chosen such that minimizes BIC criterion in (5.1). The outcome for year 2012, which involves $T = 251$ voting instances and $N = 98$ senators, is displayed in Figure 1. The estimated non-zero pairwise links are displayed as a solid line in grey, length of which does not carry any information on its intensity or direction and are purely determined by ease of visualization. The nodes are colored according to party affiliations: Democrats are represented by blue, Republicans by red, and Independents by white.

It is immediately clear from Figure 1 that the Senate behaves as two almost exclusive blocks or groups, defined exclusively along partisan lines, where the Independents behave most similarly to the Democrats. It seems that the two blocks slightly overlap each other, and the results in Theorem 4 can be applied. One Republican forms a block him/herself. Bear in mind that we are using a cross-validated tuning parameter, and hence we are being conservative already in concluding a block structure in the spatial weight matrix.

It is of interest to visualize the number of political collaborations and its evolution throughout the years. To achieve this, we build two measures of cross-partisanship association for a given year. The first is based on the ratio of links with ends on Senators from different parties to the overall number of links. We name this as "Cross-Party Connections". As seen in Figure 2, it is under 3% for all years under study. The second measure is the number of Senators who are the starting points of directed links towards colleagues from different parties, who are generically named "brokers". Both measures represent the number of Senators and links that appear in the frontier and, therefore, could represent collaborative cross-partisan political connections. Both measures show very limited collaboration if compared to the overall legislative activity. It is concluded, therefore, that political affiliations are strong determinants of group identity. It also appears that frontier between the groups and scope for collaborative legislative work is very limited throughout the recent Senates history.

Table 5: Senate Composition.

Year	Congress	Rep	Dem	Ind	Votes
1993	103rd	46	55	0	395
1994					329
1995	104th	53	46	1	613
1996					306
1997	105th	54	45	1	298
1998					314
1999	106th	55	45	1	374
2000					298
2001	107th	49	50	1	380
2002					253
2003	108th	51	48	1	459
2004					216
2005	109th	54	45	1	366
2006					279
2007	110th	49	50	2	442
2008					215
2009	111th	41	61	2	397
2010					299
2011	112th	47	51	2	235
2012					251

6 Conclusion

We developed the LASSO penalization for detecting block structure in a spatial weight matrix, when the size of the panel can be close to the sample size. One distinct feature of our model is the absence of

Figure 1: Visualization of the estimated spatial weight matrix for voting, 2012.

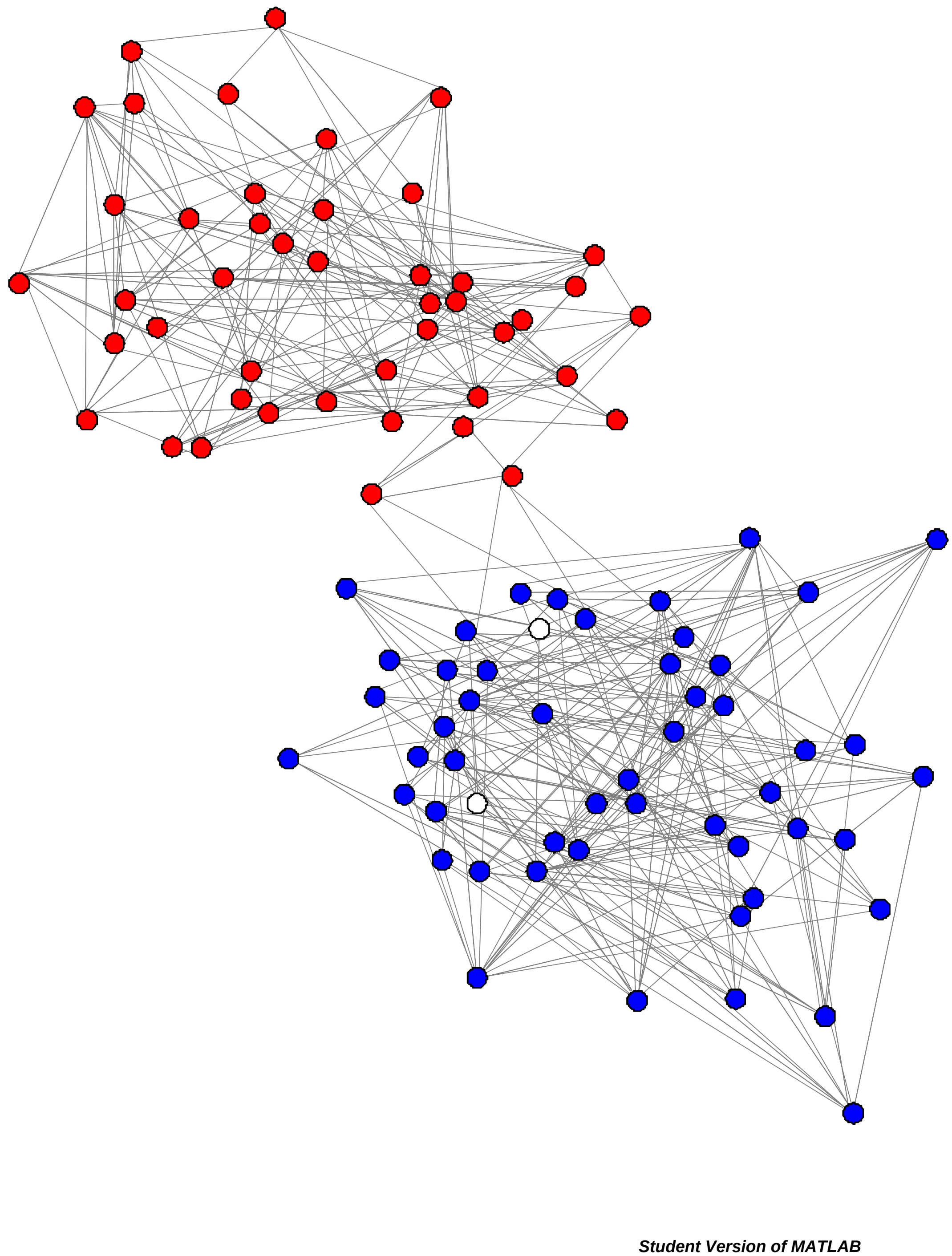
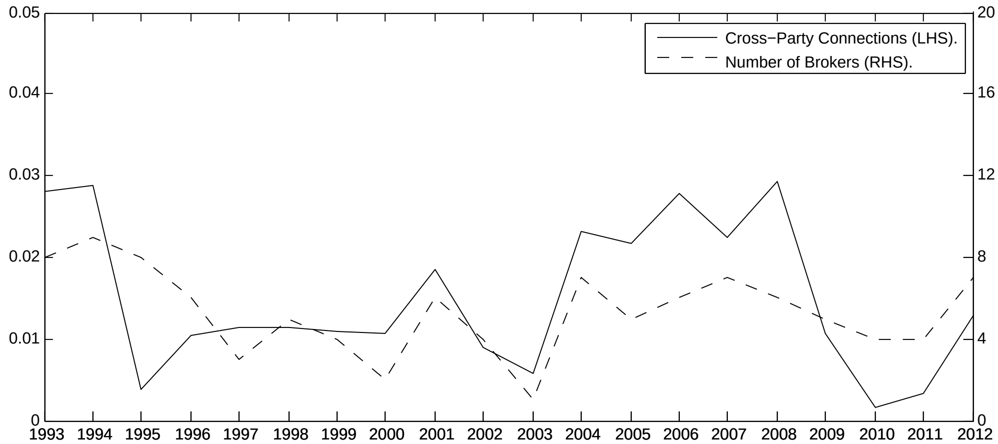


Figure 2: Cross-party collaboration.



covariates, which is motivated by the US senate voting data example analyzed in this paper. Also, there is no need for the decay of variance of the noise series, like Lam and Souza (2013) does. One contribution of the paper is the derivation of the probability lower bound for the LASSO estimator to be zero-block consistent - a concept that an estimator correctly estimates the non-diagonal zero blocks as zero. We also proved that the diagonal blocks of the estimator are not all zero with probability 1, so that block structure becomes apparent in the estimator. We use the LARS algorithm for practical computation, which is well-established for solving LASSO minimization efficiently, with computational order the same as ordinary least squares iterations. The estimated spatial weight matrix is visualized by a graph with directional edges between components. The absence of edges between two groups of components indicates two blocks. We also allow for the fact that blocks sometimes can overlap slightly, and develop the corresponding theories to show that zero-block consistency still holds in the case of slightly overlapping blocks. The US senate voting data example demonstrates clearly such a case.

Our proofs utilize results from random matrix theories for bounding extreme eigenvalues of a sample covariance matrix, as well as a Nagaev-type inequality for finding the tail probability of a general time series process. These results can be useful for the theoretical development of other time series researches.

7 Appendix

Proof of Theorem 1. For a random variable z, define the norm $\left\|z\right\|_a = [E|z|^a]^{1/a}$. We need to show that there are some constants $\mu, C > 0, w > 2$ and $\alpha > 1/2 - 1/w$ such that

$$\max_{1\le j\le N} \left\|\epsilon_{tj}\right\|_{2w} \le \mu, \tag{7.1}$$

$$\sum_{t=m}^{\infty} \max_{1\le j\le N} \left\|\epsilon_{tj} - \epsilon'_{tj}\right\|_{2w} \le Cm^{-\alpha}, \tag{7.2}$$

where $\boldsymbol{\epsilon}'_t$ has exactly the same causal definition as $\boldsymbol{\epsilon}_t$ as in assumption (iv) with the same values of $\boldsymbol{\Phi}_i$'s and $\boldsymbol{\eta}_j$'s, except for $\boldsymbol{\eta}_0$, which is replaced by an independent and identically distributed copy $\boldsymbol{\eta}'_0$. With

(7.1) and (7.2), we can use Lemma 1 of Lam and Souza (2013) for the product process $\{\epsilon_{ti}\epsilon_{tj} - E(\epsilon_{ti}\epsilon_{tj})\}$ to complete the proof.

To prove (7.1), by the Fubini's Theorem and assumption (v),

$$\begin{aligned} E|\epsilon_{tj}|^{2w} &= E\int_0^{|\epsilon_{tj}|^{2w}} ds = \int_0^\infty P(|\epsilon_{tj}| > s^{1/2w})\, ds \le \int_0^\infty D_1 \exp(-D_2 s^{q/2w})\, ds \\ &= \frac{4wD_1}{q}\int_0^\infty x^{4w/q-1}e^{-D_2x^2}\, dx = \frac{2wD_1}{qD_2^{2w/q}}\Gamma(2w/q) = \mu^{2w} < \infty, \end{aligned} \tag{7.3}$$

so that $\max_{1\le j\le N}\left\|\epsilon_{tj}\right\|_{2w} \le \mu < \infty$ for any $w > 0$. This proves (7.1).

To prove (7.2), denote $\boldsymbol{\phi}_{ij}^{\mathrm{T}}$ the j-th row of $\boldsymbol{\Phi}_i$. Then using the causal definition in assumption (iv),

$$|\epsilon_{tj} - \epsilon'_{tj}| = |\boldsymbol{\phi}_{tj}^{\mathrm{T}}(\boldsymbol{\eta}_0 - \boldsymbol{\eta}'_0)| \le \left\|\boldsymbol{\phi}_{tj}\right\|_1 \max_{i\in J_{tj}} |\eta_{0i} - \eta'_{oi}|,$$

where J_{tj} is the index set of non-zeros in $\boldsymbol{\phi}_{tj}$ as defined in assumption (vi). Hence by assumption (v) on η_{0i} and the calculations in (7.3),

$$\begin{aligned} \left\|\epsilon_{tj} - \epsilon'_{tj}\right\|_{2w} &\le \left\|\boldsymbol{\phi}_{tj}\right\|_1 \Big[E\big\{\max_{i\in J_{tj}} |\eta_{0i} - \eta'_{0i}|^{2w}\big\}\Big]^{\frac{1}{2w}} \\ &\le \left\|\boldsymbol{\phi}_{tj}\right\|_1 |J_{tj}|^{\frac{1}{2w}} \max_{i\in J_{tj}} \left\|\eta_{0i} - \eta'_{0i}\right\|_{2w} \\ &\le \left\|\boldsymbol{\phi}_{tj}\right\|_1 |J_{tj}|^{\frac{1}{2w}} (\max_{i\in J_{tj}} \left\|\eta_{0i}\right\|_{2w} + \max_{i\in J_{tj}} \left\|\eta'_{0i}\right\|_{2w}) \\ &\le 2\mu\left\|\boldsymbol{\phi}_{tj}\right\|_1 |J_{tj}|^{\frac{1}{2w}}, \end{aligned}$$

so that by assumption (vi), using the same $w > 2$ in the assumption,

$$\begin{aligned} \sum_{t=m}^\infty \max_{1\le j\le N} \left\|\epsilon_{tj} - \epsilon'_{tj}\right\|_{2w} &\le 2\mu \sum_{t=m}^\infty \max_{1\le j\le N} \left\|\boldsymbol{\phi}_{tj}\right\|_1 \max_{1\le j\le N} |J_{tj}|^{\frac{1}{2w}} \\ &\le 2\mu \max_{t,j} |J_{tj}|^{\frac{1}{2w}} \sum_{t=m}^\infty \left\|\boldsymbol{\Phi}_t\right\|_\infty \\ &\le 2\mu \max_{t,j} |J_{tj}|^{\frac{1}{2w}} Cm^{-\alpha}\big(\max_{t,j} |J_{tj}|\big)^{-\frac{1}{2w}} \\ &= 2\mu Cm^{-\alpha}, \end{aligned}$$

which is (7.2) since μ, C are constants. This completes the proof of the theorem. □

Proof of Theorem 3. Define the set

$$D = \{j : j \notin H,\ \xi_j^* \text{ does not correspond to the diagonal of } \mathbf{W}^*\},$$

and define $J = D \cup H$. Hence J contains indices for ξ_i not corresponding to the diagonal of $\mathbf{W}^*$.

The KKT condition implies that $\widetilde{\boldsymbol{\xi}}$ is a solution to (2.6) if and only if there exists a subgradient

$$\mathbf{g} = \partial|\widetilde{\boldsymbol{\xi}}| = \left\{ \mathbf{g} \in \mathbb{R}^{2N^2} : \begin{cases} g_i = 0, & i \in J^c; \\ g_i = \text{sign}(\widetilde{\xi}_i), & \widetilde{\xi}_i \neq 0; \\ |g_i| \leq 1, & \text{otherwise.} \end{cases} \right\}$$

such that, differentiating the expression to be minimized in (2.6) with respect to $\boldsymbol{\xi}_J$,

$$\frac{1}{T}\mathbf{Z}_J^{\mathrm{T}}\mathbf{Z}_J\widetilde{\boldsymbol{\xi}}_J - \frac{1}{T}\mathbf{Z}_J^{\mathrm{T}}\mathbf{y} = -\gamma_T\mathbf{g}_J,$$

where the notation $\mathbf{A}_S$ represents the matrix $\mathbf{A}$ restricted to the columns with index $j \in S$. Using $\mathbf{y} = \mathbf{Z}_J\boldsymbol{\xi}_J^* + \boldsymbol{\epsilon}$, the equation above can be written as

$$\frac{1}{T}\mathbf{Z}_J^{\mathrm{T}}\mathbf{Z}_J(\widetilde{\boldsymbol{\xi}}_J - \boldsymbol{\xi}_J^*) - \frac{1}{T}\mathbf{Z}_J^{\mathrm{T}}\boldsymbol{\epsilon} = -\gamma_T\mathbf{g}_J.$$

For $\widetilde{\boldsymbol{\xi}}$ to be zero-block consistent, we need $\widetilde{\boldsymbol{\xi}}_H = \mathbf{0}$, implying $\mathbf{Z}_J(\widetilde{\boldsymbol{\xi}}_J - \boldsymbol{\xi}_J^*) = \mathbf{Z}_D(\widetilde{\boldsymbol{\xi}}_D - \boldsymbol{\xi}_D^*)$. Hence, the KKT condition implies that $\widetilde{\boldsymbol{\xi}}$ is a zero-block consistent solution if and only if

$$\begin{aligned} &\frac{1}{T}\mathbf{Z}_H^{\mathrm{T}}\mathbf{Z}_D(\widetilde{\boldsymbol{\xi}}_D - \boldsymbol{\xi}_D^*) - \frac{1}{T}\mathbf{Z}_H^{\mathrm{T}}\boldsymbol{\epsilon} = -\gamma_T\mathbf{g}_H, \\ &\frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\mathbf{Z}_D(\widetilde{\boldsymbol{\xi}}_D - \boldsymbol{\xi}_D^*) - \frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\boldsymbol{\epsilon} = -\gamma_T\mathbf{g}_D, \end{aligned} \tag{7.4}$$

which can be simplified to

$$\left|\frac{1}{T}\mathbf{Z}_H^{\mathrm{T}}\mathbf{Z}_D\Big(\frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\mathbf{Z}_D\Big)^{-1}\Big(\frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\boldsymbol{\epsilon} - \gamma_T\mathbf{g}_D\Big) - \frac{1}{T}\mathbf{Z}_H^{\mathrm{T}}\boldsymbol{\epsilon}\right| \leq \gamma_T, \tag{7.5}$$

since $\mathbf{g}_H$ has elements less than or equal to 1.

We now show that, on the set A_ϵ as defined in (3.2), (7.5) is true for large enough T, N, thus completing the proof of zero-block consistency of $\widetilde{\boldsymbol{\xi}}$. To this end, there are four terms we need to bound. Define $I_1, \ldots, I_G \subset \{1, \ldots, N\}$ to be the index sets for the G groups of components as in (2.4). Then, consider on the set A_ϵ,

$$\begin{aligned} \Big\|\frac{1}{T}\mathbf{Z}_H^{\mathrm{T}}\boldsymbol{\epsilon}\Big\|_{\max} &= \max_{i \in I_q, j \notin I_q}\left|\frac{1}{T}\sum_{t=1}^T y_{ti}\epsilon_{tj}\right| = \max_{i \in I_q, j \notin I_q}\left|\sum_{s \in I_q}\pi_{is}^*\Big(\frac{1}{T}\sum_{t=1}^T \epsilon_{ts}\epsilon_{tj}\Big)\right| \\ &\leq \lambda_T \max_{1 \leq i \leq N}\sum_{s=1}^N |\pi_{is}^*| \leq \frac{\lambda_T}{1-\eta}, \end{aligned} \tag{7.6}$$

where we used the reduced form $\mathbf{y}_t = \boldsymbol{\Pi}^*\boldsymbol{\epsilon}_t = (\mathbf{I}_N - \mathbf{W}^*)^{-1}\boldsymbol{\epsilon}_t$ of model (2.3) and $y_{ti} = \sum_{j \in I_q}\pi_{ij}^*\epsilon_{tj}$ for $i \in I_q$ for some q, with π_{ij}^* being the (i,j)-th element of $\boldsymbol{\Pi}^* = (\mathbf{I}_N - \mathbf{W}^*)^{-1}$. The last line follows from assumption (ii) that $\text{cov}(\epsilon_{ti}, \epsilon_{tj}) = 0$ if i and j correspond to different groups, so that on A_ϵ, $|T^{-1}\sum_{t=1}^T \epsilon_{ts}\epsilon_{tj}| \leq \lambda_T$. We also used assumption (i) to arrive at

$$\max_{1 \leq i \leq N}\sum_{s=1}^N |\pi_{is}^*| = \big\|\boldsymbol{\Pi}^*\big\|_\infty \leq \big\|\mathbf{I}_N\big\|_\infty + \sum_{k \geq 1}\big\|\mathbf{W}^*\big\|_\infty^k \leq 1 + \sum_{k \geq 1}\eta^k = \frac{1}{1-\eta}.$$

A potentially larger term is, by similar calculations on A_ϵ,

$$\Big\|\frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\boldsymbol{\epsilon}\Big\|_{\max} = \max_{i\in I_q, j\in I_{q'}} \left|\sum_{s\in I_q} \pi^*_{is}\Big(\frac{1}{T}\sum_{t=1}^T \epsilon_{ts}\epsilon_{tj}\Big)\right| \leq \frac{\sigma_\epsilon^2+\lambda_T}{1-\eta}, \tag{7.7}$$

where we used assumption (ii) that $\mathrm{var}(\epsilon_{tj}) \leq \sigma_\epsilon^2$. We also have, on A_ϵ,

$$\Big\|\frac{1}{T}\mathbf{Z}_H^{\mathrm{T}}\mathbf{Z}_D\Big\|_\infty \leq n \max_{i\in I_q, j\notin I_q}\left|\frac{1}{T}\sum_{t=1}^T y_{ti}y_{tj}\right| = n \max_{\substack{i\in I_q, j\in I_{q'}\\ q\neq q'}} \left|\sum_{s\in I_q,\ell\in I_{q'}} \pi^*_{is}\pi^*_{j\ell}\Big(\frac{1}{T}\sum_{t=1}^T \epsilon_{ts}\epsilon_{t\ell}\Big)\right| \leq \frac{\lambda_T n}{(1-\eta)^2}. \tag{7.8}$$

Finally, let $\sigma_{\max}(\mathbf{A}) = \lambda_{\max}^{1/2}(\mathbf{A}^{\mathrm{T}}\mathbf{A})$ denotes the maximum singular value of the matrix $\mathbf{A}$, and $\sigma_{\min}(\mathbf{A})$ the smallest one. Then

$$\begin{aligned}\Big\|\Big(\frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\mathbf{Z}_D\Big)^{-1}\Big\|_\infty &\leq n^{1/2}\lambda_{\min}^{-1}\Big(\frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\mathbf{Z}_D\Big) \leq n^{1/2}\lambda_{\min}^{-1}\Big(\frac{1}{T}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\Big) = n^{1/2}\lambda_{\min}^{-1}\Big(\frac{1}{T}\sum_{t=1}^T \mathbf{y}_t\mathbf{y}_t^{\mathrm{T}}\Big)\\ &= n^{1/2}\lambda_{\min}^{-1}\bigg(\boldsymbol{\Pi}^*\Big(\frac{1}{T}\sum_{t=1}^T \boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{\mathrm{T}}\Big)\boldsymbol{\Pi}^{*\mathrm{T}}\bigg) \leq n^{1/2}\sigma_{\min}^{-2}(\boldsymbol{\Pi}^*)\lambda_{\min}^{-1}\Big(\frac{1}{T}\sum_{t=1}^T \boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{\mathrm{T}}\Big).\end{aligned} \tag{7.9}$$

To bound (7.9), we have

$$\sigma_{\min}^{-2}(\boldsymbol{\Pi}^*) = \sigma_{\max}^2(\mathbf{I}_N - \mathbf{W}^*) \leq (1+\sigma_{\max}(\mathbf{W}^*))^2 \leq \big(1+\|\mathbf{W}^*\|_1^{1/2}\|\mathbf{W}^*\|_\infty^{1/2}\big)^2 \leq (1+\eta^{1/2}\eta_c^{1/2})^2, \tag{7.10}$$

where we used assumption (i) for bounding $\|\mathbf{W}^*\|_1$ and $\|\mathbf{W}^*\|_\infty$.

Also, the conditions assumed in assumption (iv) for the η_{ti}'s ensure that Theorem 5.11 on the extreme eigenvalues of a sample covariance matrix in Bai and Silverstein (2010) can be applied. Hence, for each integer $i \geq 0$, we have

$$\lim_{T\to\infty}\lambda_{\min}\Big(\frac{1}{T}\sum_{t=1}^T \boldsymbol{\eta}_{t-i}\boldsymbol{\eta}_{t-i}^{\mathrm{T}}\Big) = \sigma^2(1-\sqrt{d})^2, \quad \lim_{T\to\infty}\lambda_{\max}\Big(\frac{1}{T}\sum_{t=1}^T \boldsymbol{\eta}_{t-i}\boldsymbol{\eta}_{t-i}^{\mathrm{T}}\Big) = \sigma^2(1+\sqrt{d})^2$$

almost surely, where d is specified in assumption (iii). For each i, let U_i be the almost sure set such that the above limits hold. Then on the almost sure set $U = \bigcap_{i\geq 0} U_i$, the above limits hold for all integers $i \geq 0$. Hence on U, for large enough T, N, we have

$$\lambda_{\min}^{1/2}\Big(\frac{1}{T}\sum_{t=1}^T \boldsymbol{\eta}_t\boldsymbol{\eta}_t^{\mathrm{T}}\Big) \geq \sigma(1-\sqrt{d}) - e, \quad \lambda_{\max}^{1/2}\Big(\frac{1}{T}\sum_{t=1}^T \boldsymbol{\eta}_t\boldsymbol{\eta}_t^{\mathrm{T}}\Big) \leq \sigma(1+\sqrt{d}) + e,$$

where the constant e is as in assumption (iv). Therefore, on U, for large enough T, N, we have

$$\begin{aligned}
\lambda_{\min}\Big(\frac{1}{T}\sum_{t=1}^T \boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{\mathrm{T}}\Big) &= \sigma_{\min}^2\Big(T^{-1/2}\sum_{i\geq 0}\boldsymbol{\Phi}_i(\boldsymbol{\eta}_{1-i},\ldots,\boldsymbol{\eta}_{T-i})\Big)\\
&\geq \left\{\sigma_{\min}\big(T^{-1/2}(\boldsymbol{\eta}_1,\ldots,\boldsymbol{\eta}_T)\big) - \sum_{i\geq 1}\sigma_{\max}\big(\boldsymbol{\Phi}_i T^{-1/2}(\boldsymbol{\eta}_{1-i},\ldots,\boldsymbol{\eta}_{T-i})\big)\right\}^2\\
&\geq \left\{\lambda_{\min}^{1/2}\Big(\frac{1}{T}\sum_{t=1}^T\boldsymbol{\eta}_t\boldsymbol{\eta}_t^{\mathrm{T}}\Big) - \sum_{i\geq 1}\big\|\boldsymbol{\Phi}_i\big\|\lambda_{\max}^{1/2}\Big(\frac{1}{T}\sum_{t=1}^T\boldsymbol{\eta}_{t-i}\boldsymbol{\eta}_{t-i}^{\mathrm{T}}\Big)\right\}^2\\
&\geq \left\{\sigma(1-\sqrt{d}) - e - (\sigma(1+\sqrt{d})+e)\sum_{i\geq 1}\big\|\boldsymbol{\Phi}_i\big\|\right\}^2 \geq c^2, \qquad (7.11)
\end{aligned}$$

where $c > 0$ is a constant as in assumption (iv). Combining (7.10) and (7.11), on U and for large enough T, N, (7.9) becomes

$$\Big\|\Big(\frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\mathbf{Z}_D\Big)^{-1}\Big\|_\infty \leq \frac{n^{1/2}(1+\eta^{1/2}\eta_c^{1/2})^2}{c^2}. \qquad (7.12)$$

Hence combining the bounds (7.6), (7.7), (7.8) and (7.12), on $A_\epsilon \cap U$, for large enough T, N, we have

$$\begin{aligned}
&\Big|\frac{1}{T}\mathbf{Z}_H^{\mathrm{T}}\mathbf{Z}_D\Big(\frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\mathbf{Z}_D\Big)^{-1}\Big(\frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\boldsymbol{\epsilon} - \gamma_T\mathbf{g}_D\Big) - \frac{1}{T}\mathbf{Z}_H^{\mathrm{T}}\boldsymbol{\epsilon}\Big|\\
&\leq \Big\|\frac{1}{T}\mathbf{Z}_H^{\mathrm{T}}\mathbf{Z}_D\Big\|_\infty\Big\|\Big(\frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\mathbf{Z}_D\Big)^{-1}\Big\|_\infty\Big\|\frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\boldsymbol{\epsilon} - \gamma_T\mathbf{g}_D\Big\|_{\max} + \Big\|\frac{1}{T}\mathbf{Z}_H^{\mathrm{T}}\boldsymbol{\epsilon}\Big\|_{\max}\\
&\leq \frac{\lambda_T n^{3/2}(1+\eta^{1/2}\eta_c^{1/2})^2}{(1-\eta)^2c^2}\left(\frac{\sigma_\epsilon^2+\lambda_T}{1-\eta}+\gamma_T\right) + \frac{\lambda_T}{1-\eta}\\
&= O(\lambda_T n^{3/2}) = o(\gamma_T),
\end{aligned}$$

by the assumption $n = o(\{\gamma_T/\lambda_T\}^{2/3})$. Hence on $A_\epsilon \cap U$, (7.5) is satisfied for large enough T, N, so that $\widetilde{\boldsymbol{\xi}}$ is zero-block consistent, i.e. $\widetilde{\boldsymbol{\xi}}_H = \mathbf{0}$. It is clear then for large enough T, N, $A_\epsilon \cap U \subseteq \{\widetilde{\boldsymbol{\xi}}_H = \mathbf{0}\}$, and hence

$$P(\widetilde{\boldsymbol{\xi}}_H = \mathbf{0}) \geq P(A_\epsilon \cap U) = P(A_\epsilon),$$

since U is an almost sure set. The part where $P(A_\epsilon) \to 1$ if $N = o(T^{w/4-1/2}\log^{w/4}(T))$ is given by the results of Corollary 2. This completes the proof of the first half of Theorem 3.

For the second half, suppose $\widetilde{\boldsymbol{\xi}}_D = \mathbf{0}$. Then using (7.4), we have

$$\mathbf{g}_D = \frac{1}{\gamma_T}\Big(\frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\boldsymbol{\epsilon} + \frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\mathbf{Z}_D\boldsymbol{\xi}_D^*\Big) = \frac{1}{\gamma_T}\Big(\frac{1}{T}\mathbf{Z}_D^{\mathrm{T}}\mathbf{y}\Big).$$

One of the element of $\mathbf{g}_D$ is, for some j, with T, N large enough and on U,

$$\frac{1}{\gamma_T}\Big(\frac{1}{T}\sum_{t=1}^T y_{tj}^2\Big) = \frac{1}{\gamma_T}\Big(\frac{1}{T}\sum_{t=1}^T \boldsymbol{\pi}_j^{*\mathrm{T}}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\boldsymbol{\pi}_j^*\Big) \geq \frac{\big\|\boldsymbol{\pi}_j^*\big\|^2}{\gamma_T}\lambda_{\min}\Big(\frac{1}{T}\sum_{t=1}^T\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\Big) \geq \frac{c^2}{\gamma_T},$$

where $\boldsymbol{\pi}_j^{\mathrm{T}}$ is the j-th row of $\boldsymbol{\Pi}^*$, with $\big\|\boldsymbol{\pi}_j^*\big\| > 1$, and we used (7.11). Since $\gamma_T \to 0$, we have just proved that this particular element goes to infinity as $T, N \to \infty$, which is a contradiction since all elements in

$\mathbf{g}_D$ are less than or equal to 1 in magnitude. Hence we must have $\widetilde{\boldsymbol{\xi}}_D \neq \mathbf{0}$ for large enough T, N. This completes the proof of the theorem. □

Proof of Theorem 4. Define the set

$$D' = \{j : j \notin H', \xi_j \text{ does not correspond to the diagonal of } \mathbf{W}^*\}.$$

Then the proof of this theorem is almost exactly the same as that for Theorem 3 by replacing D with D' and H with H'. The only differences are the bounds in (7.6) and (7.8). Consider, on A_ϵ,

$$\begin{aligned}
\Big\|\frac{1}{T}\mathbf{Z}_{H'}^{\mathrm{T}}\boldsymbol{\epsilon}\Big\|_{\max} &= \max_{i\in I_q, j\notin I_q}\Big|\frac{1}{T}\sum_{t=1}^T y_{ti}\epsilon_{tj}\Big| = \max_{i\in I_q, j\notin I_q}\Big|\sum_{s\in I_q}\pi_{is}^*\Big(\frac{1}{T}\sum_{t=1}^T \epsilon_{ts}\epsilon_{tj}\Big) + \sum_{s\notin I_q}\pi_{is}^*\Big(\frac{1}{T}\sum_{t=1}^T \epsilon_{ts}\epsilon_{tj}\Big)\Big| \\
&\leq \max_{s\in I_q, j\notin I_q}\Big|\frac{1}{T}\sum_{t=1}^T \epsilon_{ts}\epsilon_{tj}\Big|\big\|\boldsymbol{\Pi}^*\big\|_\infty + \max_{s\notin I_q, j\notin I_q}\Big|\frac{1}{T}\sum_{t=1}^T \epsilon_{ts}\epsilon_{tj}\Big|\max_{i\in I_q}\sum_{s\notin I_q}|\pi_{is}^*| \\
&\leq \frac{\lambda_T + c_\epsilon\lambda_T}{1-\eta} + (\sigma_\epsilon^2 + \lambda_T)c_\pi\lambda_T = O(\lambda_T),
\end{aligned} \tag{7.13}$$

where we used assumption (Rii) that $\mathrm{cov}(\epsilon_{ts}, \epsilon_{tj}) \leq c_\epsilon\lambda_T$ when $s \in I_q$ for some q and $j \notin I_\ell$ for any ℓ, and assumption (i)' that $\sum_{j\notin I_q}|\pi_{ij}^*| \leq c_\pi\lambda_T$ for $i \in I_q$. Also, on A_ϵ,

$$\begin{aligned}
\Big\|\frac{1}{T}\mathbf{Z}_{H'}^{\mathrm{T}}\mathbf{Z}_{D'}\Big\|_\infty &\leq n\max_{i\in I_q, j\notin I_q}\Big|\sum_{s\in I_q}\pi_{js}^*\Big(\frac{1}{T}\sum_{t=1}^T y_{ti}\epsilon_{ts}\Big) + \sum_{s\notin I_q}\pi_{js}^*\Big(\frac{1}{T}\sum_{t=1}^T y_{ti}\epsilon_{ts}\Big)\Big| \\
&\leq n\left(\frac{\sigma_\epsilon^2 + \lambda_T}{1-\eta}\right)c_\pi\lambda_T + n\lambda_T\left(\frac{1+c_\epsilon}{1-\eta} + c_\pi(\sigma_\epsilon^2 + \lambda_T)\right)\frac{1}{1-\eta} = O(\lambda_T n),
\end{aligned} \tag{7.14}$$

where we used (7.13) in the last line. The rates in (7.13) and (7.14) are the same as (7.6) and (7.8) respectively, and hence the results in Theorem 3 follows. □

References

Anselin, L., J. Le Gallo, and H. Jayet (2006). *Spatial panel econometrics. In: Matyas L, Sevestre P. (eds) The econometrics of panel data, fundamentals and recent developments in theory and practice* (3 ed.). Kluwer, Dordrecht.

Arbia, G. and B. Fingleton (2008). New spatial econometric techniques and applications in regional science. *Papers in Regional Science 87*(3), 311–317.

Bai, Z. and J. Silverstein (2010). *Spectral Analysis of Large Dimensional Random Matrices* (2 ed.). New York: Springer Series in Statistics.

Beenstock, M. and D. Felsenstein (2012). Nonparametric estimation of the spatial connectivity matrix using spatial panel data. *Geographical Analysis 44*(4), 386–397.

Bhattacharjee, A. and C. Jensen-Butler (2013). Estimation of the spatial weights matrix under structural constraints. *Regional Science and Urban Economics 43*(4), 617 – 634.

Brueckner, J. (2003). Strategic interaction among local governments: An overview of empirical studies. *International Regional Science Review 26*(2), 175–188.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. *Annals of Statistics 32*(2), 407–499.

Elhorst, J. (2010). Spatial panel data models. In M. M. Fischer and A. Getis (Eds.), *Handbook of Applied Spatial Analysis*, pp. 377–407. Springer Berlin Heidelberg.

Ferraty, F. and P. Vieu (2006). *Nonparametric Functional Data Analysis: Theory and Practice.* Berlin: Springer-Verlang.

Fischer, M. M. and J. Wang (2011, September). *Spatial Data Analysis: Models, Methods and Techniques (SpringerBriefs in Regional Science)* (1st Edition. ed.). Springer.

Fowler, J. (2006). Connecting the congress: A study of cosponsorship networks. *Political Analysis 71*(1), 456–487.

Lam, C. and P. C. L. Souza (2013). Regularization for spatial panel time series using the adaptive lasso. Manuscript.

LeSage, J. and R. K. Pace (2008). *Introduction to Spatial Econometrics.* Chapman and Hall.

Pinkse, J. and M. E. Slade (2010). The future of spatial econometrics. *Journal of Regional Science 50*(1), 103–117.

Pinkse, J., M. E. Slade, and C. Brett (2002). Spatial price competition: A semiparametric appraoch. *Econometrica 70*(3), 1111–1153.

Plümper, T. and E. Neumayer (2010). Model specification in the analysis of spatial dependence. *European Journal of Political Research 49*(3), 418–442.

Wang, H., B. Li, and C. Leng (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(3), 671–683.

Zou, H. (2006, December). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*, 1418–1429.