large sample VC matrix for spatially dependent data. However, in contrast to the above cited papers and our specifications, their approach relates to a panel data model in which the number of time periods T limits to infinity.

Our specifications also accommodate situations in which the researcher considers more than one distance measure and is unsure about which one to use in the specification of the SHAC estimator. To that effect we allow for the researcher to employ several distance measures, and we show that our estimator remains consistent, as long as the ''true'' distance measure is included in the set of measures employed by the researcher.[6] We also allow for measurement errors relating to the distance measures. Our consistency result is also generic in the sense that the estimated residuals used in the formulation of the SHAC estimator may correspond to a variety of linear and non-linear models, provided they are $n^{1/2}$-consistently estimated.

All of our asymptotic results are derived under the assumption of a single cross section in which the number of cross sectional units n tends to infinity. Generalizations to a finite number of cross section, T, are trivial. We also give Monte Carlo results which suggest that our SHAC estimator yields reasonable results in finite samples.

The second part of the paper considers a general spatial regression model that allows for endogenous regressors, their spatial lags, as well as exogenous regressors. The model may, in particular, represent the ith equation of a simultaneous system of equations.[7] The disturbance process allows for general patterns of correlation and heteroscedasticity. We define an IV estimator for this model and derive its asymptotic distribution. We then apply our results concerning SHAC estimation derived in the first part of this paper to obtain a consistent estimator for the asymptotic VC matrix of the IV estimator.

The (non-parametric) model of spatial correlation and heteroscedasticity is specified in Section 2. That section also contains a discussion of the model assumptions. In Section 3 we specify our SHAC estimator. We first consider the case of single and then the case of

[6]The economics of a particular application will typically suggest a set of possible distance measures. Our setup only maintains that one of these measures is the ''true'' one, and does not require for the researcher to combine different distance measures into one distance measure. The specification of our estimator is thus quite different from specifications where distance is taken to be a p-dimensional Euclidean distance in $\mathbf{R}^p$, which may, e.g., be viewed as composed of a set of lower dimensional Euclidean distances. Of course, the case of one p-dimensional distance measure is included as a special case.

[7]Among other things, this model differs from the ith equation's specifications considered in Kelejian and Prucha (2004) in that their disturbance VC matrix is parametrically specified, while ours is not, and our specifications here allow for some of the endogenous regressors to be generated by a non-linear model, while theirs do not.

multiple distance measures. In this section we give our central results concerning the consistency of SHAC estimators. In Section 4 we specify a general spatial model which allows the disturbance process to be of the general form considered in Section 2. Section 4 also contains large sample results concerning the estimator of the regression parameters of that model, as well as of our SHAC estimator of the VC matrix involved. In Section 5 we report on a limited Monte Carlo study.[8] Section 6 contains conclusions and suggestions for further work. Technical details are relegated to the appendix.

It will be helpful to introduce the following notation: Let A_n with $n \in \mathbf{N}$ be some matrix; we then denote the (i,j)th element of A_n as $a_{ij,n}$. Similarly, if v_n with $n \in \mathbf{N}$ is a vector, then $v_{i,n}$ denotes the ith element of v_n. An analogous convention is adopted for matrices and vectors that do not depend on the index n, in which case the index n is suppressed on the elements. If A_n is a square matrix, then A_n^{-1} denotes the inverse of A_n. If A_n is singular, then A_n^{-1} should be interpreted as the generalized inverse of A_n. Further, let $(B_n)_{n\in\mathbf{N}}$ be some sequence of $n \times n$ matrices. Then we say the row and column sums of the (sequence of) matrices B_n are bounded uniformly in absolute value if there exists a constant $c_B < \infty$ (that does not dependent of n) such that

$$\max_{1\leqslant i\leqslant n} \sum_{j=1}^{n} |b_{ij,n}| \leqslant c_B \quad \text{and} \quad \max_{1\leqslant j\leqslant n} \sum_{i=1}^{n} |b_{ij,n}| \leqslant c_B \ \text{ for all } n \in \mathbf{N}$$

holds. For future reference we note that if $(B_n)_{n\in\mathbf{N}}$ and $(A_n)_{n\in\mathbf{N}}$ are sequences of $n \times n$ matrices whose row and column sums are bounded uniformly in absolute value, then so are the row and column sums of $A_n + B_n$ and A_nB_n. Also, if C_n is a sequence of $q \times n$ matrices whose elements are uniformly bounded in absolute value, then so are the elements of C_nA_n, see, e.g., Kelejian and Prucha (1998).

2. A model for spatial correlation

In the following we specify a general cross sectional disturbance process, allowing for unknown forms of heteroscedasticity and correlations across spatial units. Our SHAC estimator will be based on estimated disturbances. We will maintain a generic and simple set of assumptions concerning the estimated disturbances, which should make our estimator applicable in many situations involving linear and non-linear models, provided that the model parameters are estimated $n^{1/2}$-consistently. The results do not depend on the interpretation of the spatial process as a disturbance process. The case where the process is observed is covered as a trivial special case.

2.1. Assumptions

We assume that the $n \times 1$ disturbance vector u_n is generated as follows:

$$u_n = R_n\varepsilon_n, \tag{1}$$

where ε_n is a $n \times 1$ vector of innovations and R_n is an $n \times n$ non-stochastic matrix whose elements are not known. Now let H_n be a $n \times p_h$ non-stochastic matrix of instruments. The asymptotic distribution of corresponding IV estimators will then typically

[8] In a different context, Monte Carlo results relating to non-parametric estimation of asymptotic variances are also given in Conley and Molinari (2005).

involve the VC matrix

$$\Psi_n = (\psi_{ij,n}) = VC(n^{-1/2}H_n'u_n) = n^{-1}H_n'\Sigma_n H_n, \tag{2}$$

where $\Sigma_n = (\sigma_{ij,n})$ denotes the VC matrix of u_n. The focus of the first part of this paper is to find a consistent estimator for Ψ_n and to prove the consistency of that estimator under a set of assumptions that is suitable in a spatial context.[9]

Let $u_n' = (u_{1,n}, \ldots, u_{n,n})$, $\hat{u}_n' = (\hat{u}_{1,n}, \ldots, \hat{u}_{n,n})$, and $\varepsilon_n' = (\varepsilon_{1,n}, \ldots, \varepsilon_{n,n})$ where $\hat{u}_n$ is an estimator for u_n. Also, at this point, assume there is a meaningful distance measure, with the usual properties, between units i and j, say $d_{ij,n} = d_{ji,n} \geqslant 0$. We allow for the case where the researcher measures those distances with error as say $d^*_{ij,n} = d^*_{ji,n} \geqslant 0$.[10]

We can now state the set of maintained assumptions. A detailed discussion of those assumptions will be given in the next subsection. The assumptions maintain that various quantities are uniformly bounded by some finite constants, which do not depend on i or n.

Assumption 1. For each $n > 1$, $\varepsilon_{i,n}$ is i.i.d. $(0, 1)$ with $\mathrm{E}|\varepsilon_{i,n}|^q \leqslant c_E$ for some $q \geqslant 4$, with $0 < c_E < \infty$.

Assumption 2. The (non-stochastic) matrix R_n is non-singular and the row and column sums of R_n and R_n^{-1} are bounded uniformly in absolute value by some constant c_R, $0 < c_R < \infty$.

Assumption 3. The (non-stochastic) instrument matrix H_n has full column rank p_h for n large enough, and its elements are uniformly bounded in absolute value by some constant c_H, $0 < c_H < \infty$.

We assume that the researcher can select a distance $d_n > 0$ such that $d_n \uparrow \infty$ as $n \to \infty$. For each unit $i = 1, \ldots, n$, let $\ell_{i,n}$ denote the number of units (neighbors) j for which $d^*_{ij,n} \leqslant d_n$, i.e.,

$$\ell_{i,n} = \sum_{j=1}^{n} \mathbf{1}_{d^*_{ij,n}} \leq d_n,$$

and let $\ell_n = \max_{1 \leqslant i \leqslant n}(\ell_{i,n})$.

Assumption 4. (a) $E\ell_n^2 = \mathrm{o}(n^{2\tau})$ where $\tau \leqslant \frac{1}{2}(q-2)/(q-1)$ and q is defined in Assumption 1; (b) $\sum_{j=1}^{n} |\sigma_{ij,n}| d^{\rho_S}_{ij,n} \leqslant c_S$ for some $\rho_S \geqslant 1$ and $0 < c_S < \infty$, where $\sigma_{ij,n}$ is the (i,j)th element of Σ_n.

Assumption 5. The distance measure employed by the researcher is given by

$$d^*_{ij,n} = d_{ij,n} + v_{ij,n} \geqslant 0,$$

where $v_{ij,n} = v_{ji,n}$ denotes the measurement errors, $|v_{ij,n}| \leqslant c_V$ with $0 < c_V < \infty$, and $(v_{ij,n})$ is independent of $(\varepsilon_{i,n})$.

Our next assumption relates to the estimator $\hat{u}_n$.

[9] We note that the elements of respective vectors and matrices are allowed to depend on the sample size, i.e., to form triangular arrays. This accommodates, among others, the case where the disturbances are generated from Cliff–Ord type models; see, e.g., Kelejian and Prucha (1999) on this point.

[10] At a later point we extend our results to the case involving multiple distances, each measured with error.

Assumption 6. There exist finite-dimensional vectors $z_{i,n}$ and Δ_n such that $\hat{u}_{i,n} - u_{i,n} = z_{i,n}\Delta_n$, and

$$n^{-1}\sum_{i=1}^{n}\|z_{i,n}\|^2 = O_p(1) \quad \text{and} \quad n^{1/2}\|\Delta_n\| = O_p(1).^{11}$$

2.2. Discussion of the assumptions

A widely used model for spatial correlation is the spatial AR(1) model introduced by Cliff and Ord (1973, 1981):

$$u_n = \rho W_n u_n + \varepsilon_n, \quad |\rho| < 1,$$

where ρ is a scalar parameter and W_n is an $n \times n$ spatial weighting matrix. This model is a variant of a model introduced by Whittle (1954), and can be viewed as a special case of (1) with[12]

$$R_n = (I_n - \rho W_n)^{-1}.$$

A typical assumption in the literature for this model is that R_n satisfies Assumption 2; see, e.g., Lee (2002, 2003, 2004) and Kelejian and Prucha (1998, 1999, 2004).[13] More generally, a special case of (1) is the spatial ARMA(p, q) model in which case

$$R_n = (I_n - \rho_1 W_{1,n} - \cdots - \rho_q W_{p,n})^{-1}[I_n + \lambda_1 M_{1,n} + \cdots + \lambda_q M_{q,n}],$$

where for $r = 1, \ldots, p$ and $s = 1, \ldots, q$, ρ_r and λ_s are scalar parameters, and $W_{r,n}$ and $M_{s,n}$ are spatial weighting matrices. On an intuitive level, if $a > b$ the ith row of $W_{a,n}$ selects neighbors which are more distant in some relevant space to the ith unit than does the ith row of $W_{b,n}$, and similarly for the ith rows of $M_{a,n}$ and $M_{b,n}$; see Anselin (1988, 2001b) for a further discussion.

Assumption 1 implies that the VC matrix of u_n is given by $\Sigma_n = R_n R_n'$. Assumption 2 then implies that the row and column sums of Σ_n are uniformly bounded in absolute value, and so the extent of correlation is restricted. In a time series context this condition ensures that the process possesses a fading memory. We note that the extent of correlation is necessarily restricted in virtually all large sample theory,—see, e.g. Amemiya (1985, Chapter 3, 4) and Pötscher and Prucha (1997, Chapter 5, 6).

Another implication of Assumptions 1 and 2 is, as is readily seen, that the qth moments of $u_{i,n}$ are uniformly bounded.

The spatial model specified in Section 4 is conditional on the exogenous variables and the weighting matrix, which are therefore taken as matrices of constants. The IVs used to estimate such spatial models are typically formulated in terms of the exogenous variables and the weighting matrix. Our Assumption 3 is consistent with this scenario.

Assumption 4(a) relates to the bandwidth of the SHAC estimator considered below. In essence, as will be seen, ℓ_n plays the same role as the bandwidth parameter (multiplied

[11] For definiteness, let A be some vector or matrix, then $\|A\| = [\mathrm{Tr}(A'A)]^{1/2}$. We note that this norm is submultiplicative, i.e., $\|AB\| \leq \|A\|\|B\|$.

[12] For a review of some applications of this model see, e.g., Anselin (2001a).

[13] Of course, if $\varepsilon_n = \Phi_n^{1/2}\eta_n$, where $\Phi_n^{1/2}$ is a diagonal matrix with non-negative uniformly bounded elements and the elements of η_n are i.i.d., then we can take $R_n = (I_n - \rho W_n)^{-1}\Phi_n^{1/2}$. The matrix R_n then satisfies Assumption 2, provided the row and column sums of $(I_n - \rho W_n)^{-1}$ are uniformly bounded in absolute value.

by two) in the time series literature in that in conjunction with a kernel function specified below it limits the number of sample covariances entering into the SHAC estimator to no more than $n\ell_n$. Clearly Assumption 4 implies that $\ell_n = o_p(n^\tau)$. Also observe that the bound $\tau_m = \frac{1}{2}(q-2)/(q-1) \leqslant \frac{1}{2}$; for $q = 4$ we have $\tau_m = \frac{1}{3}$ and as $q \to \infty$ we have $\tau_m \to \frac{1}{2}$.

Our consistency proof relies, in part, upon Chebyshev's inequality. Among other things, Assumptions 4(a) and (b) ensure that the bias and variance terms limit to zero. Along with our other assumptions, part (a) also ensures that the probability limit of the HAC estimator based on estimated disturbances is asymptotically equivalent to one which has the same form but is based on true disturbances. Among other things, part (b) restricts the extent of correlation in relation to the distances between cross sectional units, which are implicitly assumed to (eventually) increase as the sample size increases.[14] As an illustration and comparison with the time series literature consider the case where all units are arranged in an ordered fashion on a line, with the distance between the ith and jth unit equal to $|i-j|$ and the covariance between them equal to $\sigma_u^2\rho^{|i-j|}$ for some $|\rho|<1$. Then Assumption 4(b) would be satisfied for, e.g., $\rho_S = 1$ since $\sum_{i=1}^{\infty}|\rho^i|i = |\rho|/(1-|\rho|)^2 < \infty$. Assumption 4(b) generalizes this feature.

Assumption 5 specifies that the measurement errors are uniformly bounded and independent of the model disturbances. Clearly, the non-negativity of the measured distances implies that the measurement errors depend in part upon the "true" distances involved.[15]

Assumption 6 should be satisfied for most cases in which $\hat{u}_n$ is based upon $n^{1/2}$-consistent estimators of regression coefficients. For example, using evident notation, consider the non-linear regression model $y_{i,n} = f(x_{i,n},\beta) + u_{i,n}$. Let $\hat{\beta}_n$ denote the non-linear least squares estimator and let $\hat{u}_{i,n} = y_{i,n} - f(x_{i,n},\hat{\beta}_n)$. Assuming that f is differentiable and applying the mean value theorem it is readily seen that $\hat{u}_{i,n} - u_{i,n} = z_{in}\Delta_n$ with $z_{in} = [\partial f(x_{i,n},\underline{\beta})/\partial\underline{\beta}]_{\beta_n^*}$ where β_n^* is, element by element, between $\hat{\beta}_n$ and β, and $\Delta_n = \hat{\beta}_n - \beta$. Under typical assumptions maintained for the non-linear regression model, z_{in} and Δ_n will satisfy the conditions postulated in Assumption 6; cp., e.g., Pötscher and Prucha (1986).

3. Spatial HAC estimators

In this section, we first specify a class of kernel functions. We then suggest consistent SHAC estimators for Ψ_n based on this class, which determines weights for the different covariances as a function of measured distances between respective units. We first consider the case where the researcher employs a single distance measure, possibly measured with error, and provide a result for the consistency of the corresponding SHAC estimator. We then extend the discussion to the case where the researcher is unsure about the proper choice of a distance measure. We formulate a SHAC estimator that employs several distance measures, and demonstrate the consistency of the estimator as long as the "true" distance measure, possibly measured with error, is among those considered by the researcher.

[14] As an illustration, Assumption 4(b) would not be reasonable for cases in which the sample size increases because of more intensive sampling within a given distance, e.g., increased sampling within a given neighborhood. Cressie (1993, p. 57) refers to this case as " infill asymptotics".

[15] We note that Conley (1999) also maintains bounded measurement errors.

3.1. A single distance measure

Let $K(.)$ denote the kernel function. Then the weights for the different covariances will be of the form $K(d^*_{ij,n}/d_n)$ where $d^*_{ij,n} \geqslant 0$ and $d_n > 0$ are as in Assumptions 4 and 5.

Assumption 7. The kernel $K : \mathbf{R} \to [-1, 1]$, with $K(0) = 1$, $K(x) = K(-x)$, $K(x) = 0$ for $|x| > 1$, satisfies

$$|K(x) - 1| \leqslant c_K |x|^{\rho_K}, \quad |x| \leqslant 1, \tag{3}$$

for some $\rho_K \geqslant 1$ and $0 < c_K < \infty$.

Note that $K(d^*_{ij,n}/d_n) = K(d^*_{ji,n}/d_n)$ since $d^*_{ij,n} = d^*_{ji,n}$. Of course, since $d^*_{ij,n}/d_n \geqslant 0$ it would have sufficed to define K on the $\mathbf{R}_+$. We have specified the Kernel as is usual in the time series literature where frequently the difference between two time periods (rather than the absolute difference) divided by the truncation lag is used as an argument in the Kernel function.

Clearly if (3) holds for some $\rho_K \geqslant 1$, then it also holds for $\rho_K = 1$. The larger the value of ρ_K for which this condition is satisfied, the flatter and smoother the kernel will be at zero; compare, e.g., Pötscher and Prucha (1997, p. 129). We note that this condition is satisfied for many of the usual kernels such as the rectangular kernel, Bartlett or triangular kernel, the Parzen kernel, Tukey–Hanning kernel, Blackman–Tukey kernel, quadratic spectral kernel, exponential density kernel, etc.; see Brockwell and Davis (1991, pp. 359–361), and Andrews (1991).

Using evident notation, the (r, s)th element of the true VC matrix Ψ_n in (2) and our corresponding SHAC estimator of it are given by, respectively,

$$\psi_{rs,n} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ir,n} h_{js,n} \sigma_{ij,n}, \tag{4}$$

$$\hat{\psi}_{rs,n} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ir,n} h_{js,n} \hat{u}_{i,n} \hat{u}_{j,n} K(d^*_{ij,n}/d_n). \tag{5}$$

For later reference, let $\hat{\Psi}_n = (\hat{\psi}_{rs,n})$. For purposes of comparison with the time series literature consider, from a spatial perspective, the degenerate case where all units are arranged on a line in an ordered fashion and $d_{ij,n} = |i - j|$ which is measured without error. Let $h_{i.,n}$ denote the ith row of H_n. In this case, $\hat{\Psi}_n$ would reduce to

$$\begin{aligned} \hat{\Psi}_n = {} & K(0) n^{-1} \sum_{i=1}^{n} h'_{i.,n} h_{i.,n} \hat{u}^2_{i,n} \\ & + \sum\nolimits_{j=1}^{n-1} K(j/d_n) \sum_{i=1}^{n-j} [h'_{i.,n} h_{i+j.,n} + h'_{i+j.,n} h_{i.,n}] \hat{u}_{i,n} \hat{u}_{i+j,n}, \end{aligned} \tag{6}$$

which is an expression familiar from the time series literature, see, e.g., Pötscher and Prucha (1997, Chapter 12). On the other hand, it will typically be the case in a spatial context that $d_{ij,n} \neq d_{i+r,j+r,n}$ so that expressions comparable to (6) will not exist. The next theorem establishes the consistency of our estimator $\hat{\Psi}_n$ defined by (5).

Theorem 1. *Let* $\Psi_n = (\psi_{rs,n})$ *and* $\hat{\Psi}_n = (\hat{\psi}_{rs,n})$ *be as defined by* (4) *and* (5). *Given the model in* (1) *and Assumptions* 1–7,

$$\hat{\Psi}_n - \Psi_n = o_p(1).$$

Remark 1. Although the estimator $\hat{\Psi}_n$ is symmetric and consistent, it may not be positive semi-definite in finite samples. In many cases the distance measure will correspond to a Euclidean norm in $\mathbf{R}^p$, $p \geqslant 1$, and $d_{ij}^*/d = \|z_i - z_j\|_p = [\sum_{l=1}^p (z_{li} - z_{lj})^2]^{1/2}$, where the vectors $z_i = [z_{1i}, \ldots, z_{pi}]'$ describe certain characteristics of unit i, and where we have dropped subscripts n for notational convenience. A leading example would be the case of geographic distances, in which case z_i would be of dimension 2×1 and would contain the (normalized) geographic coordinates of unit i. We next discuss the case of Euclidean distance measures in more detail, and provide for this case general conditions regarding kernel functions that ensures that $\hat{\Psi}_n$ is positive semi-definite. Given the non-negativity of distance measures, our discussion will focus w.l.o.g. on the restriction of the kernel function to $[0, \infty)$. Let $\mathfrak{P}_p$ be the class of continuous functions $\varphi : [0, \infty) \to \mathbf{R}$ with $\varphi(0) = 1$ and where for all positive integers n the matrix

$$[\varphi(\|z_i - z_j\|_p)]_{i,j=1}^n$$

is positive semi-definite for any points $z_1, \ldots, z_n$ in $\mathbf{R}^p$. Clearly, if a kernel function K belongs to $\mathfrak{P}_p$ and $d_{ij}^*/d = \|z_i - z_j\|_p$, then $\hat{\Psi}_n$ is positive semi-definite. A complete description of the class $\mathfrak{P}_p$ is given in the seminal article by Schoenberg (1938).[16] He established that $\mathfrak{P}_p \supseteq \mathfrak{P}_{p+1}$ and that φ is an element of $\mathfrak{P}_p$ if and only if it is of the form

$$\varphi(x) = \Gamma\left(\frac{p}{2}\right) \int_0^\infty \left(\frac{2}{rx}\right)^{(p-2)/2} J_{(p-2)/2}(rx)\, \mathrm{d}F(r), \quad x \geqslant 0, \tag{7}$$

where F is a probability distribution function on $[0, \infty)$ and $J_{(p-2)/2}$ is a Bessel function of order $(p-2)/2$. The functions $\varphi(x)$ defined by (7) are $[(p-2)/2]$-times differentiable on $(0, \infty)$, where $[a]$ denotes the greatest integer less than or equal to a. The following result covers the triangular kernel and generalizations thereof: Consider the class of kernel functions

$$K_v(x) = \begin{cases} (1-x)^v, & 0 \leqslant x \leqslant 1, \\ 0, & x > 1. \end{cases}$$

Then $K_v(x)$ is an element of $\mathfrak{P}_p$ if and only if $v \geqslant (p+1)/2$. This result is due to by Golubov (1981). It establishes in particular that $K_1(x)$ belongs to $\mathfrak{P}_1$ and $K_2(x)$ belongs $\mathfrak{P}_1$ and $\mathfrak{P}_2$; for a recent reference and discussion of functions of the from (7) see, e.g., Gneiting (2002).

Remark 2. As a by-product of proving Theorem 1 we also obtain information on the rate of convergence. In particular, suppose the following additional assumptions

[16] Schoenberg's (1938) results utilizes Bochner's (1933) theorem, that establishes the equivalence between positive definite functions and characteristic functions of finite measures on $\mathbf{R}^p$. Bochner's theorem has been used widely, including by Priestley (1981) and Yaglom (1962) within the context of spectral density estimation. It also underlies results given, e.g., in Conley (1999) and Pötscher and Prucha (1997), regarding the positive semidefiniteness of smoothed periodogram estimators.

hold: $E\ell_n^2(d_n) = o(d_n^\eta)$, $d_n = O(n^{2\tau/\eta})$ and $d_n^{-1} = O(n^{-2\tau/\eta})$ for some $\eta > 0$.[17] Then we have $\hat{\psi}_{rs,n} - \psi_{rs,n} = O_p(\gamma_n)$ with

$$\gamma_n = \max\{n^{-1/2+1/q+\tau(1-1/q)}, n^{-2\rho_*\tau/\eta}\}$$

with $\rho_* = \min\{\rho_S, \rho_K\}$. The implied optimal rate for γ_n is achieved for $\tau_* = \frac{1}{2}[q-2]/[q-1+2q(\rho_*/\eta)] < \frac{1}{2}(q-2)/(q-1)$ and is given by $n^{-[(q-2)(\rho_*/\eta)]/[q-1+2q(\rho_*/\eta)]}$. We note that τ_* falls within the range postulated for τ in Assumption 4. As $q \to \infty$ we have $\tau_* \to 1/[2+4(\rho_*/\eta)]$ and the optimal rate is given by $n^{-(\rho_*/\eta)/[1+2(\rho_*/\eta)]}$.

3.2. Multiple distance measures

We now generalize the above result by allowing the researcher to consider several distance measure between units i and j, namely $d_{ij,m,n} = d_{ji,m,n}$, $m = 1, \ldots, M$. We allow for measurement errors in the measurement of these distances. In the following let $d^*_{ij,m,n} = d^*_{ji,m,n} \geqslant 0$ denote the distance measures employed by the researcher. Corresponding to each of these measures, we assume that the researcher can select a distance $d_{m,n} > 0$ such that $d_{m,n} \uparrow \infty$ as $n \to \infty$, $m = 1, \ldots, M$. For each unit $i = 1, \ldots, n$, let $\ell_{i,n}$ denote the number of units (neighbors) j for which $d^*_{ij,m,n} \leqslant d_{m,n}$ for at least one $m = 1, \ldots, M$, i.e.,

$$\ell_{i,n} = \sum_{j=1}^{n} \left(1 - \prod_{m=1}^{M} \mathbf{1}_{d^*_{ij,m,n} > d_{m,n}}\right),$$

and let $\ell_n = \max_{1 \leqslant i \leqslant n}(\ell_{i,n})$. Of course, if $M = 1$, then, dropping subscript m, we have $\ell_{i,n} = \sum_{j=1}^{n}(1 - \mathbf{1}_{d^*_{ij,n} > d_n}) = \sum_{j=1}^{n} \mathbf{1}_{d^*_{ij,n} \leqslant d_n}$, which is the expression for $\ell_{i,n}$ employed above in our discussion of the case of a single distance measure. We now replace Assumptions 4 and 5 by the following two assumptions.

Assumption 4*. (a) $E\ell_n^2 = o(n^{2\tau})$ where $\tau \leqslant \frac{1}{2}(q-2)/(q-1)$ and q is defined in Assumption 1; (b) $\sum_{j=1}^{n} |\sigma_{ij,n}| d^{\rho_S}_{ij,1,n} \leqslant c_S$ for some $\rho_S \geqslant 1$ and $0 < c_S < \infty$.

Assumption 5*. The distance measures employed by the researcher are given by

$$d^*_{ij,m,n} = d_{ij,m,n} + v_{ij,m,n} \geqslant 0,$$

where $v_{ij,m,n} = v_{ji,m,n}$ denotes the measurement errors, $|v_{ij,m,n}| \leqslant c_V$ with $0 < c_V < \infty$, and $\{(v_{ij,m,n}), m = 1, \ldots, M\}$ is independent of $(\varepsilon_{i,n})$.

Remark 3. Assumption 4*(b) is postulated to hold w.l.o.g. for the first distance measure, since we can always relabel the measures. It is important to note that we do not assume that the researcher knows the distance measure for which Assumption 4*(b) holds. We only postulate that the set of measures considered by the researcher contains the "true" distance measure, i.e., the measure for which Assumption 4*(b) holds.

[17] For example, and as discussed in more detail in our Monte Carlo study below, if spatial units are located on a square grid (of respective length one) and the distance $d_{ij,n}$ between units is given by the Euclidean distance, then $\ell_n(d_n) \leqslant 4(d_n + c_V)^2 - 4(d_n + c_V) + 4$, and thus $E\ell_n^2(d_n) = o(d_n^\eta)$ for $\eta > 4$.

Our SHAC estimator for the (r,s)th element of the true VC matrix Ψ_n defined by (4) is, in the present case of multiple distance measures, now given by

$$\hat{\psi}_{rs,n} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ir,n} h_{js,n} \hat{u}_{i,n} \hat{u}_{j,n} K\left(\min_{m} \{ d^{*}_{ij,m,n} / d_{m,n} \} \right). \tag{8}$$

As is evident from the specification, the estimator includes all covariance terms for which at least one of the ratios $d^{*}_{ij,m,n}/d_{m,n}$ is less than one. The next theorem establishes the consistency of our estimator $\hat{\Psi}_n$ defined by (8).

Theorem 2. *Let* $\Psi_n = (\psi_{rs,n})$ *and* $\hat{\Psi}_n = (\hat{\psi}_{rs,n})$ *be as defined by* (4) *and* (8). *Given the model in* (1) *and Assumptions* 1–3, 4*, 5*, 6, 7,

$$\hat{\Psi}_n - \Psi_n = o_p(1).$$

Clearly Theorem 2 is a generalization of Theorem 1. The importance of this generalization is that in practice researchers can consider "many" distance measures between units and base their SHAC estimator on just the minimum of the measured ratios described in (8).

4. A general spatial regression model

In this section we derive the limiting distribution of an IV estimator for the parameters of a single Cliff–Ord type spatial equation, which may be part of a system of equations, and may contain spatial lags of the exogenous, as well as endogenous variables. Formal estimation of such models in the single equation case has been recently considered by, e.g., Lee (2002, 2003, 2004) and Kelejian and Prucha (1998, 1999), and in the systems case by Kelejian and Prucha (2004). In contrast to this earlier literature, in this paper we do not impose any specific structure on the disturbance process apart from that provided by (1). We also allow for endogenous regressors which could be generated by a non-linear model. We derive the limiting distribution of the IV estimator of the regression parameters. It will become clear that, in light of Theorem 2, the VC matrix involved in that distribution can be consistently estimated.

4.1. Model specification

Consider the Cliff–Ord type regression model

$$y_n = X_n \beta_0 + \lambda_0 W_n y_n + Y_n \gamma_0 + u_n, \quad |\lambda_0| < 1, \tag{9}$$

where y_n is an $n \times 1$ vector of observations on the dependent variable corresponding to n cross sectional units, X_n is a corresponding $n \times k_x$ matrix of observations on k_x non-stochastic regressors, W_n is an $n \times n$ weighting matrix of known constants, Y_n is a corresponding $n \times r_y$ matrix of observations on r_y endogenous variables, u_n is the disturbance vector, and β_0, λ_0, and γ_0 are correspondingly defined parameters. The disturbance vector u_n is assumed to be generated according to (1), which allows for general patterns of spatial correlation and heteroscedasticity.

In the above model the (i,j)th element of the weighting matrix, $w_{ij,n}$, would typically be taken to be non-zero only if units i and j are related in a meaningful way, in which case these units are said to be neighbors. The non-zero elements of W_n would also typically be assumed to decline as a measure of distance between the corresponding units increases.

The distance measure could relate to geographic space, technology space, etc.[18] In the literature $W_n y_n$ is said to be the spatial lag of y_n. Our specification allows for the elements of all data vectors and matrices to depend on the sample size. Consequently the specification allows for X_n and Y_n to contain, respectively, the spatial lags of some or all of the considered exogenous and endogenous variables. In the following it will be convenient to express (9) more compactly as

$$y_n = Z_n \delta_0 + u_n, \tag{10}$$

where $Z_n = (X_n, W_n y_n, Y_n)$ and $\delta_0' = (\beta_0', \lambda_0, \gamma_0')$.

In the following we specify a set of assumptions for model (10), which are in addition to Assumptions 1, 2 and 4* for the disturbance process.

Assumption 8. All diagonal elements of W_n are zero.

Assumption 9. (a) $(I - \lambda_0 W_n)$ is non-singular. (b) The row and column sums of W_n, and $(I - \lambda_0 W_n)^{-1}$ are bounded uniformly in absolute value by some constant c_W, $0 < c_W < \infty$.

Assumption 10. X_n has full column rank (for n large enough), and its elements are uniformly bounded in absolute value by constant c_X, $0 < c_X < \infty$.

We will estimate the model by an IV procedure. Towards this end again let H_n be the $n \times p_h$ matrix of instruments. In this section we maintain the following assumptions concerning H_n which are an extension of those given in Assumption 3.

Assumption 11. The (non-stochastic) instrument matrix H_n has full column rank $p_h \geqslant k_x + r_y + 1$ for n large enough, and its elements are uniformly bounded in absolute value by some constant c_H, $0 < c_H < \infty$. It contains at least the linearly independent columns of $(X_n, W_n X_n)$. Furthermore, H_n has the following properties:

(a) $Q_{HH} = \lim_{n\to\infty} n^{-1} H_n' H_n$ is a finite non-singular matrix;
(b) $Q_{HZ} = p \lim_{n\to\infty} n^{-1} H_n' Z_n$ is a finite matrix which has full column rank $k_x + r_y + 1$;
(c) $\Psi = \lim_{n\to\infty} n^{-1} H_n' \Sigma_n H_n$ is a finite non-singular matrix where, again, $\Sigma_n = E(u_n u_n') = R_n R_n'$.

The above assumptions are consistent with those maintained for Cliff–Ord type models in the recent literature. For further discussions and interpretations see, e.g., Kelejian and Prucha (2004). It seems of interest to further comment on the choice of the instruments. Clearly the optimal instruments for Z_n are $EZ_n = [X_n, E(W_n y_n), E(Y_n)]$. Solving (9) "partially" for y_n and assuming that the roots of $\lambda_0 W_n$ are less than unity in absolute value[19] yields

$$\begin{aligned} Ey_n &= (I_n - \lambda_0 W_n)^{-1}(X_n \beta_0 + EY_n \gamma_0) \\ &= \sum_{s=0}^{\infty} \lambda_0^i W_n^i (X_n \beta_0 + EY_n \gamma_0). \end{aligned} \tag{11}$$

[18] For a discussion of some weighting matrix formulations see Anselin (1988, Chapter 3) and Kelejian and Robinson (1995).

[19] In many models the weighting matrix is row normalized and it is assumed that $|\lambda_0| < 1$. In this case, as well as in others, the roots of $\lambda_0 W_n$ would be less than unity in absolute value and so the expansion in (11) holds.

Under reasonable conditions, if (9) represents one equation of a linear system of equations, the discussion in Kelejian and Prucha (2004) implies that $EY_n = \sum_{s=0}^{\infty} W_n^s \underline{X}_n \Pi_s$, where $\underline{X}_n$ denotes the matrix of all exogenous variables in the system and Π_s are (reduced form) parameter matrices. In applications we may now try to approximate the optimal instruments by selecting H_n such that it includes the independent columns of $(\underline{X}_n, W_n\underline{X}_n, \ldots, W_n^g\underline{X}_n)$ for some $g \geqslant 1$. Of course, if all exogenous variables in the system are not observed, we may construct instruments based only on the observed set. We also note that if the elements of the exogenous variables are uniformly bounded, then so will be the elements of $W_n^s\underline{X}_n$ for $s \geqslant 1$, and thus our selection of instruments will be consistent with Assumption 11. Similarly, if one or more of the elements of Y_n are generated by a non-linear model, we are effectively assuming that the elements of $E(Y_n)$ are uniformly bounded and relate to X_n, W_nX_n and, perhaps, to other variables that may or may not be included in H_n. A final point should be noted concerning Assumption 11(b), which in essence ensures that the instruments H_n identify the parameters. Suppose $\beta_0 = 0$ and $\gamma_0 = 0$ so that $E(W_n y_n) = 0$. It should be clear, in this case, that Assumption 11(b) will not hold. Therefore, the estimation theory presented below will not enable the researcher to test the hypothesis that $\beta_0 = 0$ and $\gamma_0 = 0$. However, unlike for the case described in Kelejian and Prucha (1998), our results do enable the researcher to test the hypothesis $\beta_0 = 0$. Indeed, given our other assumptions, an analysis which is virtually identical to that in Kelejian and Prucha (1998) will demonstrate that our results and corresponding testing procedures only require that at least one element of β_0 or γ_0 be non-zero.[20]

The next assumption bounds the third absolute moments of the elements $y_{ir,n}$ of Y_n.

Assumption 12. The expectations $E|y_{ir,n}|^3$ are uniformly bounded by some constant c_Y, $0 < c_Y < \infty$.

Assumption 1 maintains that the innovations have uniformly bounded fourth moments. Given this, Assumption 12 should be satisfied for typical specifications of Cliff–Ord type models. In particular, if (9) represents one equation of a linear system of equations such as that considered in Kelejian and Prucha (2004), but with the disturbance processes allowed to be of the more general form considered in this paper, then Assumption 12 holds. This follows since the demonstration in the earlier paper that the third absolute moments of the endogenous variables as well as their spatial lags are uniformly bounded did not depend on the specific structure of the disturbance process and only used the features of Assumptions 1 and 2 in this paper.

4.2. Instrumental variable estimation

We next define a spatial 2SLS estimator which is based on the instruments H_n. We derive its asymptotic distribution and provide a consistent estimator for its VC matrix that utilizes the SHAC estimator considered in Section 4. In particular, let $\hat{Z}_n = P_n Z_n$ where $P_n = H_n(H_n'H_n)^{-1}H_n'$; then the spatial 2SLS estimator for the parameter vector δ of (10) is given by

$$\hat{\delta}_n = (\hat{Z}_n' Z_n)^{-1} \hat{Z}_n' y_n. \tag{12}$$

[20]If the weighting matrix is row normalized, that non-zero element must correspond to a non-constant regressor—see, e.g., Kelejian and Prucha (1998).

Let $\hat{u}_n = (\hat{u}_{1,n}, \ldots, \hat{u}_{n,n})$ denote the 2SLS residuals, i.e., $\hat{u}_n = y_n - Z_n\hat{\delta}_n$. Based on those residuals and H_n, let $\hat{\Psi}_n = (\hat{\psi}_{rs,n})$, as given in (8), be the corresponding SHAC estimator of $\Psi_n = n^{-1}H_n'\Sigma_n H_n$ and of its limit, $\Psi = \lim_{n\to\infty} n^{-1}H_n'\Sigma_n H_n$. We can now give the following theorem concerning the asymptotic distribution of $\hat{\delta}_n$ and the consistent estimation of its asymptotic VC matrix.

Theorem 3. *Assume the disturbance specification in* (1), *the model in* (10), *and Assumptions* 1, 2, 4*, 5*, *and* 7–12. *Then* (a) $n^{1/2}(\hat{\delta}_n - \delta_0) \xrightarrow{d} N(0, \Phi)$ *and* (b) $\hat{\Phi}_n \xrightarrow{p} \Phi$ *as* $n \to \infty$, *where*

$$\Phi = (Q_{HZ}'Q_{HH}^{-1}Q_{HZ})^{-1}Q_{HZ}'Q_{HH}^{-1}\Psi Q_{HH}^{-1}Q_{HZ}(Q_{HZ}'Q_{HH}^{-1}Q_{HZ})^{-1},$$

$$\hat{\Phi}_n = n^2(\hat{Z}_n'\hat{Z}_n)^{-1}Z_n'H_n(H_n'H_n)^{-1}\hat{\Psi}_n(H_n'H_n)^{-1}H_n'Z_n(\hat{Z}_n'\hat{Z}_n)^{-1}.$$

Given Theorem 3, small sample inferences concerning δ_0 can be based on the approximation $\hat{\delta}_n \sim N(\delta_0, n^{-1}\hat{\Phi}_n)$.

5. A Monte Carlo study

In this section, we give some illustrative Monte Carlo results which suggest that our SHAC estimator performs reasonably well in finite samples.

5.1. Monte Carlo design

Our Monte Carlo design is influenced by the widely used format for the analysis of spatial regression models by Anselin and Rey (1991) and Anselin and Florax (1995), as well as by the format used by Andrews (1991) and Andrews and Monahan (1992) for the analysis of HAC estimators. In particular, we consider the following special case of the spatial model considered above:

$$y_n = X_n\beta_0 + u_n,$$

$$u_n = \rho_0 W_n u_n + \varepsilon_n, \quad |\rho_0| < 1, \tag{13}$$

with $X_n = [e_n, x_n]$ consisting of two regressors (one of which is the intercept), and $\beta_0 = [a_0, b_0]' = [1, 5]'$. That is, we consider a linear regression model where the disturbances follow a first-order Cliff–Ord spatial autoregressive process. The spatial units are assumed to be located on a square grid at locations $\{(r, s) : r, s = 0, 1, \ldots, m\}$, and thus the total number of units is $n = (m+1)^2$. The distance d_{ij} between units is given by the Euclidean distance. As, e.g., in Baltagi et al. (2003) the weights matrix W_n is taken to be a rook-type matrix where two units are neighbors if their Euclidean distance is less than or equal to one. The weights matrix is normalized such that the weights in each row sum to one. Given this $I_n - \rho_0 W_n$ is non-singular, and hence $u_n = (I_n - \rho_0 W_n)^{-1}\varepsilon_n$. The $\varepsilon_{i,n}$ are taken to be i.i.d. standardized normal, and thus

$$\Sigma_n = Eu_n u_n' = (I_n - \rho_0 W_n)^{-1}(I_n - \rho_0 W_n')^{-1}. \tag{14}$$

The OLS estimator and its (normalized) VC matrix, conditional on the regressors, are given by

$$\hat{\beta}_n = [\hat{a}_n, \hat{b}_n]' = (X_n'X_n)^{-1}X_n'y_n,$$

$$VC(\sqrt{n}(\hat{\beta}_n - \beta_0) \mid X_n) = (n^{-1}X_n'X_n)^{-1}[n^{-1}X_n'Eu_nu_n'X_n](n^{-1}X_n'X_n)^{-1}. \tag{15}$$

Analogous to Andrews (1991) and Andrews and Monahan (1992) we consider the case where $n^{-1}X_n'X_n = I_n$, and the estimand of interest in our Monte Carlo study is taken to be the variance of the least squares estimator corresponding to the slope parameter b_0, i.e., the (normalized) variance of $\hat{b}_n$. Given our setup this variance is given by

$$\begin{aligned}\psi_n &= var(\sqrt{n}(\hat{b}_n - b_0)|x_n) \\ &= n^{-1}x_n'Eu_nu_n'x_n = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} x_{i,n}x_{j,n}Eu_{i,n}u_{j,n},\end{aligned} \tag{16}$$

where $x_{i,n}$ denotes the ith element of x_n. The SHAC estimator for ψ_n is given by

$$\hat{\psi}_{\mathrm{HAC},n} = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} x_{i,n}x_{j,n}\hat{u}_{i,n}\hat{u}_{j,n}K(d_{ij,n}^*/d_n), \tag{17}$$

where $\hat{u}_{i,n}$ denotes the ith element of the OLS residual vector $\hat{u}_n = y_n - X_n\hat{\beta}_n$. We also compute the estimator for ψ_n corresponding to the " classical" OLS VC matrix estimator. Since $n^{-1}X_n'X_n = I_n$ this estimator is simply given by $\hat{\psi}_{\mathrm{OLS},n} = n^{-1}\sum_{i=1}^{n}\hat{u}_{i,n}^2$. Our Monte Carlo results relate to five experimental values for ρ_0, namely $(-0.8, -0.5, 0, 0.5, 0.8)$, and two sample sizes, namely $n = 400$ and 1024. We also consider a case in which measurement errors are absent, and one in which they are not. In all 20 of our experiments we used the Parzen kernel, which is consistent with Assumption 7.

For each Monte Carlo iteration we draw a set of innovations ε_n from a standardized normal distribution. For a given regressor vector x_n we can then generate y_n from the above model. The elements of the $n \times 1$ vector x_n are generated via the following simple spatial autoregressive model: $x_n = 0.3W_nx_n + \zeta_n$ where the elements of ζ_n are i.i.d. draws from a uniform distribution over the interval $[0, 1]$. The elements of x_n are then further standardized by subtracting the sample mean and dividing each observation by the sample standard deviation so that $n^{-1}X_n'X_n = I_n$. Although the estimand of interest is the variance ψ_n, which is conditional on x_n, we randomly draw, similar to Andrews (1991), a new set of regressor vectors x_n for each repetition of the experiment (in the above described manner) to reduce the dependence of the results on particular realizations of x_n. As a result the value of the estimand ψ_n will vary across repetitions. Among other things, in the tables we will report its average value across repetitions.

Our first set of experiments relates to the case in which measurement errors are absent in the distance measure, i.e., $d_{ij,n}^* = d_{ij}$ where d_{ij} stands for the Euclidean distance between units i and j. It is not difficult to check that if $\ell_n(d_n)$ is the maximum numbers of neighbors that satisfy $d_{ij} \leqslant d_n$, then $\ell_n(d_n) \leqslant 4d_n^2 - 4d_n + 4$ and thus if $d_n = o(n^\kappa)$, then $\ell_n = o(n^{2\kappa})$. In our Monte Carlo study we took $d_n = [n^{1/4}]$, where $[z]$ denotes the nearest integer that is less than or equal to z.[21]

In our second set of experiments we allow for errors of measurement. Specifically, in this case we assume the "true" Euclidean distance d_{ij} is measured as $d_{ij,n}^* = d_{ij} + v_{ij,n}$ where $v_{ij,n}$ denotes a random measurement error. For the case in which $d_{ij} = 1$, we take $P(v_{ij,n} = 0) = P(v_{ij,n} = 1) = \frac{1}{2}$; for the case in which $d_{ij} \geqslant 2$ we take $P(v_{ij,n} = -1) = P(v_{ij,n} = 0) + P(v_{ij,n} = 1) = \frac{1}{3}$.

[21] In a future larger Monte Carlo study it may be of interest to consider specifications where $d_n = c[n^{1/4}]$ for different values of c, as well as various other variations of the Monte Carlo design.

For each replication we generate a set of measurement errors which are independent of the disturbances and regressors.

5.2. Monte Carlo results

Table 1 gives results relating to small sample biases and RMSEs for our SHAC estimator $\hat{\psi}_{\mathrm{HAC},n}$ based on the Parzen kernel for sample sizes $n = 400$ and 1024 for the case in which the measurement errors are zero. These results are based on 1000 Monte Carlo replications. As expected, for all cases considered, the RMSEs are lower for the sample size 1024 than for $n = 400$. For sample size $n = 400$ the magnitude of the biases and RMSEs corresponding to the Parzen kernel are on average roughly 8% and 23%, respectively, of the true value of ψ_n. For sample size $n = 1024$ their magnitudes are on average roughly 4% and 16%, respectively. These results are encouraging. The RMSEs are lowest for the case in which $\rho = 0$.

Table 1 also reports results relating to small sample biases and RMSEs of the "classical" OLS VC matrix estimator $\hat{\psi}_{\mathrm{OLS},n}$. For the case in which $\rho = 0$ this estimator is consistent, and known to perform well. As expected, for $\rho = 0$ both bias and RMSE of the estimator $\hat{\psi}_{\mathrm{OLS},n}$ are small. For $\rho \neq 0$ the estimator is generally inconsistent. As expected, the bias increases with $|\rho|$, and the bias remains high even as the sample size increases.

The results in Table 1 support our theoretical findings for $\hat{\psi}_{\mathrm{HAC},n}$. Given the limited nature of our experiments, these results can only provide limited information regarding the relative performance of $\hat{\psi}_{\mathrm{HAC},n}$ and $\hat{\psi}_{\mathrm{OLS},n}$. More informative small sample comparisons would have to be based on a wider Monte Carlo study involving, among other things,

Table 1
Bias and RMSE of spatial HAC estimator $\hat{\psi}_{\mathrm{HAC},n}$ and OLS estimator $\hat{\psi}_{\mathrm{OLS},n}$ of the variance of $\hat{b}_n$, ψ_n: distance without measurement error

ρ	ψ_n	HAC (Parzen kernel)		OLS	
		Bias	RMSE	Bias	RMSE
$n = 400$					
0.8	3.428	−0.452	0.911	−1.073	1.201
0.5	1.516	−0.125	0.315	−0.261	0.292
0	1.000	−0.038	0.186	−0.002	0.073
−0.5	1.062	−0.002	0.194	0.204	0.234
−0.8	1.722	0.051	0.375	0.704	0.792
Column average (of absolute values)		0.134	0.396	0.449	0.518
$n = 1024$					
0.8	3.352	−0.248	0.622	−1.014	1.068
0.5	1.506	−0.067	0.231	−0.250	0.263
0	1.000	−0.020	0.140	−0.001	0.044
−0.5	1.058	−0.001	0.147	0.199	0.211
−0.8	1.682	0.024	0.262	0.672	0.707
Column average (of absolute values)		0.072	0.280	0.427	0.458

Table 2
Bias and RMSE of spatial HAC estimator $\hat{\psi}_{HAC,n}$ of the variance of $\hat{b}_n$, ψ_n: distance with measurement errors

ρ	ψ_n	HAC (Parzen kernel)	
		Bias	RMSE
$n = 400$			
0.8	3.417	−0.634	1.006
0.5	1.513	−0.173	0.345
0	1.000	−0.038	0.203
−0.5	1.064	0.044	0.228
−0.8	1.731	0.227	0.497
Column averages (of absolute values)		0.223	0.456
$n = 1024$			
0.8	3.340	−0.388	0.652
0.5	1.504	−0.104	0.245
0	1.000	−0.021	0.150
−0.5	1.058	0.029	0.167
−0.8	1.682	0.138	0.331
Column averages (of absolute values)		0.136	0.309

heteroscedastic innovations ε_n, various weighting matrices, as well as various dependence structures for x_n, etc.

Finally, for illustrative purposes, Table 2 gives results relating to the use of the Parzen kernel for the case in which there are errors of measurement concerning distances. The cases considered are the same as those in Table 1. All of the results are based on 500 Monte Carlo repetitions.[22] A glance at Table 2 suggests that for all cases considered the biases and RMSEs decrease as the sample size increases. Also, although the RMSEs are generally larger than corresponding values in Table 1, on average the difference is, roughly, only 15% and 10% for $n = 400$ and 1024, respectively.

6. Conclusion and suggestions for future research

In this paper we suggested a spatial HAC (SHAC) estimator of a VC matrix in a spatial framework, and demonstrated the consistency of that estimator. An important aim of this paper was to establish that consistency under a set of relatively simple assumptions that covers, among others, the important and widely used class of Cliff–Ord type models. Our assumptions allow the researcher to be unsure about which distance measure to use in the SHAC estimator, as well as for measurement errors in the distance measures considered. Our consistency result is also generic in the sense that residuals may correspond to a variety of linear and non-linear models, provided they are $n^{1/2}$-consistently estimated.

[22]The smaller number of Monte Carlo repetitions was considered because for $n = 1024$ even for this smaller number of iterations, each Monte Carlo experiment corresponding to a particular value of ρ_0 took four days to complete on a fast PC. We also note that because of the difference in the number of Monte Carlo repetitions the numbers for ψ_n are slightly different in the two tables.

In this paper we also derived the asymptotic distribution of an IV estimator for the parameters of a general spatial model and demonstrated that a consistent estimator of the VC matrix involved can be based on our suggested SHAC procedure.

Finally, we gave Monte Carlo results which suggest that our SHAC estimator performs reasonably well in small samples. Our Monte Carlo study was based on a limited number of experiments relating to model parameter values, only one kernel, one weighting matrix, one process generating the regressors x_n, one process for the innovations ε_n, and only one specification of measurement errors relating to the measured distances underlying our SHAC estimator. Therefore, one suggestion for future research would be to expand that Monte Carlo study to one which has a wider scope of experimental model parameter values, various spatial weights matrices as well as kernels, various generating mechanisms for x_n and ε_n, and more than one specification of distance measurement errors. As part of such a study it would be of interest to explore bandwidth selection issues as well as the small sample properties of Wald-type test statistics that involve the SHAC estimator.

Appendix A

Proof of Theorem 1. Theorem 1 is a special case of Theorem 2, which is proven below. □

Proof of Theorem 2. The (r,s)th element of Ψ_n and its corresponding SHAC estimator $\hat{\Psi}_n$ as given in (4) and (8) are

$$\psi_{rs,n} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ir,n} h_{js,n} \sigma_{ij,n}, \tag{A.1}$$

$$\hat{\psi}_{rs,n} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ir,n} h_{js,n} \hat{u}_{i,n} \hat{u}_{j,n} K(\min_m \{d^*_{ij,m,n}/d_{m,n}\}). \tag{A.2}$$

Clearly

$$\hat{\psi}_{rs,n} - \psi_{rs,n} = a_{rs,n} + b_{rs,n} + c_{rs,n}, \tag{A.3}$$

with

$$a_{rs,n} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ir,n} h_{js,n} [\hat{u}_{i,n} \hat{u}_{j,n} - u_{i,n} u_{j,n}] K\Big(\min_m \{d^*_{ij,m,n}/d_{m,n}\}\Big),$$

$$b_{rs,n} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ir,n} h_{js,n} [u_{i,n} u_{j,n} - \sigma_{ij}] K\Big(\min_m \{d^*_{ij,m,n}/d_{m,n}\}\Big),$$

$$c_{rs,n} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ir,n} h_{js,n} \sigma_{ij,n} \Big[K\Big(\min_m \{d^*_{ij,m,n}/d_{m,n}\}\Big) - 1\Big].$$

To prove that $\hat{\psi}_{rs,n} - \psi_{rs,n} = o_p(1)$ we show that each term on the r.h.s of (A.3) is $o_p(1)$.

(a) *Proof that* $a_{rs,n} = o_p(1)$: Observe that in light of Assumption 6

$$|\hat{u}_{i,n}\hat{u}_{j,n} - u_{i,n}u_{j,n}| \leqslant |u_{i,n}| \|z_{j,n}\| \|\Delta_n\| + |u_{j,n}| \|z_{i,n}\| \|\Delta_n\| + \|z_{i,n}\| \|z_{j,n}\| \|\Delta_n\|^2.$$

Recalling that $|h_{ir,n}| \leqslant c_H$, observing that $|K(\min_m \{d^*_{ij,m,n}/d_{m,n}\})| \leqslant 1 - \prod_{m=1}^M \mathbf{1}_{d^*_{ij,m,n} > d_{m,n}}$, and utilizing the above inequality we have

$$|a_{rs,n}| \leqslant n^{-1} \sum_{i=1}^n \sum_{j=1}^n \left| h_{ir,n} \| h_{js,n} \| \hat{u}_{i,n} \hat{u}_{j,n} - u_{i,n} u_{j,n} \| K\left(\min_m \{ d^*_{ij,m,n}/d_{m,n} \} \right) \right|$$
$$\leqslant A^{(1)}_{rs,n} + A^{(2)}_{rs,n} + A^{(3)}_{rs,n},$$

$$A^{(1)}_{rs,n} = c_H^2 \| \Delta_n \| n^{-1} \sum_{i=1}^n \sum_{j=1}^n \left(1 - \prod_{m=1}^M \mathbf{1}_{d^*_{ij,m,n} > d_{m,n}} \right) |u_{i,n}| \| z_{j,n} \|,$$

$$A^{(2)}_{rs,n} = c_H^2 \| \Delta_n \| n^{-1} \sum_{i=1}^n \sum_{j=1}^n \left(1 - \prod_{m=1}^M \mathbf{1}_{d^*_{ij,m,n} > d_{m,n}} \right) |u_{j,n}| \| z_{i,n} \|,$$

$$A^{(3)}_{rs,n} = c_H^2 \| \Delta_n \|^2 n^{-1} \sum_{i=1}^n \sum_{j=1}^n \left(1 - \prod_{m=1}^M \mathbf{1}_{d^*_{ij,m,n} > d_{m,n}} \right) \| z_{i,n} \| \| z_{j,n} \|.$$

Also observe that in light of the definition of ℓ_n given before Assumption 4* we have $\sum_{j=1}^n (1 - \prod_{m=1}^M \mathbf{1}_{d^*_{ij,m,n} > d_{m,n}}) \leqslant \ell_n$. Let q be as in Assumption 1. It then follows from Hölder's inequality that

$$A^{(1)}_{rs,n} \leqslant c_H^2 \| \Delta_n \| n^{-1} \sum_{j=1}^n \| z_{j,n} \| \left[\sum_{i=1}^n \left(1 - \prod_{m=1}^M \mathbf{1}_{d^*_{ij,m,n} > d_{m,n}} \right) \right]^{1-1/q} \left[\sum_{i=1}^n |u_{i,n}|^q \right]^{1/q}$$
$$\leqslant c_H^2 n^{-1/2+1/q} \ell_n^{1-1/q} [n^{1/2} \| \Delta_n \|] \left[n^{-1} \sum_{j=1}^n \| z_{j,n} \| \right] \left[n^{-1} \sum_{i=1}^n |u_{i,n}|^q \right]^{1/q}.$$

As remarked in the text, in light of Assumptions 1 and 2 it is readily seen that $\mathrm{E}|u_{i,n}|^q \leqslant const < \infty$. Because of this and the other maintained assumptions all terms in square brackets in the last inequality are seen to be $\mathrm{O_p}(1)$. Now consider

$$\pi^{(1)}_n = n^{-1/2+1/q} \ell_n^{1-1/q} = n^{-1/2+1/q+\tau(1-1/q)} [n^{-\tau} \ell_n]^{1-1/q}.$$

In light of Assumption 4*(a) we have $\ell_n = \mathrm{o_p}(n^\tau)$ with $\tau \leqslant \tau_m = \frac{1}{2}(q-2)/(q-1)$. Observing further that $-\frac{1}{2} + 1/q + \tau(1-1/q) \leqslant -\frac{1}{2} + 1/q + \tau_m(1-1/q) = 0$ clearly $\pi^{(1)}_n = \mathrm{o_p}(1)$ and hence $A^{(1)}_{rs,n} = \mathrm{o_p}(1)$. By the same arguments it follows that also $A^{(2)}_{rs,n} = \mathrm{o_p}(1)$. Applying Hölder's inequality we see further that

$$A^{(3)}_{rs,n} \leqslant c_H^2 \| \Delta_n \|^2 n^{-1} \sum_{j=1}^n \| z_{j,n} \| \left[\sum_{i=1}^n \left(1 - \prod_{m=1}^M \mathbf{1}_{d^*_{ij,m,n} > d_{m,n}} \right) \right]^{1/2} \left[\sum_{i=1}^n \| z_{i,n} \|^2 \right]^{1/2}$$
$$\leqslant c_H^2 n^{-1/2} \ell_n^{1/2} \left[n^{1/2} \| \Delta_n \| \right]^2 \left[n^{-1} \sum_{j=1}^n \| z_{j,n} \| \right] \left[n^{-1} \sum_{i=1}^n \| z_{i,n} \|^2 \right]^{1/2}.$$

In light of Assumption 6 all terms in square brackets in the last inequality are $\mathrm{O_p}(1)$. Now consider

$$\pi^{(3)}_n = n^{-1/2} \ell_n^{1/2} = n^{-(1/2)(1-\tau)} [n^{-\tau} \ell_n]^{1/2}.$$

Since $\ell_n = o_p(n^\tau)$ with $\tau \leq \frac{1}{2}(q-2)/(q-1) \leq \frac{1}{2}$ by Assumption 4*(a) clearly $\pi_n^{(3)} = o_p(1)$ and hence $A_{rs,n}^{(3)} = o_p(1)$. Thus $a_{rs,n} = o_p(1)$.

(b) *Proof that* $b_{rs,n} = o_p(1)$: Substitution of $u_{i,n} = \sum_{l=1}^n r_{il,n}\varepsilon_{l,n}$ into the expression for $b_{rs,n}$ yields

$$b_{rs,n} = n^{-1} \sum_{l=1}^n \sum_{k=1}^n \gamma_{lk,n}[\varepsilon_{l,n}\varepsilon_{k,n} - E\varepsilon_{l,n}\varepsilon_{k,n}]$$

with $\gamma_{lk,n} = \sum_{i=1}^n \sum_{j=1}^n h_{ir,n}h_{js,n}r_{il,n}r_{jk,n}K(\min_m\{d^*_{ij,m,n}/d_{m,n}\})$. Let $V_n = \{(v_{ij,m,n}), m = 1,\ldots,M\}$ be the matrix of measurement errors and let ε_n be the vector of innovations. Since V_n and ε_n are independent by Assumption 5* clearly $Eb_{rs,n} = 0$. It hence suffices to show that $var(b_{rs,n}) = o(1)$. The variance of $b_{rs,n}$ conditional on V_n is—see, e.g., Kelejian and Prucha (2001, p. 227)—given by

$$var(b_{rs,n}|V_n) = 2n^{-2} \sum_{l=1}^n \sum_{k=1}^n \gamma_{lk,n}^2 + n^{-2} \sum_{l=1}^n \gamma_{ll}^2[E\varepsilon_{l,n}^4 - 3].$$

Next observe that

$$\sum_{k=1}^n |\gamma_{lk,n}| \leq c_H^2 \sum_{i=1}^n |r_{il,n}| \sum_{j=1}^n \left(1 - \prod_{m=1}^M \mathbf{1}_{d^*_{ij,m,n} > d_{m,n}}\right) \sum_{k=1}^n |r_{jk,n}| \leq c_H^2 c_R^2 \ell_n,$$

where we have utilized that $\sum_{j=1}^n (1 - \prod_{m=1}^M \mathbf{1}_{d^*_{ij,m,n} > d_{m,n}}) \leq \ell_n$ as well as Assumption 2; hence $\sum_{k=1}^n |\gamma_{lk,n}|^2 \leq [\sum_{k=1}^n |\gamma_{lk,n}|]^2 \leq c_H^4 c_R^4 \ell_n^2$. The variance of $b_{rs,n}$ is now given by

$$E\, var(b_{rs,n}|V_n) \leq E\left\{2n^{-2} \sum_{l=1}^n c_H^4 c_R^4 \ell_n^2 + n^{-2}(c_E + 4) \sum_{l=1}^n c_H^4 c_R^4 \ell_n^2\right\} \leq const * n^{-1} E\ell_n^2.$$

Observing again that $E\ell_n^2 = o(n^{2\tau})$ with $\tau \leq \frac{1}{2}(q-2)/(q-1) \leq 1/2$ shows that $var(b_{rs,n}) = o(1)$ and hence $b_{rs,n} = o_p(1)$.

(c) *Proof that* $c_{rs,n} = o_p(1)$: Let $\rho_* = \min\{\rho_S, \rho_K\}$, where ρ_S and ρ_K are as in Assumptions 4*(b) and 7. Then condition (3) for $K(.)$ also hold with ρ_K replaced by ρ_*. Furthermore observe that $|K(x) - 1| \leq (c_K + 1)|x|^{\rho_*}$ for all x. Given Assumptions 1, 2 and 4*(b) we have $\sum_{j=1}^n |\sigma_{ij,n}| \leq c_R^2 < \infty$ and

$$\begin{aligned}\sum_{j=1}^n |\sigma_{ij,n}|[d_{ij,1,n} + c_V]^{\rho_*} &\leq \sum_{j=1}^n |\sigma_{ij,n}|[d_{ij,1,n} + c_V + 1]^{\rho_S} \\ &\leq 2^{\rho_S}[c_V + 1]^{\rho_S} \sum_{j=1}^n |\sigma_{ij,n}| + 2^{\rho_S} \sum_{j=1}^n |\sigma_{ij,n}| d_{ij,1,n}^{\rho_S} \leq const < \infty.\end{aligned}$$

Hence

$$\begin{aligned}|c_{rs,n}| &\leq n^{-1} \sum_{i=1}^n \sum_{j=1}^n |h_{ir,n}\|h_{js,n}\|\sigma_{ij,n}| \left|K\left(\min_m\{d^*_{ij,m,n}/d_{m,n}\}\right) - 1\right| \\ &\leq c_H^2(c_K + 1)n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{ij,n}| \left[\min_m\{d^*_{ij,m,n}/d_{m,n}\}\right]^{\rho_*}\end{aligned}$$

$$\leqslant c_H^2(c_K+1)\,d_{1,n}^{-\rho_*}n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|\sigma_{ij,n}|[d_{ij,1,n}+c_V]^{\rho_*}$$

$$\leqslant const^* d_{1,n}^{-\rho_*}=\mathrm{o}(1)$$

since $d_{1,n}\to\infty$. This establishes that $c_{rs,n}=\mathrm{o_p}(1)$.

Having demonstrated that all terms on the r.h.s. of (A.3) are $\mathrm{o_p}(1)$ this concludes the proof. □

Derivation of optimal rate given in Remark 2: Observe that under the assumptions of the remark we have $E\ell_n^2(d_n)=\mathrm{o}(n^{2\tau})$ and hence $\ell_n=\mathrm{o_p}(n^{\tau})$, and $d_n^{-1}=\mathrm{O}(n^{-2\tau/\eta})$. From the proof of Theorem 1 it is then readily seen that

$$|\hat{\psi}_{rs,n}-\psi_{rs,n}|\leqslant|a_{rs,n}|+|b_{rs,n}|+|c_{rs,n}|$$

with $a_{rs,n}=\mathrm{O_p}(1)*[n^{-1/2+1/q}\ell_n^{1-1/q}]=\mathrm{o_p}(n^{-1/2+1/q+\tau(1-1/q)})$, $b_{rs,n}=\mathrm{O_p}(n^{-1/2}(E\ell_n^2)^{1/2})=\mathrm{o_p}(n^{\tau-1/2})$ and $c_{rs,n}=\mathrm{O_p}(d_n^{-\rho_*})=\mathrm{O_p}(n^{-2\rho_*\tau/\eta})$. Thus $\hat{\psi}_{rs,n}-\psi_{rs,n}=\mathrm{O_p}(\gamma_n)$ with γ_n as given in the remark. Clearly, γ_n is minimized for τ_* as given in the remark.

Proof of Theorem 3. From (1), (10), and (12) we have

$$n^{1/2}(\hat{\delta}_n-\delta_0)=M_n n^{-1/2}L_n'\varepsilon_n, \tag{A.4}$$

where $M_n=(n^{-1}\hat{Z}_n'Z_n)^{-1}n^{-1}Z_n'H_n(n^{-1}H_n'H_n)^{-1}$ and $L_n'=H_n'R_n$. Observing that $\hat{Z}_n'Z_n=\hat{Z}_n'\hat{Z}_n=Z_n'H_n(H_n'H_n)^{-1}H_n'Z_n$ and $R_nR_n'=\Sigma_n$ it follows from Assumption 11 that

$$M_n\xrightarrow{\mathrm{p}}(Q_{HZ}'Q_{HH}^{-1}Q_{HZ})^{-1}Q_{HZ}'Q_{HH}^{-1},$$

$$n^{-1}L_nL_n'\xrightarrow{\mathrm{p}}\Psi, \tag{A.5}$$

where both limiting matrices are finite and non-singular. Assumptions 2 and 11 imply that the elements of L_n are uniformly bounded in absolute value. Assumption 1 and the central limit theorem for triangular arrays given in Kelejian and Prucha (1998, p. 112) then imply that

$$n^{-1/2}L_n'\varepsilon_n\xrightarrow{\mathrm{d}}\mathrm{N}(0,\Psi). \tag{A.6}$$

Part (a) of Theorem 3 follows trivially from (A.4) to (A.6).

Consider now part (b) of Theorem 3. We established above that $n^{-1}\hat{Z}_n'\hat{Z}_n\xrightarrow{\mathrm{p}}Q_{HZ}'Q_{HH}^{-1}Q_{HZ}$. Next observe that

$$\hat{\Phi}_n=(n^{-1}\hat{Z}_n'\hat{Z}_n)^{-1}n^{-1}Z_n'H_n(n^{-1}H_n'H_n)^{-1}\hat{\Psi}_n(n^{-1}H_n'H_n)^{-1}n^{-1}H_n'Z_n(n^{-1}\hat{Z}_n'\hat{Z}_n)^{-1}.$$

Part (b) of Theorem 3 then follows in light of Assumption 11, provided we can establish the consistency of the SHAC estimator $\hat{\Psi}_n$. To show that this is indeed the case we verify the assumptions of Theorem 2. Assumptions 1, 2, 4*, 5* and 7 are assumed to hold. Assumption 3 is clearly implied by Assumption 11. Hence we only have to verify that the 2SLS residuals satisfy Assumption 6. Let $z_{i,n}$ denote the ith row of Z_n, then $\hat{u}_{i,n}-u_{i,n}=z_{i.,n}\Delta_n$ with $\Delta_n=\delta_0-\hat{\delta}_n$. Given part (a) of the theorem clearly $n^{1/2}\|\Delta_n\|=\mathrm{O_p}(1)$. A sufficient condition for $z_{i,n}$ to satisfy the conditions of Assumption 6 is that all elements have uniformly bounded third absolute moments; see, e.g., Lemma A.2 in Kelejian and Prucha (1998). Since the elements of X_n are uniformly bounded in absolute value

by Assumption 10 and the third absolute moments of the elements of Y_n are uniformlybounded by Assumption 12 it only remains to be shown that $E|\bar{y}_{i,n}|^3 \leqslant const < \infty$, where $\bar{y}_{i,n}$ denotes the ith element of the spatial lag $W_n y_n$. Observe that

$$W_n y_n = W_n(I_n - \lambda_0 W_n)^{-1} X_n \beta_0 + W_n(I_n - \lambda_0 W_n)^{-1} Y_n \gamma_0 + W_n(I_n - \lambda_0 W_n)^{-1} R_n \varepsilon_n.$$

Assumptions 2 and 9 imply that the row and column sums of $W_n(I_n - \lambda_0 W_n)^{-1}$ and $W_n(I_n - \lambda_0 W_n)^{-1} R_n$ are uniformly bounded in absolute value. Assumption 10 then implies that the elements of $W_n(I_n - \lambda_0 W_n)^{-1} X_n \beta_0$ are uniformly bounded in absolute value. It now follow immediately from Lemma A.2 in Kelejian and Prucha (2004) that the elements of $W_n y_n$ have third absolute moments which are uniformly bounded. This completes the proof of part (b) of Theorem 3. □

References

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, MA.

Andrews, D.W.K., 1991. Heteroscedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Andrews, D.W.K., Monahan, J.C., 1992. An improved heteroscedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60, 953–966.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Boston, MA.

Anselin, L., 2001a. Spatial econometrics. In: Baltagi, B. (Ed.), A Companion in Theoretical Econometrics. Basil Blackwell, New York, NY.

Anselin, L., 2001b. Rao's score test in spatial econometrics. Journal of Statistical Planning and Inference 97, 113–139.

Anselin, L., Florax, R.J.G.M., 1995. Small sample properties of tests for spatial dependence in regression models. In: Anselin, L., Florax, R.J.G.M. (Eds.), New Directions in Spatial Econometrics. Springer, New York, NY.

Anselin, L., Rey, S., 1991. Properties of tests for spatial dependence in linear regression models. Geographic Analysis 23, 110–131.

Audretsch, D., Feldmann, M., 1996. R&D spillovers and the geography of innovation and production. American Economic Review 86, 630–640.

Baltagi, B.H., Li, D., 2001a. Double length artificial regressions for testing spatial dependence. Econometric Reviews 20, 31–40.

Baltagi, B.H., Li, D., 2001b. LM tests for functional form and spatial error correlation. International Regional Science Review 24, 194–225.

Baltagi, B.H., Li, D., 2004. Prediction in panel data models with spatial correlation. In: Anselin, L., Florax, R.J.G.M., Rey, S.J. (Eds.), Advances in Spatial Econometrics. Springer, New York, NY.

Baltagi, B.H., Song, S.H., Koh, W., 2003. Testing panel data regression models with spatial error correlation. Journal of Econometrics 117, 123–150.

Bell, K.P., Bockstael, N.E., 2000. Applying the generalized moments estimation approach to spatial problems involving micro level data. Review of Economics and Statistics 82, 72–82.

Bernat Jr., G., 1996. Does manufacturing matter? A spatial econometric view of Kaldor's laws. Journal of Regional Science 36, 463–477.

Besley, T., Case, A., 1995. Incumbent behavior: vote-seeking, tax-setting, and yardstick competition. American Economic Review 85, 25–45.

Bochner, S., 1933. Monotone funktionen, Stieltjessche integrale und harmonische analyse. Mathematische Annalen 108, 378–410.

Bollinger, C., Ihlanfeldt, K., 1997. The impact of rapid rail transit on economic development: the case of Atlanta's Marta. Journal of Urban Economics 42, 179–204.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. Springer, New York, NY.

Buettner, T., 1999. The effect of unemployment, aggregate wages, and spatial contiguity on local wages: an investigation with German district level data. Papers in Regional Science 78, 47–67.

Case, A., 1991. Spatial patterns in household demand. Econometrica 59, 953–966.
Case, A., Hines Jr., J., Rosen, H., 1993. Budget spillovers and fiscal policy independence: evidence from the states. Journal of Public Economics 52, 285–307.
Cliff, A., Ord, J., 1973. Spatial Autocorrelation. Pion, London.
Cliff, A., Ord, J., 1981. Spatial Process: Models and Applications. Pion, London.
Conley, T., 1996. Econometric modelling of cross sectional dependence. Ph.D. Thesis, University of Chicago, Chicago.
Conley, T., 1999. GMM estimation with cross sectional dependence. Journal of Econometrics 92, 1–45.
Conley, T., Molinari, F., 2005. Spatial correlation robust inference with errors in location or distances. Journal of Econometrics, this issue, in press, doi:10.1016/j.jeconom.2006.09.003.
Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley, New York, NY.
de Jong, R.M., Davidson, J., 2000. Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. Econometrica 68, 407–423.
Dowd, M.R., LeSage, J.P., 1997. Analysis of spatial contiguity influences on state price level formation. International Journal of Forecasting 13, 245–253.
Driscoll, J., Kraay, A., 1998. Consistent covariance matrix estimation with spatially dependent panel data. The Review of Economics and Statistics 80, 549–560.
Fuentes, M., 2002a. Interpolation of nonstationary air pollution process: a spatial spectral approach. Statistical Modelling 2, 281–298.
Fuentes, M., 2002b. Spectral methods for nonstationary spatial processes. Biometrika 89, 197–210.
Gallant, R., White, H., 1988. A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Basil Blackwell, New York, NY.
Gneiting, T., 2002. Compactly supported correlation functions. Journal of Multivariate Analysis 83, 493–508.
Golubov, B.I., 1981. On Abel-Poisson type and Riesz means. Annals of Mathematics 7, 161–194.
Grenander, U., Rosenblatt, M., 1957. Statistical Analysis of Stationary Time Series. Wiley, New York, NY.
Holtz-Eakin, D., 1994. Public sector capital and the productivity puzzle. Review of Economics and Statistics 76, 12–21.
Kelejian, H.H., Prucha, I.R., 1997. Estimation of spatial regression models with autoregressive errors by two-stage least squares procedures: a serious problem. International Regional Science Review 20, 103–111.
Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17, 99–121.
Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533.
Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics 104, 219–257.
Kelejian, H.H., Prucha, I.R., 2002. 2SLS and OLS in a spatial autoregressive model with equal spatial weights. Regional Science and Urban Economics 32, 691–707.
Kelejian, H.H., Prucha, I.R., 2004. Estimation of systems of spatially interrelated cross sectional equations. Journal of Econometrics 118, 27–50.
Kelejian, H.H, Prucha, I.R., Yuzefovich, Y., 2004. Instrumental variable estimation of a spatial autorgressive model with autoregressive disturbances: large and small sample results. In: Pace, K., LeSage, J. (Eds.), Advances in Econometrics, vol. 18, Spatial Spatiotemporal Econometrics. Elsevier, London.
Kelejian, H., Robinson, D., 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. Papers in Regional Science 72, 297–312.
Kelejian, H., Robinson, D., 1995. Spatial correlation: a suggested alternative to the autoregressive model. In: Anselin, L., Florax, R.J.G.M. (Eds.), New Direction in Spatial Econometrics. Springer, New York, NY.
Kelejian, H.H., Robinson, D., 1997. Infrastructure productivity estimation and its underlying econometric specifications: a sensitivity analysis. Papers in Regional Science 76, 115–131.
Kelejian, H., Robinson, D., 2000. Returns to investment in navigation infrastructure: an equilibrium approach. Annals of Regional Science 34, 83–108.
Lee, L.F., 2001a. Generalized Method of Moments Estimation of Spatial Autoregressive Processes. Ohio State University, Ohio.
Lee, L.F., 2001b. GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models. Ohio State University, Ohio.

Lee, L.F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregessive models. Econometric Theory 18, 252–277.
Lee, L.F., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22, 307–335.
Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1926.
LeSage, J.P., 1997. Bayesian estimation of spatial autoregressive models. International Regional Science Review 20, 113–129.
LeSage, J.P., 2000. Bayesian estimation of limited dependent variable spatial autoregressive models. Geographic Analysis 32, 19–35.
Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedastic and autocorrelated consistent covariance matrix. Econometrica 55, 703–708.
Pace, R., Barry, R., 1997. Sparse spatial autoregressions. Statistics and Probability Letters 33, 291–297.
Pinkse, J., Slade, M., 1998. Contracting in space: an application of spatial statistics to discrete-choice models. Journal of Econometrics 85, 125–154.
Pinkse, J., Slade, M.E., Brett, C., 2002. Spatial price competition: a semiparametric approach. Econometrica 70, 1111–1153.
Pötscher, B., Prucha, I.R., 1986. A class of partially adaptive one-step m-estimators for the nonlinear regression model with dependent observations. Journal of Econometrics 32, 219–251.
Pötscher, B.M., Prucha, I.R., 1997. Dynamic Nonlinear Econometric Models, Asymptotic Theory. Springer, New York, NY.
Priestley, M.B., 1964. The analysis of two-dimensional stationary processes with discontinuous spectra. Biometrika 51, 195–217.
Priestley, M.B., 1981. Spectral Analysis and Time Series: vols. I and II. Academic Press, New York, NY.
Pulvino, T., 1998. Do asset fire sales exit? An empirical investigation of commercial aircraft transactions. The Journal of Finance, 939–978.
Rey, S.J., Boarnet, M.G., 2004. A taxonomy of spatial econometric models for simultaneous equation systems. In: Anselin, L., Florax, R.J.G.M., Rey, S.J. (Eds.), Advances in Spatial Econometrics: Methodology, Tools and Applications. Springer, New York, NY.
Schoenberg, I.J., 1938. Metric spaces and completely monotone functions. Annals of Mathematics 39, 811–841.
Shroder, M., 1995. Games the states don't play: welfare benefits and the theory of fiscal federalism. Review of Economics and Statistics 77, 183–191.
Vigil, R., 1998. Interactions among municipalities in the provision of police services: a spatial econometric approach, Ph.D. Thesis, University of Maryland, Maryland.
Whittle, P., 1954. On statinory processes in the plane. Biometrika 41, 434–449.
Yaglom, A.M., 1962. An Introduction to the Theory of Stationary Random Functions. Prentice-Hall, Englewood Cliffs, NJ.