Economics Letters 133 (2015) 100–104

Contents lists available at ScienceDirect

Economics Letters

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

Robust estimation under error cross section dependence[☆]

F. Moscone [*], Elisa Tosetti

Brunel University, United Kingdom

HIGHLIGHTS

- Propose a variance estimator for fixed effects and mean group estimators in panels.
- Robust to various forms of serial and cross sectional dependence in errors.
- Useful in applied work for large N short T panels when little is known about the process generating the errors.
- Shown to be consistent for N going to infinity, with T fixed.

ARTICLE INFO

Article history:
Received 4 February 2015
Received in revised form
13 May 2015
Accepted 14 May 2015
Available online 22 May 2015

JEL classification:
C10
C31
C33

Keywords:
Panels
Weak cross section dependence
Fixed effects estimator
Mean group estimator

ABSTRACT

We propose a robust, partial sample estimator for the covariance matrix of the fixed effects and mean group estimators of the slope coefficients in a short T panel data model with group-specific effects and errors that are weakly cross sectionally dependent and serially correlated.

1. Introduction

Recently, a number of studies have focused on robust estimation of the slope parameters of a regression model where errors are cross sectionally dependent. 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 has been extended by Moscone and Tosetti (2012) in the context of a panel data model with unobserved fixed effects, where errors are allowed to be both spatially and serially correlated for N and/or T going to infinity. Rather than using an arbitrary measure of distance between units, Bester et al. (2009) 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. However, the validity of this approach relies on the capacity of the researcher to construct groups whose averages are approximately independent. Robust inference with clustered data has also been considered by Ibragimov and Müller (2009), Cameron and Miller (2011), Cameron et al. (2011) and MacKinnon and Webb (2014). Robinson (2007) considers smoothed nonparametric kernel regression estimation. Under this approach, rather than employing mixing conditions, it is assumed that regression errors follow a general linear process representation covering both weak (spatial) dependence as well as dependence at longer ranges. Hence, the author establishes con-

☆ The authors acknowledge financial support from EPSRC (grant number EP/L021250/1).

* Corresponding author.
E-mail addresses: francesco.moscone@brunel.ac.uk (F. Moscone), elisa.tosetti@brunel.ac.uk (E. Tosetti).

http://dx.doi.org/10.1016/j.econlet.2015.05.020

sistency of the Nadaraya–Watson kernel estimate and derives its asymptotic distribution.

In this paper, we propose a *partial sample* estimator for the covariance matrix of the fixed effects (FE) and mean group (MG) estimators of the slope coefficients in a panel data model with unobserved fixed effects and errors that are weakly cross sectionally dependent and serially correlated. The idea is approximate the true covariance matrix with a (weighted) average of cross products of regression errors, computed over a subset of n cross sectional units, where $n \to \infty$ as N rises, and $n/N \to 0$.[1] We prove that the suggested partial sample estimator tends to the true covariance matrix calculated over the partial sample observations, for N going to infinity, with T fixed. A small Monte Carlo exercise reported in the paper shows that this approach is quite robust to various forms of weak cross sectional dependence, when N is large. The proposed method can be very useful for robust estimation in the context of micro-dataset where N is very large and there is little knowledge on the process generating cross sectional dependence.

2. The econometric framework

Consider the panel data model

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

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

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

where r_{ij} are (unknown) elements, possibly function of a smaller set of coefficients, of a $N \times N$ 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 $E\left(\varepsilon_{it}\varepsilon_{jt}\right) = 0$, but can be serially correlated. We make the following assumptions on the error term and regressors.

Assumption 1. ε_{it} are independently distributed across i, with $E\left(\varepsilon_{it}\right) = 0$, $E\left(\epsilon_{it}^4\right) < \infty$ and $E(\boldsymbol{\varepsilon}_{i.}\boldsymbol{\varepsilon}_{i.}') = \boldsymbol{\Omega}_i$, with $\boldsymbol{\Omega}_i$ being a finite, positive definite T-dimensional matrix with (t, s)th element, $\omega_{i,ts}$.

Assumption 2. $\max_{1\leq j\leq N} \sum_{i=1}^N \left|r_{ij}\right| < \infty$; $\max_{1\leq i\leq N} \sum_{j=1}^N \left|r_{ij}\right| < \infty$.

Assumption 3. $\mathbf{x}_{it}$ has finite elements, and $\lim_{(N,T)\to\infty} \frac{1}{NT} \sum_{i=1}^N \widetilde{\mathbf{X}}_{i.}' \widetilde{\mathbf{X}}_{i.} = \mathbf{Q}$ finite and non-singular, with $\widetilde{\mathbf{X}}_{i.} = \mathbf{M}\mathbf{X}_{i.}$, $\mathbf{X}_{i.} = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT})'$, $\mathbf{M} = \mathbf{I}_T - \mathbf{1}_T\left(\mathbf{1}_T'\mathbf{1}_T\right)^{-1}\mathbf{1}_T'$.

Under Assumptions 1–2, $0 \leq \left|E\left(e_{it}e_{js}\right)\right| = \left|\sum_{h=1}^N r_{ih}r_{jh}\omega_{h,ts}\right| < \infty$, for all i, j, t, s, and the covariance matrix of $\mathbf{e}_{.t} = (e_{1t}, e_{2t}, \ldots, e_{Nt})'$, for each t, have absolute summable elements, i.e., $\sum_{j=1}^N \left|E\left(e_{it}e_{jt}\right)\right| \leq \sum_{j=1}^N \sum_{h=1}^N |r_{ih}| \left|r_{jh}\right| \left|\omega_{h,ts}\right| < \infty$, thus carrying weak cross section dependence. A large variety of models can be cast in this framework, for example, the spatial autoregressive process having AR or MA errors, or a common factor structure with weak factors.

3. Robust estimation

The FE and MG estimators of $\boldsymbol{\beta}$ in Eq. (1) are:

$$\hat{\boldsymbol{\beta}}_P = \left(\sum_{i=1}^N \widetilde{\mathbf{X}}_{i.}'\widetilde{\mathbf{X}}_{i.}\right)^{-1} \sum_{i=1}^N \widetilde{\mathbf{X}}_{i.}'\widetilde{\mathbf{y}}_{i.}, \tag{3}$$

$$\hat{\boldsymbol{\beta}}_{MG} = \frac{1}{N}\sum_{i=1}^N \hat{\boldsymbol{\beta}}_i, \tag{4}$$

with $\widetilde{\mathbf{y}}_{i.} = \mathbf{M}\mathbf{y}_{i.}$ and $\hat{\boldsymbol{\beta}}_i = \left(\widetilde{\mathbf{X}}_{i.}'\widetilde{\mathbf{X}}_{i.}\right)^{-1}\widetilde{\mathbf{X}}_{i.}'\widetilde{\mathbf{y}}_{i.}$. Under Assumptions 1–3, it is easy to show that, as N tends to infinity and for T fixed, $\sqrt{N}\left(\hat{\boldsymbol{\beta}}_P - \boldsymbol{\beta}\right) \overset{a}{\sim} N\left(\mathbf{0}, \boldsymbol{\Sigma}_P\right)$, and $\sqrt{N}\left(\hat{\boldsymbol{\beta}}_{MG} - \boldsymbol{\beta}\right) \overset{a}{\sim} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{MG}\right)$, where (see Hansen, 2007; Pesaran and Tosetti, 2011)

$$\boldsymbol{\Sigma}_P = \mathbf{Q}^{-1}\boldsymbol{\Psi}\mathbf{Q}^{-1}, \tag{5}$$

$$\boldsymbol{\Sigma}_{MG} = \lim_{N\to\infty}\frac{1}{N}\sum_{i,j=1}^N\sum_{t,s=1}^T \mathbf{w}_{it}\mathbf{w}_{js}'\mathbf{r}_{i.}'\boldsymbol{\omega}_{st}\mathbf{r}_{j.}, \tag{6}$$

with $\mathbf{Q} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \widetilde{\mathbf{X}}_{i.}'\widetilde{\mathbf{X}}_{i.}$, $\mathbf{w}_{it}$ is the tth column of $\mathbf{W}_{i.}' = \left(\widetilde{\mathbf{X}}_{i.}'\widetilde{\mathbf{X}}_{i.}\right)^{-1}\widetilde{\mathbf{X}}_{i.}'$, $\boldsymbol{\omega}_{ts} = diag\left\{\omega_{1,ts}, \omega_{2,ts}, \ldots, \omega_{N,ts}\right\}$, and

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

Note that $\boldsymbol{\Sigma}_{MG}$ and $\boldsymbol{\Sigma}_P$ depend on the nuisance parameters in the matrices $\mathbf{R}$ and $\boldsymbol{\omega}_{ts}$. Let n be a scalar such that $n \to \infty$ as $N \to \infty$ with $n/N \to 0$, we propose the following partial sample estimators for (5) and (6), respectively:

$$\hat{\boldsymbol{\Sigma}}_P^{(n)} = \mathbf{Q}_N^{-1}\hat{\boldsymbol{\Psi}}^{(n)}\mathbf{Q}_N^{-1}, \tag{8}$$

$$\hat{\boldsymbol{\Sigma}}_{MG}^{(n)} = \frac{1}{n}\sum_{i,j=1}^n\left(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{MG}\right)\left(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}_{MG}\right)', \tag{9}$$

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

$$\hat{\boldsymbol{\Psi}}^{(n)} = \frac{1}{n}\sum_{i,j=1}^n \widetilde{\mathbf{X}}_{i.}'\hat{\mathbf{e}}_{i.}\hat{\mathbf{e}}_{j.}'\widetilde{\mathbf{X}}_{j.}, \tag{10}$$

$\hat{e}_{it} = \widetilde{y}_{it} - \hat{\boldsymbol{\beta}}_P'\widetilde{\mathbf{x}}_{it}$. Note that, in absence of cross section correlation, if we set $n = N$ and take the sum only over $i = j = 1, \ldots, N$, (8) reduces to the Arellano (1987) cluster-robust variance estimator; while the variance estimator (9) is based on the robust estimator considered in Pesaran and Smith (1995), and Pesaran (2006). The following theorem shows that the suggested partial sample estimators (8)–(9) tend to a positive definite matrix, which is the true covariance matrix computed using the partial sample observations. Such matrix is a fraction of the true variance based on N observations (see Appendix for a proof).

Theorem 1. *Let $\hat{\boldsymbol{\Sigma}}_P$ and $\hat{\boldsymbol{\Sigma}}_{MG}$ given by (8) and (9), respectively, and let n be such that $n \to \infty$, and $n/N \to 0$. Then under 1–3, for fixed T,*

$$\hat{\boldsymbol{\Psi}}^{(n)} = \boldsymbol{\Psi}^{(n)} + O_p\left(\sqrt{n/N}\right) + O_p\left(1/\sqrt{n}\right), \tag{11}$$

$$\hat{\boldsymbol{\Sigma}}_{MG}^{(n)} = \boldsymbol{\Sigma}_{MG}^{(n)} + O_p\left(\sqrt{n/N}\right) + O_p\left(1/\sqrt{n}\right), \tag{12}$$

where $\boldsymbol{\Psi}^{(n)} = \frac{1}{n}\sum_{i,j=1}^n\sum_{t,s=1}^T \widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{js}'\mathbf{r}_{i.}'\boldsymbol{\omega}_{ts}\mathbf{r}_{j.}$, and $\boldsymbol{\Sigma}_{MG}^{(n)} = \frac{1}{n}\sum_{i,j=1}^n\sum_{t,s=1}^T \mathbf{w}_{it}\mathbf{w}_{js}'\mathbf{r}_{i.}'\boldsymbol{\omega}_{ts}\mathbf{r}_{j.}$. Further, for $N \to \infty$,

$$\boldsymbol{\Psi}^{(n)} = \frac{n}{N}\boldsymbol{\Psi}^{(N)} + O_p\left(\sqrt{n}/N\right), \tag{13}$$

$$\Sigma_{MG}^{(n)} = \frac{n}{N}\Sigma_{MG}^{(N)} + O_p\left(\sqrt{n}/N\right). \tag{14}$$

[1] We found a similar idea based on partial sample briefly outlined in Bai (2009), Remark 8, although no formal proof has been provided.

Table 1
Monte Carlo results, FE estimator.

N	T	(I): Cluster-robust estimator		(II): Partial sample estimator $n = N^{0.3}$		$n = N^{0.4}$		$n = N^{0.5}$	
		Size	Power	Size	Power	Size	Power	Size	Power
		Experiment 1: $\delta_i = 0$							
300	5	0.066	0.941	0.058	0.268	0.079	0.331	0.051	0.341
500	5	0.046	0.996	0.048	0.266	0.053	0.347	0.045	0.335
1000	5	0.049	1.000	0.043	0.315	0.051	0.369	0.046	0.359
300	10	0.045	0.999	0.045	0.322	0.068	0.426	0.096	0.529
500	10	0.057	1.000	0.046	0.350	0.055	0.477	0.077	0.591
1000	10	0.051	1.000	0.042	0.387	0.048	0.504	0.057	0.602
		Experiment 2 $\delta_i \sim IIDU(0.2, 0.4)$							
300	5	0.081	0.921	0.076	0.268	0.074	0.3	0.078	0.325
500	5	0.098	0.986	0.049	0.244	0.061	0.31	0.097	0.380
1000	5	0.069	1.000	0.046	0.269	0.051	0.335	0.061	0.437
300	10	0.092	0.998	0.058	0.314	0.074	0.382	0.093	0.463
500	10	0.089	1.000	0.044	0.322	0.053	0.431	0.071	0.481
1000	10	0.083	1.000	0.045	0.354	0.047	0.447	0.056	0.589
		Experiment 3 $\delta_i \sim IIDU(0.7, 0.9)$							
300	5	0.144	0.523	0.066	0.146	0.097	0.166	0.098	0.168
500	5	0.162	0.684	0.050	0.124	0.063	0.143	0.076	0.185
1000	5	0.152	0.793	0.049	0.146	0.049	0.172	0.060	0.194
300	10	0.152	0.745	0.062	0.178	0.058	0.182	0.084	0.245
500	10	0.148	0.892	0.043	0.158	0.056	0.214	0.073	0.249
1000	10	0.148	0.948	0.041	0.166	0.046	0.220	0.056	0.301

Note: The size and power reported in Column I use the robust estimator for the variance, while those in column (II) use expression (8).

Clearly, for a small n the variance estimators converge fast to their true counterparts, but these only estimate a small fraction of the total variance. On the contrary, a large n implies slower convergence but a larger estimated fraction of the total variance. Using the above results, the Student't statistics for the unknown parameter associated to the ℓth regressor using the FE and MG estimators and their partial sample variance estimator are:

$$t_{P,\ell} = \frac{\sqrt{n/N}\left(\hat{\beta}_{P,\ell} - \beta_\ell\right)}{\sqrt{\hat{\boldsymbol{\Sigma}}^{(n)}_{P,\ell\ell}}}, \qquad t_{MG,\ell} = \frac{\sqrt{n/N}\left(\hat{\beta}_{MG,\ell} - \beta_\ell\right)}{\sqrt{\hat{\boldsymbol{\Sigma}}^{(n)}_{MG,\ell\ell}}},$$
$$\ell = 1, 2, \dots, k.$$

4. Monte Carlo experiments

Suppose y_{it} for $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$ are generated by the following panel data model

$$y_{it} = \alpha_i + \beta x_{it} + e_{it},$$

where $\beta = 1$, and the parameters α_i are generated as $\alpha_i \sim IIDN(1, 1)$, $i = 1, 2, \dots, N$ and do not change across replications. The regressor is generated as:

$$x_{it} = \alpha_i + 0.4x_{i,t-1} + u_{it},$$
$$\text{for } i = 1, 2, \dots, N, \ t = -49, \dots, 0, 1, \dots, T,$$
$$u_{it} = 0.4\sum_{j=1}^{N} s_{ij}u_{jt} + \epsilon_{it},$$

where $\epsilon_{it} \sim N(0, \sigma^2_{\epsilon i})$, $\sigma^2_{\epsilon i} \sim IIDU\ (0.5, 1.5)$, for $i = 1, 2, \dots, N$. As for the individual-specific errors, e_{it}, we assume the following spatio-temporal process:

$$e_{it} = 0.4e_{i,t-1} + v_{it},$$
$$\text{for } i = 1, 2, \dots, N, \ t = -49, \dots, 0, 1, \dots, T, \quad (15)$$
$$v_{it} = \delta_i \sum_{j=1}^{N} s_{ij}v_{jt} + \varepsilon_{it} \quad (16)$$
$$\varepsilon_{it} \sim N(0, \sigma^2_{\varepsilon i}), \quad \sigma^2_{\varepsilon i} \sim IIDU\ (0.5, 1.5)\ , \text{ for } i = 1, 2, \dots, N, \quad (17)$$

where s_{ij} are elements of a $N \times N$, time-invariant, row-normalised, 2nd order regular lattice. The first 50 observations are discarded. We carry three experiments. In Experiment 1 we consider a pure temporal process and set $\delta_i = 0$. In Experiment 2 we assume a moderate degree of spatial correlation, $\delta_i \sim IIDU(0.2, 0.4)$, while in Experiment 3 we consider a more sizeable degree of spatial dependence, setting $\delta_i \sim IIDU(0.7, 0.9)$. The Arellano (1987) cluster-robust estimator and the robust estimator considered in Pesaran and Smith (1995) deliver correct inference under Experiment 1, while they are biased under Experiments 2 and 3. The number of replications is set to 1000, experiments are carried for $N = 300, 500, 1000$, and $T = 5, 10$, and we try three alternative choices of n, by setting $n = N^{0.3}, N^{0.4}, N^{0.5}$.

4.1. Results

We report size and power for both FE and MG estimators using (8) and (9), for various choices of n.[2] For a comparison, we also report the Arellano (1987) cluster-robust estimator and the robust estimator considered in Pesaran and Smith (1995). The nominal size is set to 5%, while the power of the various tests is computed under the alternative $H_1 : \beta = 0.90$. Results reported in Tables 1 and 2 show that the proposed estimators work well.

Appendix

Lemma 1. *Consider the process e_{it} in (2). Under Assumptions 1–2, for fixed T,*

$$E\left(\frac{1}{n}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T} e_{it}e_{js}\right) = O(1), \quad \textit{and}$$
$$Var\left(\frac{1}{n}\sum_{i,j=1}^{n} e_{it}e_{jt}\right) = O\left(\frac{1}{n}\right). \quad (18)$$

[2] We have omitted to report Bias and RMSE of Pooled and MG estimators but these are available upon request.

Table 2
Monte Carlo results, Mean Group estimator.

N	T	(I): Robust estimator		(II): Partial sample estimator					
				$n=N^{0.3}$		$n=N^{0.4}$		$n=N^{0.5}$	
		Size	Power	Size	Power	Size	Power	Size	Power
		Experiment 1 $\delta_i=0$							
300	5	0.045	0.606	0.065	0.270	0.073	0.327	0.057	0.344
500	5	0.060	0.680	0.056	0.261	0.061	0.331	0.049	0.351
1000	5	0.050	0.744	0.038	0.303	0.048	0.374	0.041	0.456
300	10	0.050	0.787	0.051	0.323	0.059	0.417	0.097	0.539
500	10	0.063	0.844	0.049	0.341	0.057	0.470	0.056	0.571
1000	10	0.050	0.942	0.041	0.390	0.048	0.506	0.058	0.582
		Experiment 2 $\delta_i \sim IIDU(0.2, 0.4)$							
300	5	0.079	0.714	0.068	0.249	0.070	0.304	0.080	0.333
500	5	0.074	0.760	0.049	0.236	0.065	0.296	0.063	0.375
1000	5	0.071	0.786	0.042	0.261	0.056	0.341	0.055	0.419
300	10	0.073	0.753	0.057	0.326	0.061	0.381	0.103	0.468
500	10	0.078	0.808	0.047	0.322	0.051	0.434	0.059	0.471
1000	10	0.071	0.864	0.041	0.348	0.045	0.442	0.054	0.603
		Experiment 3 $\delta_i \sim IIDU(0.7, 0.9)$							
300	5	0.097	0.74	0.052	0.149	0.081	0.180	0.089	0.179
500	5	0.115	0.782	0.055	0.126	0.068	0.141	0.059	0.200
1000	5	0.109	0.764	0.041	0.150	0.052	0.175	0.055	0.202
300	10	0.112	0.758	0.056	0.169	0.072	0.196	0.089	0.238
500	10	0.126	0.752	0.048	0.161	0.054	0.214	0.054	0.238
1000	10	0.131	0.79	0.040	0.164	0.047	0.219	0.051	0.298

Note: The size and power reported in Column I use the Pesaran and Smith (1995) robust estimator for the variance, while those in column (II) use expression (9).

Further, we have

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{N}\sum_{t,s=1}^{T} e_{is}e_{jt} = O_p\left(\frac{n}{N}\right). \qquad (19)$$

Proof. To prove (18), we note that:

$$0 < \frac{1}{n}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T} e_{it}e_{js} = \frac{1}{n}\left[\sum_{t=1}^{T}\left(\sum_{i=1}^{n} e_{it}\right)\right]^2 \leq T\frac{1}{n}\sum_{t=1}^{T}\left(\sum_{i=1}^{n} e_{it}\right)^2$$
$$= T\frac{1}{n}\sum_{t=1}^{T}\boldsymbol{\varepsilon}'_{.t}\mathbf{R}'\mathbf{R}\boldsymbol{\varepsilon}_{.t},$$

where elements in $\boldsymbol{\varepsilon}_{.t} = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})'$ are distributed independently across i with mean zero, finite variance and finite fourth-order moments, and the matrix $\mathbf{A} = \mathbf{R}'\mathbf{R}$ has absolute summable row and column sums. It follows that we can apply Lemma 2 in Kelejian and Prucha (1999) for $\frac{1}{n}\boldsymbol{\varepsilon}'_{.t}\mathbf{A}\boldsymbol{\varepsilon}_{.t}$, $t = 1, 2, \ldots, T$, which implies (18). As for (19), let $1_{\{i\leq n\}}$ be an indicator function equal to 1 if $i \leq n$, and zero otherwise. Noting that $E\left(1_{\{i\leq n\}}\right) = \frac{n}{N}$ and $Var\left(1_{\{i\leq n\}}\right) = \frac{n}{N}\left(1-\frac{n}{N}\right)$, it follows that

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{N}\sum_{t,s=1}^{T} e_{is}e_{jt} = \frac{1}{N}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} e_{is}e_{jt}1_{\{i\leq n\}} = O_p\left(\frac{n}{N}\right). \quad \blacksquare$$

Proof of Theorem 1. We first prove the theorem for $\hat{\boldsymbol{\Sigma}}_{MG}$. Consider

$$\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{MG} = \mathbf{W}'_{i.}\widetilde{\mathbf{y}}_{i.} - \frac{1}{N}\sum_{k=1}^{N}\mathbf{W}'_{k.}\widetilde{\mathbf{y}}_{k.} = \mathbf{W}'_{i.}\widetilde{\mathbf{e}}_{i.} - \frac{1}{N}\sum_{k=1}^{N}\mathbf{W}'_{k.}\widetilde{\mathbf{e}}_{k.}$$

and noting that $\mathbf{W}'_{i.}\widetilde{\mathbf{e}}_{i.} = \mathbf{W}'_{i.}\mathbf{e}_{i.}$, we have:

$$\hat{\boldsymbol{\Sigma}}^{(n)}_{MG} = \frac{1}{n}\sum_{i,j=1}^{n}\left(\mathbf{W}'_{i.}\mathbf{e}_{i.} - \frac{1}{N}\sum_{k=1}^{N}\mathbf{W}'_{k.}\mathbf{e}_{k.}\right)\left(\mathbf{W}'_{j.}\mathbf{e}_{j.} - \frac{1}{N}\sum_{k=1}^{N}\mathbf{W}'_{k.}\mathbf{e}_{k.}\right)'$$
$$= \frac{1}{n}\sum_{i,j=1}^{n}\mathbf{W}'_{i.}\mathbf{e}_{i.}\mathbf{e}'_{j.}\mathbf{W}_{j.} - \frac{1}{n}\frac{1}{N}\sum_{i,j=1}^{n}\sum_{k=1}^{N}\mathbf{W}'_{i.}\mathbf{e}_{i.}\mathbf{e}'_{k.}\mathbf{W}_{k.}$$
$$- \frac{1}{n}\frac{1}{N}\sum_{i,j=1}^{n}\sum_{k=1}^{N}\mathbf{W}'_{k.}\mathbf{e}_{k.}\mathbf{e}'_{j.}\mathbf{W}_{j.}$$
$$+ \frac{1}{nN^2}\sum_{i,j=1}^{n}\sum_{k,h=1}^{N}\mathbf{W}'_{k.}\mathbf{e}_{k.}\mathbf{e}'_{h.}\mathbf{W}_{h.}.$$

In view of (18)–(19), we obtain[3]

$$\hat{\boldsymbol{\Sigma}}^{(n)}_{MG} = \frac{1}{n}\sum_{i,j=1}^{n}\mathbf{W}'_{i.}\mathbf{e}_{i.}\mathbf{e}'_{j.}\mathbf{W}_{j.} - \frac{1}{N}\sum_{i=1}^{n}\sum_{k=1}^{N}\mathbf{W}'_{i.}\mathbf{e}_{i.}\mathbf{e}'_{k.}\mathbf{W}_{k.}$$
$$- \frac{1}{N}\sum_{i=1}^{n}\sum_{k=1}^{N}\mathbf{W}'_{k.}\mathbf{e}_{k.}\mathbf{e}'_{i.}\mathbf{W}_{i.}$$
$$+ \frac{n}{N^2}\sum_{k,h=1}^{N}\mathbf{W}'_{k.}\mathbf{e}_{k.}\mathbf{e}'_{h.}\mathbf{W}_{h.}$$
$$= \frac{1}{n}\sum_{i,j=1}^{n}\mathbf{W}'_{i.}\mathbf{e}_{i.}\mathbf{e}'_{j.}\mathbf{W}_{j.} + O_p\left(\frac{n}{N}\right).$$

Using (18) we also have

$$E\left(\frac{1}{n}\sum_{i,j=1}^{n}\mathbf{W}'_{i.}\mathbf{e}_{i.}\mathbf{e}'_{j.}\mathbf{W}_{j.}\right) = \frac{1}{n}\sum_{i,j=1}^{n}\sum_{t,s=1}^{T}\mathbf{w}_{it}\mathbf{w}'_{js}\mathbf{r}'_{i.}\boldsymbol{\omega}_{ts}\mathbf{r}_{j.}$$
$$= \Sigma^{(n)}_{MG} = O(1), \qquad (20)$$

$$Var\left(\frac{1}{n}\sum_{i,j=1}^{n}\mathbf{W}'_{i.}\mathbf{e}_{i.}\mathbf{e}'_{j.}\mathbf{W}_{j.}\right) = O\left(\frac{1}{n}\right), \qquad (21)$$

[3] It is interesting to observe that for $n = N$ we have

$$\hat{\boldsymbol{\Sigma}}^{(n)}_{MG} = \frac{1}{N}\sum_{i,j=1}^{N}\mathbf{W}'_{i.}\mathbf{e}_{i.}\mathbf{e}'_{j.}\mathbf{W}_{j.} - \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\mathbf{W}'_{i.}\mathbf{e}_{i.}\mathbf{e}'_{k.}\mathbf{W}_{k.} - \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{N}\mathbf{W}'_{k.}\mathbf{e}_{k.}\mathbf{e}'_{i.}\mathbf{W}_{i.}$$
$$+ \frac{1}{N}\sum_{k,h=1}^{N}\mathbf{W}'_{k.}\mathbf{e}_{k.}\mathbf{e}'_{h.}\mathbf{W}_{h.} = \mathbf{0}.$$

Hence, using all observations would yield an estimator that collapses to zero.

which implies (12). To prove (14), note that

$$\Sigma_{MG}^{(n)} = \frac{1}{n}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\mathbf{w}_{it}\mathbf{w}_{js}'\mathbf{r}_{i.}'\boldsymbol{\omega}_{ts}\mathbf{r}_{j.}1_{\{i\leq n\}}1_{\{j\leq n\}},$$

having mean

$$\frac{n}{N^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\mathbf{w}_{it}\mathbf{w}_{js}'\mathbf{r}_{i.}'\boldsymbol{\omega}_{ts}\mathbf{r}_{j.} = \frac{n}{N}\Sigma_{MG}^{(N)},$$

and variance

$$\frac{1}{n^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\mathbf{w}_{it}\mathbf{w}_{js}'\mathbf{w}_{jt}\mathbf{w}_{is}'\left(\mathbf{r}_{i.}'\boldsymbol{\omega}_{ts}\mathbf{r}_{j.}\right)^2\frac{n^2}{N^2}\left(1-\frac{n}{N}\right)^2 = O\left(\frac{n}{N^2}\right).$$

It follows that, for large N, $\Sigma_{MG}^{(n)} \approx \frac{n}{N}\Sigma_{MG}^{(N)}$. Results (11) and (13) can be proved using a similar line of reasoning as above, by noting that

$$\hat{\mathbf{e}}_{i.} = \widetilde{\mathbf{e}}_{i.} + \widetilde{\mathbf{X}}_{i.}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_P\right) = \widetilde{\mathbf{e}}_{i.} - \frac{1}{N}\widetilde{\mathbf{X}}_{i.}\sum_{k=1}^{N}\mathbf{Z}_{k.}'\widetilde{\mathbf{e}}_{k.}, \tag{22}$$

where $\mathbf{Z}_{k.} = \widetilde{\mathbf{X}}_{k.}\mathbf{Q}_N^{-1} < K < \infty$, and substituting it in the expression for $\hat{\boldsymbol{\Psi}}^{(n)}$. ■

References

Arellano, M., 1987. Practitioners' corner: Computing robust standard errors for within-groups estimators. Oxford Bull. Econ. Stat. 49, 431–434.

Bai, J., 2009. Panel data modes with interactive fixed effects. Econometrica 77, 1229–1279.

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

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2011. Robust inference with multiway clustering. J. Bus. Econom. Statist..

Cameron, A.C., Miller, D.L., 2011. Robust inference with clustered data. In: Ullah, A., Giles, D.E. (Eds.), Handbook of Empirical Economics and Finance. CRC Press, pp. 1–28.

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

Driscoll, J.C., Kraay, A.C., 1998. Consistent covariance matrix estimation with spatially dependent panel data. Rev. Econ. Stat. 80, 549–560.

Hansen, C.B., 2007. Asymptotic properties of a robust variance matrix estimator for panel data when t is large. J. Econometrics 141, 597–620.

Ibragimov, R., Müller, U.K., 2009. t-statistic based correlation and heterogeneity robust inference. J. Bus. Econom. Statist..

Kelejian, H.H., Prucha, I., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. Internat. Econom. Rev. 40, 509–533.

Kelejian, H.H., Prucha, I.R., 2007. HAC estimation in a spatial framework. J. Econometrics 140, 131–154.

MacKinnon, J.G., Webb, M.D., 2014. Wild Bootstrap Inference for Wildly Different Cluster Sizes. Queen's Economics Department Working Paper 1314.

Moscone, F., Tosetti, E., 2012. Hac estimation in spatial panels. Econom. Lett. 117, 60–65.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Pesaran, M.H., 2006. Estimation and inference in large heterogenous panels with multifactor error structure. Econometrica 74, 967–1012.

Pesaran, M.H., Smith, R., 1995. Estimation of long-run relationships from dynamic heterogeneous panels. J. Econometrics 68, 79–113.

Pesaran, M.H., Tosetti, E., 2011. Large panels with common factors and spatial correlation. J. Econometrics 161, 182–202. CESifo Working Paper Series No. 2103.

Pinkse, J., Slade, M., Brett, C., 2002. Spatial price competition: a semiparametric approach. Econometrica 70, 1111–1153.

Robinson, P.M., 2007. Nonparametric spectrum estimation for spatial data. J. Statist. Plann. Inference 137, 1024–1034.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.