Economics Letters 117 (2012) 60–65

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

HAC estimation in spatial panels

Francesco Moscone *, Elisa Tosetti

Brunel University, United Kingdom

ARTICLE INFO

Article history:
Received 11 October 2011
Received in revised form
31 March 2012
Accepted 6 April 2012
Available online 24 April 2012

JEL classification:
C10
C31
C33

Keywords:
Panels
HAC estimator
Spatial dependence
Serial dependence

ABSTRACT

We propose a HAC estimator for the covariance matrix of the fixed effects estimator in a panel data model with unobserved fixed effects and errors that are both serially and spatially correlated.

1. Introduction

Recently, a number of works have focused on robust estimation of the slope parameters of a regression model where errors are spatially correlated. Variants of the Newey and West (1987) spectral density estimator in time series have been suggested by Conley (1999) and Driscoll and Kraay (1998) in the context of GMM estimators of spatial panels where T is large relative to N (see also Pinkse et al., 2002). More recently, Kelejian and Prucha (2007) have proposed a spatial version of the non-parametric heteroskedasticity–autocorrelation consistent (HAC) estimator introduced by White (1980) for a single cross section regression with spatially correlated errors. This approach permits to approximate the true covariance matrix with a weighted average of cross products of regression errors, where each element is weighted by a function of (possible multiple) distance between cross section units. Rather than using a measure of distance between units, Bester et al. (2011) have recently suggested to split the sample into groups so that group-level averages are approximately independent, and then use the HAC estimator based on a discrete group-membership metric.

In this paper, following the work by Kelejian and Prucha (2007), we suggest a HAC covariance matrix estimator in the context of a panel data model with unobserved fixed effects, where errors are allowed to be both spatially and serially correlated. Such estimator is useful in applied work, when dealing with large data sets, and little is known about the spatio-temporal process generating the error term. We show that the suggested HAC estimator is consistent for N going to infinity, with T fixed or T going to infinity. A small Monte Carlo exercise reported in the paper shows that this approach is quite robust to various forms of serial and cross sectional dependence.

2. The framework

Consider the panel data model

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + e_{it}, \quad i = 1, 2, \ldots, N; t = 1, 2, \ldots, T, \qquad (1)$$

where α_i are fixed parameters, $\mathbf{x}_{it}$ are strictly exogenous regressors, and e_{it} follows the general spatial process:

$$e_{it} = r_{i1}\varepsilon_{1t} + r_{i2}\varepsilon_{2t} + \cdots + r_{iN}\varepsilon_{Nt}, \qquad (2)$$

where r_{ij} are (unknown) elements, possibly function of a smaller set of coefficients, of an $N \times N$ non-stochastic matrix, $\mathbf{R} = (\mathbf{r}_{1.}, \mathbf{r}_{2.}, \ldots, \mathbf{r}_{N.})'$, with $\mathbf{r}_{i.} = (r_{i1}, r_{i2}, \ldots, r_{iN})'$, and ε_{it}, for each i, follows the general linear process:

$$\varepsilon_{it} = \sum_{a=0}^{\infty} c_{ia}\epsilon_{i,t-a}. \qquad (3)$$

Following Kelejian and Prucha (2007), we also assume that there is a meaningful distance measure, d_{ij}, between units i and j, with

* Correspondence to: Brunel Business School, Room: EJ057, Uxbridge, Middlesex UB8 3PH, United Kingdom. Tel.: +44 0 1895266833; fax: +44 0 1142685766.
E-mail addresses: moscone.francesco@brunel.ac.uk, francesco.moscone@brunel.ac.uk (F. Moscone).

0165-1765/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2012.04.006

$d_{ij} = d_{ji} \geq 0$, and the researcher can select a threshold distance, d_N, such that $d_N \to \infty$ as $N \to \infty$, and

$$\max_{1\leq i\leq N} \sum_{j=1}^{N} 1_{d_{ij}\leq d_N} \leq s_N, \tag{4}$$

i.e., s_N is the number of units for which $d_{ij} \leq d_N$. We make the following assumptions on the error term, regressors, and s_N.

Assumption 1. $\epsilon_{it} \sim IID(0, 1)$ with $E(\epsilon_{it}^4) < \infty$; $\max_{1\leq i\leq N} \sum_{a=0}^{\infty} |c_{ia}| < \infty$.

Assumption 2. $\max_{1\leq j\leq N} \sum_{i=1}^{N} |r_{ij}| < \infty$; $\max_{1\leq i\leq N} \sum_{j=1}^{N} |r_{ij}| < \infty$, $\sum_{j=1}^{N} |\mathbf{r}_{i.}'\mathbf{r}_{j.}| {d_{ij}}^{\rho} < \infty$.

Assumption 3. $s_N = O(N^{\alpha})$, with $0 \leq \alpha < 0.5$.

Assumption 4. $\mathbf{x}_{it}$ and ε_{is} are independently distributed for all i, t, s. $\mathbf{x}_{it}$ has finite elements, and $\lim_{(N,T)\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \widetilde{\mathbf{X}}_{i.}'\widetilde{\mathbf{X}}_{i.} = \mathbf{Q}$ is finite and non-singular, with $\widetilde{\mathbf{X}}_{i.} = \mathbf{M}\mathbf{X}_{i.}$, $\mathbf{X}_{i.} = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT})'$, $\mathbf{M} = \mathbf{I}_T - \mathbf{1}_T \left(\mathbf{1}_T'\mathbf{1}_T\right)^{-1} \mathbf{1}_T'$.

Assumption 5. $K(x) : \mathbb{R} \to [0, 1]$ is a kernel function satisfying $K(x) = K(-x)$, $K(0) = 1$, $K(x) = 0$ for $x > 1$, and $\frac{|K(x)-1|}{|x|^{\rho}} \leq C < \infty$, for $|x| \leq 1$ and with $\rho \geq 1$.

Under specification (2) and (3), errors are both cross sectionally and serially correlated, and $0 \leq |E(e_{it}e_{js})| = |\sum_{h=1}^{N} r_{ih}r_{jh} \sum_{a=0}^{\infty} c_{ia}c_{j,a+|s-t|}| < \infty$, for all i, j, t, s. A large variety of spatio-temporal models can be cast in this model, for example, the SAR or SMA processes having AR or MA errors. We observe that the clustered covariance matrix estimator advanced by Arellano (1987) is inconsistent under this specification, given that it ignores the cross section dependence present in the data. Under Assumptions 1 and 2, the covariance matrix of $\mathbf{e}_{i.} = (e_{i1}, e_{i2}, \ldots, e_{iT})'$, for each i, and that of $\mathbf{e}_{.t} = (e_{1t}, e_{2t}, \ldots, e_{Nt})'$, for each t, have absolute summable elements, i.e., $\sum_{s=1}^{T} |E(e_{it}e_{is})| \leq \sum_{h=1}^{N} r_{ih}^2 \sum_{s=1}^{T} \sum_{a=0}^{\infty} |c_{ia}| \left|c_{i,a+|s-t|}\right| < \infty$, and $\sum_{j=1}^{N} \left|E\left(e_{it}e_{jt}\right)\right| \leq \sum_{j=1}^{N} \sum_{h=1}^{N} |r_{ih}| \left|r_{jh}\right| \sum_{a=0}^{\infty} |c_{ia}| \left|c_{ja}\right| < \infty$. Finally, Assumption 5 is satisfied for many of the commonly used kernels (see Pötscher and Prucha, 1997, p. 129).

3. Robust estimation

The FE estimator of $\boldsymbol{\beta}$ in Eq. (1) is:

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^{N} \widetilde{\mathbf{X}}_{i.}'\widetilde{\mathbf{X}}_{i.}\right)^{-1} \sum_{i=1}^{N} \widetilde{\mathbf{X}}_{i.}'\widetilde{\mathbf{y}}_{i.}, \tag{5}$$

with $\widetilde{\mathbf{y}}_{i.} = \mathbf{M}\mathbf{y}_{i.}$. Under Assumptions 1–4, it is easily seen that

$$Asy.Cov\left(\hat{\boldsymbol{\beta}}_{FE}\right) = \frac{1}{NT}\mathbf{Q}^{-1}\boldsymbol{\Psi}\mathbf{Q}^{-1}, \tag{6}$$

with

$$\boldsymbol{\Psi} = \lim_{(N,T)\to\infty} \frac{1}{NT} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{js}'\mathbf{r}_{i.}'\boldsymbol{\gamma}_{ts}\mathbf{r}_{j.}, \tag{7}$$

$\boldsymbol{\gamma}_{ts} = \boldsymbol{\gamma}_{st} = \mathrm{diag}\{\gamma_1(|t-s|), \gamma_2(|t-s|), \ldots, \gamma_N(|t-s|)\}$, $E(\varepsilon_{ht}\varepsilon_{hs}) = \gamma_h(|t-s|)$ (see Lemma A.1). We suggest the following HAC estimator for (6):

$$\widehat{Asy.Cov}\left(\hat{\boldsymbol{\beta}}_{FE}\right) = \frac{1}{NT}\mathbf{Q}_{NT}^{-1}\hat{\boldsymbol{\Psi}}\mathbf{Q}_{NT}^{-1}, \tag{8}$$

where $\mathbf{Q}_{NT} = \frac{1}{NT} \sum_{i=1}^{N} \widetilde{\mathbf{X}}_{i.}'\widetilde{\mathbf{X}}_{i.}$,

$$\hat{\boldsymbol{\Psi}} = \frac{1}{NT} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{js}'\hat{e}_{it}\hat{e}_{js}K(d_{ij}/d_N), \tag{9}$$

$\hat{e}_{it} = \widetilde{y}_{it} - \hat{\boldsymbol{\beta}}_{FE}'\widetilde{\mathbf{x}}_{it}$. Note that for $T = 1$ expression (8) reduces to the Kelejian and Prucha (2007) HAC estimator, while in absence of spatial correlation (i.e., setting $K(d_{ij}/d_N) = 1$ when $i = j$ and zero otherwise) it reduces to the Arellano (1987) clustered estimator. The following theorem establishes the asymptotic normality of $\hat{\boldsymbol{\beta}}_{FE}$ and the consistency of (8) (see the Appendix for a proof).

Theorem 1. *Suppose Assumptions 1–5 are satisfied. Then as $N \to \infty$, for fixed T or $T \to \infty$*

$$\sqrt{NT}\left(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}\right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}). \tag{10}$$

Further, let $\hat{\boldsymbol{\Sigma}} = NT \cdot \widehat{Asy.Cov}\left(\hat{\boldsymbol{\beta}}_{FE}\right)$, where $\widehat{Asy.Cov}\left(\hat{\boldsymbol{\beta}}_{FE}\right)$ is given by (8), then

$$\hat{\boldsymbol{\Sigma}} \xrightarrow{p} \boldsymbol{\Sigma}. \tag{11}$$

4. Monte Carlo experiments

The data generating process is:

$$y_{it} = \alpha_i + \beta x_{it} + e_{it}, \quad \text{with } x_{it} = \alpha_i + \upsilon_{it}, \tag{12}$$

where $\beta = 1$, $\alpha_i \sim IIDN(1, 1)$ do not change across replications, and

$$e_{it} = \delta_i \sum_{j=1}^{N} s_{ij}e_{jt} + \varepsilon_{it},$$
$$\varepsilon_{it} = \rho_i\varepsilon_{i,t-1} + \left(1-\rho_i^2\right)^{1/2}\epsilon_{it}, \quad \epsilon_{it} \sim IIDN(0, 1), \tag{13}$$

and s_{ij} are elements of a $N \times N$ spatial weights matrix, $\mathbf{S}$. The data generating process for the regressor error, υ_{it}, does not change across experiments and is given by:

$$\upsilon_{it} = 0.5 \sum_{j=1}^{N} s_{ij}\upsilon_{jt} + \xi_{it},$$
$$\xi_{it} = 0.5\xi_{i,t-1} + \left(1-0.5^2\right)^{1/2}\varkappa_{it}, \quad \varkappa_{it} \sim IIDN(0, 1). \tag{14}$$

We follow Kelejian and Prucha (2007) and assume that units are located on a grid at locations (r, s), for $r, s = 1, \ldots, \sqrt{N}$, and $\mathbf{S}$ is taken to be a row-normalized, rook-type matrix where two units are neighbors if their Euclidean distance, d_{ij}, is less than or equal to one. We try $\delta_i = 0$, $\delta_i \sim U(0.2, 0.4)$, $\delta_i \sim U(0.5, 0.7)$, in all its combinations with $\rho_i = 0$, $\rho_i \sim U(0.2, 0.4)$, and $\rho_i \sim U(0.5, 0.7)$. We also try with $\delta_i \sim U(-0.4, -0.2)$, $\delta_i \sim U(-0.7, -0.5)$, in all its combinations with $\rho_i = 0$, $\rho_i \sim U(-0.4, -0.2)$, and $\rho_i \sim U(-0.7, -0.5)$. The number of replications is set to 2,000, and experiments are carried for $N = 400, 625, 900$ and $T = 5, 50$. We adopted the Parzen kernel function.

Table 1 reports the relative bias, computed as the bias of the proposed HAC estimator divided by the bias of the Arellano (1987) clustered estimator, the relative RMSE, computed as the ratio of the RMSEs, as well as size and power of the FE estimator[1] both adopting clustered standard errors, and the proposed HAC standard errors, for various combinations of δ_i and ρ_i. The nominal size is set to 5%, while power of the FE estimator is computed under the hypothesis that $\beta = 0.90$. Results show that, as expected, when $\delta_i = 0$ test statistics using the clustered standard errors have the correct size. Under this case, the bias and RMSE of the two estimators are very small, causing the relative bias and RMSE to be volatile. However, when $\delta_i \neq 0$, the bias and RMSE of the proposed HAC estimator are always smaller than those of the clustered estimator, making the relative bias and RMSE smaller

1 Bias and RMSE of the FE estimator are available upon request.

Table 1
Small sample properties for the Clustered and HAC estimator.

N \ T	Relative bias		Relative RMSE		Clustered estimator				HAC estimator			
	5	50	5	50	Size		Power		Size		Power	
					5	50	5	50	5	50	5	50
	$\delta_i = 0$, $\rho_i \sim U(-0.4, -0.2)$											
400	7.063	4.726	1.972	2.250	0.054	0.049	0.730	1.00	0.053	0.048	0.753	1.00
625	6.029	−4.132	2.234	2.474	0.052	0.055	0.831	1.00	0.050	0.052	0.842	1.00
900	−1.169	3.941	2.436	2.739	0.046	0.055	0.975	1.00	0.046	0.051	0.985	1.00
	$\delta_i \sim U(-0.4, -0.2)$, $\rho_i \sim U(-0.4, -0.2)$											
400	−0.014	0.020	0.806	0.808	0.038	0.021	0.890	1.00	0.051	0.043	0.892	1.00
625	0.043	0.044	0.783	0.745	0.029	0.031	0.965	1.00	0.049	0.053	0.951	1.00
900	0.052	−0.003	0.729	0.701	0.025	0.038	1.00	1.00	0.047	0.054	1.00	1.00
	$\delta_i \sim U(-0.7, -0.5)$, $\rho_i \sim U(-0.4, -0.2)$											
400	0.101	0.117	0.392	0.379	0.022	0.013	0.910	1.00	0.046	0.045	0.921	1.00
625	0.108	0.101	0.368	0.347	0.013	0.014	0.987	1.00	0.052	0.046	0.976	1.00
900	0.095	0.072	0.341	0.319	0.014	0.023	1.00	1.00	0.048	0.049	1.00	1.00
	$\delta_i \sim U(-0.7, -0.5)$, $\rho_i \sim U(-0.7, -0.5)$											
400	0.100	0.121	0.396	0.389	0.017	0.018	0.979	1.00	0.043	0.046	0.987	1.00
625	0.111	0.094	0.375	0.342	0.018	0.012	0.990	1.00	0.046	0.048	0.999	1.00
900	0.090	0.075	0.335	0.313	0.019	0.015	1.00	1.00	0.045	0.050	1.00	1.00
	$\delta_i = 0$, $\rho_i \sim U(0.2, 0.4)$											
400	5.375	3.945	1.862	2.197	0.053	0.051	0.751	1.00	0.052	0.053	0.771	1.00
625	−5.162	−5.412	2.039	2.498	0.052	0.048	0.852	1.00	0.053	0.050	0.860	1.00
900	−1.527	4.055	2.318	2.787	0.049	0.051	0.979	1.00	0.050	0.052	0.980	1.00
	$\delta_i \sim U(0.2, 0.4)$, $\rho_i \sim U(0.2, 0.4)$											
400	0.480	0.408	0.963	0.907	0.075	0.068	0.880	1.00	0.054	0.050	0.930	1.00
625	0.347	0.333	0.859	0.860	0.072	0.070	0.985	1.00	0.052	0.055	0.990	1.00
900	0.280	0.345	0.848	0.808	0.070	0.074	1.00	1.00	0.051	0.051	1.00	1.00
	$\delta_i \sim U(0.5, 0.7)$, $\rho_i \sim U(0.2, 0.4)$											
400	0.389	0.353	0.598	0.540	0.122	0.106	0.950	1.00	0.070	0.051	0.944	1.00
625	0.296	0.287	0.500	0.478	0.112	0.096	1.00	1.00	0.058	0.053	1.00	1.00
900	0.246	0.277	0.471	0.442	0.108	0.089	1.00	1.00	0.055	0.053	1.00	1.00
	$\delta_i \sim U(0.5, 0.7)$, $\rho_i \sim U(0.5, 0.7)$											
400	0.382	0.345	0.609	0.541	0.149	0.103	0.985	1.00	0.064	0.049	0.990	1.00
625	0.292	0.287	0.510	0.478	0.121	0.011	1.00	1.00	0.056	0.053	1.00	1.00
900	0.245	0.271	0.481	0.443	0.109	0.010	1.00	1.00	0.051	0.050	1.00	1.00

than 1, with a decreasing pattern as δ_i in absolute value of gets large. Further, test statistics based on the clustered standard errors are undersized for values of $\delta_i < 0$, and oversized for values of $\delta_i > 0$. On the contrary, test statistics based on the proposed HAC estimator seem to be quite robust to various patterns of serial and cross sectional dependence, also when these are sizable.

Acknowledgment

Elisa Tosetti and Francesco Moscone acknowledge financial support from the Economic and Social Research Council (grant RES-061-25-0317).

Appendix

Lemma A.1. *Consider e_{it} in (2) and (3). Then under Assumptions 1 and 2, for $i, j = 1, 2, \ldots, N$, $t, s = 1, 2, \ldots, T$*

$$E\left(e_{it}e_{js}\right) = \mathbf{r}_{i.}'\boldsymbol{\gamma}_{ts}\mathbf{r}_{j.}, \tag{A.1}$$

and for $i, j, u, v = 1, 2, \ldots, N$, $t, s, t', s' = 1, 2, \ldots, T$

$$Cov\left(e_{it}e_{js}, e_{ut'}e_{vs'}\right) = \mathbf{r}_{i.}'\boldsymbol{\gamma}_{tt'}\mathbf{r}_{u.}\mathbf{r}_{j.}'\boldsymbol{\gamma}_{ss'}\mathbf{r}_{v.} + \mathbf{r}_{i.}'\boldsymbol{\gamma}_{ts'}\mathbf{r}_{v.}\mathbf{r}_{j.}'\boldsymbol{\gamma}_{st'}\mathbf{r}_{u.} + \sum_{\ell=1}^{N} r_{i\ell}r_{j\ell}r_{u\ell}r_{v\ell}\omega_{\ell,tst's'}. \tag{A.2}$$

where $\omega_{\ell,tst's'} = [\mu_{\ell,tst's'} - \gamma_\ell(|t-s|)\gamma_\ell(|t'-s'|) - \gamma_\ell(|t-t'|)\gamma_\ell(|s-s'|) - \gamma_\ell(|t-s'|)\gamma_\ell(|s-t'|)]$, and $\mu_{\ell,tst's'} = E(\varepsilon_{\ell t}\varepsilon_{\ell s}\varepsilon_{\ell t'}\varepsilon_{\ell s'})$.

Proof. Noting that, under Assumption 1, $E\left(\epsilon_{it}^2\right) = 1$, we have

$$E\left(e_{it}e_{js}\right) = \sum_{h=1}^{N} r_{ih}r_{jh}E\left(\varepsilon_{ht}\varepsilon_{hs}\right) = \sum_{h=1}^{N} r_{ih}r_{jh}\sum_{a=0}^{\infty}|c_{ha}|\left|c_{h,a+|t-s|}\right|$$
$$= \sum_{h=1}^{N} r_{ih}r_{jh}\gamma_h\left(|s-t|\right) = \mathbf{r}_{i.}'\boldsymbol{\gamma}_{ts}\mathbf{r}_{j.},$$

which proves (A.1). As for (A.2), we have (see also Ullah, 2004)

$$
\begin{aligned}
&Cov\left(e_{it}e_{js}, e_{ut'}e_{vs'}\right)\\
&= \sum_{k,h,p,q=1}^{N} r_{ik}r_{jh}r_{up}r_{vq}E\left(\varepsilon_{kt}\varepsilon_{hs}\varepsilon_{pt'}\varepsilon_{qs'}\right) - \mathbf{r}'_{i.}\boldsymbol{\gamma}_{ts}\mathbf{r}_{j.}\mathbf{r}'_{u.}\boldsymbol{\gamma}_{t's'}\mathbf{r}_{v.}\\
&= \sum_{k\neq p=1}^{N} r_{ik}r_{jk}E\left(\varepsilon_{kt}\varepsilon_{ks}\right)\sum_{p=1}^{N} r_{up}r_{vp}E\left(\varepsilon_{pt'}\varepsilon_{ps'}\right)\\
&\quad+\sum_{k\neq p=1}^{N} r_{ik}r_{uk}E\left(\varepsilon_{kt}\varepsilon_{kt'}\right)\sum_{p=1}^{N} r_{jp}r_{vp}E\left(\varepsilon_{ps}\varepsilon_{ps'}\right)\\
&\quad+\sum_{k\neq p=1}^{N} r_{ik}r_{vk}E\left(\varepsilon_{kt}\varepsilon_{ks'}\right)\sum_{p=1}^{N} r_{jp}r_{up}E\left(\varepsilon_{ps}\varepsilon_{pt'}\right)\\
&\quad+\sum_{\ell=1}^{N} r_{i\ell}r_{j\ell}r_{u\ell}r_{v\ell}E\left(\varepsilon_{\ell t}\varepsilon_{\ell s}\varepsilon_{\ell t'}\varepsilon_{\ell s'}\right)\\
&- \mathbf{r}'_{i.}\boldsymbol{\gamma}_{ts}\mathbf{r}_{j.}\mathbf{r}'_{u.}\boldsymbol{\gamma}_{t's'}\mathbf{r}_{v.}\\
&= \mathbf{r}'_{i.}\boldsymbol{\gamma}_{tt'}\mathbf{r}_{u.}\mathbf{r}'_{j.}\boldsymbol{\gamma}_{ss'}\mathbf{r}_{v.} + \mathbf{r}'_{i.}\boldsymbol{\gamma}_{ts'}\mathbf{r}_{v.}\mathbf{r}'_{j.}\boldsymbol{\gamma}_{st'}\mathbf{r}_{u.}\\
&\quad+\sum_{\ell=1}^{N} r_{i\ell}r_{j\ell}r_{u\ell}r_{v\ell}\omega_{\ell,tst's'}. \quad \square
\end{aligned}
$$

Lemma A.2. *Consider e_{it} in (2) and (3). Then, under Assumptions 1 and 2,*

$$
\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} e_{it}e_{js} = O_p\left(\frac{1}{NT}\right). \tag{A.3}
$$

Proof. Note that $\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} e_{it}e_{js}$ has mean

$$
\begin{aligned}
E\left(\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} e_{it}e_{js}\right) &= \frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\mathbf{r}'_{i.}\boldsymbol{\gamma}_{ts}\mathbf{r}_{j.}\\
&= O\left(\frac{1}{NT}\right),
\end{aligned}
$$

since, under Assumptions 1 and 2, $\sum_{s=1}^{T}\gamma_{h,ts} = O(1)$, and $\sum_{j=1}^{N}\mathbf{r}'_{i.}\mathbf{r}_{j.} = \sum_{j=1}^{N}\sum_{h=1}^{N} r_{ih}r_{jh} = O(1)$. Further, using (A.2) its variance satisfies

$$
\begin{aligned}
&Var\left(\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} e_{it}e_{js}\right)\\
&= \frac{1}{(NT)^4}\sum_{i,j,u,v=1}^{N}\sum_{t,s,t',s'=1}^{T} Cov\left(e_{it}e_{js}, e_{ut'}e_{vs'}\right)\\
&= \frac{1}{(NT)^4}\sum_{i,j,u,v=1}^{N}\sum_{t,s,t',s'=1}^{T}\left[\mathbf{r}'_{i.}\boldsymbol{\gamma}_{tt'}\mathbf{r}_{u.}\mathbf{r}'_{j.}\boldsymbol{\gamma}_{ss'}\mathbf{r}_{v.}\right.\\
&\quad\left.+\mathbf{r}'_{i.}\boldsymbol{\gamma}_{ts'}\mathbf{r}_{v.}\mathbf{r}'_{j.}\boldsymbol{\gamma}_{st'}\mathbf{r}_{u.}\right]\\
&\quad+\frac{1}{(NT)^4}\sum_{\ell=1}^{N}\sum_{i,j,u,v=1}^{N}\sum_{t,s,t',s'=1}^{T} r_{i\ell}r_{j\ell}r_{u\ell}r_{v\ell}\omega_{\ell,tst's'}\\
&= O\left(\frac{1}{N^2T^2}\right),
\end{aligned}
$$

given that, under Assumptions 1 and 2, $\frac{1}{T^2}\sum_{t,s,t',s'=1}^{T}\mu_{\ell,tst's'} = \frac{1}{T^2}\sum_{t,s,t',s'=1}^{T}E(\varepsilon_{\ell t}\varepsilon_{\ell s}\varepsilon_{\ell t'}\varepsilon_{\ell s'}) = O(1)$. $\square$

Proof of Theorem 1. Consider

$$
\begin{aligned}
&\sqrt{NT}\left(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}\right)\\
&= \left(\frac{1}{NT}\sum_{t=1}^{T}\widetilde{\mathbf{X}}'_{.t}\widetilde{\mathbf{X}}_{.t}\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\widetilde{\mathbf{X}}'_{.t}\mathbf{R}\boldsymbol{\varepsilon}_{.t},
\end{aligned}
\tag{A.4}
$$

where $\widetilde{\mathbf{X}}_{.t} = (\widetilde{\mathbf{x}}_{1t}, \ldots, \widetilde{\mathbf{x}}_{Nt})'$. Asymptotic normality of (A.4) when $N, T \to \infty$ can be proved by applying the Beveridge–Nelson decomposition to ε_{it} (see Phillips and Solo, 1992, for details):

$$
\varepsilon_{it} = c_i(1)\epsilon_{it} + \tilde{\epsilon}_{i,t-1} - \tilde{\epsilon}_{it},
$$

where $c_i(1) = \sum_{a=0}^{\infty} c_{ia}$, $\tilde{\epsilon}_{it} = \sum_{a=0}^{\infty}\tilde{c}_{ia}\epsilon_{i,t-a}$ and $\tilde{c}_{ia} = \sum_{k=a+1}^{\infty} c_{ik}$. Hence

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\widetilde{\mathbf{X}}'_{.t}\mathbf{R}\boldsymbol{\varepsilon}_{.t} &= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\widetilde{\mathbf{X}}'_{.t}\mathbf{R}\mathbf{c}(1)\boldsymbol{\epsilon}_{.t}\\
&\quad+\frac{1}{\sqrt{NT}}\widetilde{\mathbf{X}}'_{.t}\mathbf{R}\tilde{\epsilon}_{.1} - \frac{1}{\sqrt{NT}}\widetilde{\mathbf{X}}'_{.t}\mathbf{R}\tilde{\epsilon}_{.T},
\end{aligned}
$$

where $\mathbf{c}(1)$ is a diagonal matrix with diagonal elements $c_i(1) < \infty$, and $\frac{1}{\sqrt{NT}}\widetilde{\mathbf{X}}'_{.t}\mathbf{R}\tilde{\epsilon}_{.1}$, $\frac{1}{\sqrt{NT}}\mathbf{R}\tilde{\epsilon}_{.T}$ tend to zero as $T \to \infty$ under Assumptions 1 and 2. Hence, asymptotic normality follows by applying to $\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\widetilde{\mathbf{X}}'_{.t}\mathbf{R}\mathbf{c}(1)\boldsymbol{\epsilon}_{.t}$ the central limit theorem for triangular arrays provided in Kelejian and Prucha (1998, see p. 112), and noting that the matrices $\widetilde{\mathbf{X}}'_{.t}\mathbf{R}\mathbf{c}(1)$ and $\widetilde{\mathbf{X}}'_{.t}\mathbf{R}\mathbf{c}(1)\mathbf{c}(1)'\mathbf{R}'\widetilde{\mathbf{X}}_{.t}/N$ have finite elements. To prove consistency of $\hat{\boldsymbol{\Psi}}$, consider

$$
\begin{aligned}
\hat{e}_{it} = \tilde{y}_{it} - \hat{\boldsymbol{\beta}}'_{FE}\widetilde{\mathbf{x}}_{it} &= \tilde{e}_{it} + \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{FE}\right)'\widetilde{\mathbf{x}}_{it}\\
&= \tilde{e}_{it} + \frac{1}{NT}\sum_{k=1}^{N} z_{ki,t}\tilde{e}_{kt},
\end{aligned}
\tag{A.5}
$$

where $z_{ki,t} = \widetilde{\mathbf{x}}'_{kt}\mathbf{Q}_{NT}^{-1}\widetilde{\mathbf{x}}_{it} < K < \infty$. Replace the expression for $\hat{e}_{it}$ into (9), to obtain:

$$
\begin{aligned}
\hat{\boldsymbol{\Psi}} &= \frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}'_{js}e_{it}e_{js}K(d_{ij}/d_N)\\
&\quad+\frac{2}{N^2T^2}\sum_{i,j=1}^{N}\sum_{k=1}^{N}\sum_{t,s=1}^{T}\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}'_{js}e_{is}e_{kt}z_{kj,t}K(d_{ij}/d_N)\\
&\quad+\frac{1}{N^3T^3}\sum_{i,j=1}^{N}\sum_{k,h=1}^{N}\sum_{t,s=1}^{T}\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}'_{js}z_{ki,t}z_{hj,s}e_{kt}e_{hs}K(d_{ij}/d_N).
\end{aligned}
$$

Note that we have dropped $\sim$ from e_{it} given that $\widetilde{\mathbf{X}}'_{i.}\widetilde{\mathbf{e}}_{i.} = \widetilde{\mathbf{X}}'_{i.}\mathbf{e}_{i.}$. We now focus on the (g, m)th element of $\hat{\boldsymbol{\Psi}}$, given by

$$
\begin{aligned}
\hat{\Psi}_{gm} &= \frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\tilde{x}_{g,it}\tilde{x}_{m,js}e_{it}e_{js}K(d_{ij}/d_N)\\
&\quad+\frac{2}{N^2T^2}\sum_{i,j=1}^{N}\sum_{k=1}^{N}\sum_{t,s=1}^{T}\tilde{x}_{g,it}\tilde{x}_{m,js}z_{kj,t}e_{is}e_{kt}K(d_{ij}/d_N)\\
&\quad+\frac{1}{N^3T^3}\sum_{i,j=1}^{N}\sum_{k,h=1}^{N}\sum_{t,s=1}^{T}\tilde{x}_{g,it}\tilde{x}_{m,js}z_{ki,t}z_{hj,s}e_{kt}e_{hs}\\
&\quad\times K(d_{ij}/d_N).
\end{aligned}
\tag{A.6}
$$

Note that

$$\hat{\Psi}_{gm} - \Psi_{gm} = \frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\tilde{x}_{g,it}\tilde{x}_{m,js} \times \left(\hat{e}_{it}\hat{e}_{js} - e_{it}e_{js}\right)K(d_{ij}/d_N) \tag{A.7}$$

$$+ \frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\tilde{x}_{g,it}\tilde{x}_{m,js}\left(e_{it}e_{js} - \mathbf{r}'_{i.}\boldsymbol{\gamma}_{ts}\mathbf{r}_{j.}\right)K(d_{ij}/d_N) \tag{A.8}$$

$$+ \frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\tilde{x}_{g,it}\tilde{x}_{m,js}\mathbf{r}'_{i.}\boldsymbol{\gamma}_{ts}\mathbf{r}_{j.}\left[K(d_{ij}/d_N) - 1\right]$$

$$= A + B + C. \tag{A.9}$$

We now prove that A, B, C go to zero. First note that, since $K(d_{ij}/d_N) \leq 1$, we have, under Assumption 3,

$$\sum_{j=1}^{N}K(d_{ij}/d_N) \leq \sum_{j=1}^{N}1_{d_{ij}\leq d_N} \leq s_N = O(N^{\alpha}). \tag{A.10}$$

Also, using (A.6), term A satisfies:

$$A = \frac{2}{N^2T^2}\sum_{i,j=1}^{N}\sum_{k=1}^{N}\sum_{t,s=1}^{T}\tilde{x}_{g,it}\tilde{x}_{m,js}e_{kt}e_{is}z_{kj,t}K(d_{ij}/d_N) \tag{A.11}$$

$$+ \frac{1}{N^3T^3}\sum_{i,j=1}^{N}\sum_{k,h=1}^{N}\sum_{t,s=1}^{T}\tilde{x}_{g,it}\tilde{x}_{m,js}z_{ki,t}z_{hj,s}e_{kt}e_{hs} \times K(d_{ij}/d_N) = A1 + A2. \tag{A.12}$$

Hence $E(A1)$ satisfies

$$|E(A1)| \leq \frac{2}{N^2T^2}\sum_{i,j=1}^{N}\sum_{k=1}^{N}\sum_{t,s=1}^{T}\left|\tilde{x}_{g,it}\tilde{x}'_{m,js}\right| \times \left|\mathbf{r}'_{i.}\boldsymbol{\gamma}_{ts}\mathbf{r}_{k.}\right|\left|z_{kj,t}\right|K(d_{ij}/d_N)$$

$$\leq \frac{2C}{N^2T^2}\sum_{i,j=1}^{N}\sum_{k=1}^{N}\sum_{t,s=1}^{T}\left|\mathbf{r}'_{i.}\boldsymbol{\gamma}_{ts}\mathbf{r}_{k.}\right|K(d_{ij}/d_N)$$

$$\leq \frac{2C}{N^2T}\sum_{i,k=1}^{N}\left|\mathbf{r}'_{i.}\mathbf{r}_{k.}\right|\sum_{j=1}^{N}K(d_{ij}/d_N)$$

$$= \frac{2C}{N^{2-\alpha}T}\sum_{i,k=1}^{N}\left|\mathbf{r}'_{i.}\mathbf{r}_{k.}\right|$$

$$\leq \frac{2C}{N^{2-\alpha}T}\sum_{h=1}^{N}\sum_{i=1}^{N}|r_{ih}|\sum_{k=1}^{N}|r_{kh}| = O\left(\frac{1}{N^{1-\alpha}T}\right),$$

from which it follows that $E(A1) = O\left(\frac{1}{N^{1-\alpha}T}\right)$. Using Lemma A.1, the variance of $A1$ satisfies

$$Var(A1) \leq \frac{2}{N^4T^4}\sum_{i,j,i',j'=1}^{N}\sum_{k,k'=1}^{N}\sum_{t,s,t',s'=1}^{T} \times \left|Cov\left(e_{is}e_{kt}, e_{i's'}e_{k't'}\right)\right|\left|z_{kj,t}\right|\left|z_{k'j',t'}\right| \times K(d_{ij}/d_N)K(d_{i'j'}/d_N)$$

$$\leq \frac{2C}{N^4T^4}\sum_{i,i',k,k'=1}^{N}\sum_{t,s,t',s'=1}^{T} \times \left|\mathbf{r}'_{i.}\boldsymbol{\gamma}_{tt'}\mathbf{r}_{i'.}\mathbf{r}'_{k.}\boldsymbol{\gamma}_{ss'}\mathbf{r}_{k'.} + \mathbf{r}'_{i.}\boldsymbol{\gamma}_{ts'}\mathbf{r}_{k'.}\mathbf{r}'_{k.}\boldsymbol{\gamma}_{st'}\mathbf{r}_{i'.}\right| \times \sum_{j=1}^{N}K(d_{ij}/d_N)\sum_{j'=1}^{N}K(d_{i'j'}/d_N)$$

$$+ \frac{2C}{N^4T^4}\sum_{i,i',k,k'=1}^{N}\sum_{t,s,t',s'=1}^{T}\sum_{\ell=1}^{N}|r_{i\ell}|\,|r_{i'\ell}|\,|r_{k\ell}|\,|r_{k'\ell}| \times \left|\omega_{\ell,tst's'}\right|\sum_{j=1}^{N}K(d_{ij}/d_N)\sum_{j'=1}^{N}K(d_{i'j'}/d_N)$$

$$= O\left(\frac{1}{N^{(2-2\alpha)}T^2}\right),$$

which implies $A1 = O_p\left(\frac{1}{N^{1-\alpha}T}\right)$. Using similar lines of reasoning, it can be proved that $A2 = O_p\left(\frac{1}{N^{1-\alpha}T^2}\right)$. Focusing on B, we have $E(B) = 0$, and its variance satisfies

$$Var(B) \leq \frac{1}{N^2T^2}\sum_{i,j,u,v=1}^{N}\sum_{t,s,t',s'=1}^{T} \times \left|\mathbf{r}'_{i.}\boldsymbol{\gamma}_{tt'}\mathbf{r}_{u.}\mathbf{r}'_{j.}\boldsymbol{\gamma}_{ss'}\mathbf{r}_{v.} + \mathbf{r}'_{i.}\boldsymbol{\gamma}_{ts'}\mathbf{r}_{v.}\mathbf{r}'_{j.}\boldsymbol{\gamma}_{st'}\mathbf{r}_{u.}\right| \times K(d_{ij}/d_N)K(d_{uv}/d_N)$$

$$+ \frac{1}{N^2T^2}\sum_{i,j,u,v=1}^{N}\sum_{t,s,t',s'=1}^{T}\sum_{\ell=1}^{N}\left|r_{i\ell}r_{j\ell}r_{u\ell}r_{v\ell}\omega_{\ell,tst's'}\right| \times K(d_{ij}/d_N)K(d_{uv}/d_N) = O\left(\frac{1}{N^{1-2\alpha}}\right).$$

Finally, using condition $\sum_{j=1}^{N}\left|\mathbf{r}'_{i.}\mathbf{r}_{j.}\right|d_{ij}^{\rho} < \infty$ in Assumption 3, C satisfies

$$|C| \leq \frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left|\tilde{x}_{g,it}\tilde{x}_{m,js}\right|\left|\mathbf{r}'_{i.}\boldsymbol{\gamma}_{ts}\mathbf{r}_{j.}\right|\left|K(d_{ij}/d_N) - 1\right|$$

$$\leq \frac{C}{NT}\sum_{i,j=1}^{N}\left|\mathbf{r}'_{i.}\boldsymbol{\gamma}_{ts}\mathbf{r}_{j.}\right|\left|K(d_{ij}/d_N) - 1\right|$$

$$\leq \frac{C}{N}\sum_{i,j=1}^{N}\left|\mathbf{r}'_{i.}\mathbf{r}_{j.}\right|\left|K(d_{ij}/d_N) - 1\right|$$

$$\leq \frac{C}{Nd_N^{\rho}}\sum_{i,j=1}^{N}\left|\mathbf{r}'_{i.}\mathbf{r}_{j.}\right|d_{ij}^{\rho} = O\left(d_N^{-\rho}\right).$$

It follows that

$$\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi} = O_p\left(\frac{1}{N^{(1-\alpha)}T}\right) + O_p\left(\frac{1}{N^{0.5-\alpha}}\right) + O_p\left(\frac{1}{d_N^{\rho}}\right), \tag{A.13}$$

and $\hat{\boldsymbol{\Sigma}} \xrightarrow{p} \boldsymbol{\Sigma}$. □

References

Arellano, M., 1987. Practitioners' corner: computing robust standard errors for within-groups estimators. Oxford Bulletin of Economics and Statistics 49, 431–434.

Bester, C.A., Conley, T.G., Hansen, C.B., 2011. Inference with dependent data using cluster covariance estimators. Journal of Econometrics 165, 137–151.

Conley, T.G., 1999. GMM estimation with cross sectional dependence. Journal of Econometrics 92, 1–45.

Driscoll, J.C., Kraay, A.C., 1998. Consistent covariance matrix estimation with spatially dependent panel data. The Review of Economics and Statistics 80, 549–560.

Kelejian, H.H., Prucha, I., 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., 2007. HAC estimation in a spatial framework. Journal of Econometrics 140, 131–154.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Phillips, P., Solo, V., 1992. Asymptotics for linear processes. Annals of Statistics 20, 971–1001.
Pinkse, J., Slade, M., Brett, C., 2002. Spatial price competition: a semiparametric approach. Econometrica 70, 1111–1153.
Pötscher, B.M., Prucha, I.R., 1997. Dynamic Nonlinear Econometric Models: Asymptotic Theory. Springer-Verlag.
Ullah, A., 2004. Finite Sample Econometrics. Oxford University Press, Oxford.
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.