doi:http://dx.doi.org/10.5705/ss.202024.0341 Statistica Sinica: Supplement

Regularized Estimation of High-Dimensional Matrix-Variate Autoregressive Models

Hangjin Jiang[1], Baining Shen[1], Yuzhou Li[1], and Zhaoxing Gao[2] *

[1]*Center for Data Science, Zhejiang University*

[2]*School of Mathematical Sciences, University of Electronic Science and Technology of China*

Supplementary Material

Section S1 presents additional results from simulation.

Section S2 contains all the proofs of theoretical results in the main paper.

S1 Additional Results

Table S1 reports the estimation errors between starting with initial estimators $\widehat{\mathbf{B}}^{(0)}$ and $\widehat{\mathbf{A}}^{(0)}$ in **Algorithm 1**. Figure S1 illustrates the error of Kronecker product $\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$ of **Algorithm 1** and ALSE method in Chen et al. (2021) under various settings of (p_1, p_2) and (k_1, k_2). Table S3 reports cr$(\widehat{\mathbf{A}})$, cr$(\widehat{\mathbf{B}})$ and the estimation error under different settings and tuning methods from 100 inde-

*Corresponding author: zhaoxing.gao@uestc.edu.cn (Z. Gao), School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, 611731 P.R. China.

pendent replications. Figure S3 Estimation error $\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$ of ALSE and **Algorithm 2** with different tuning methods (sdCV, mCV and KSC) under different settings of (p_1, p_2) and (r_1, r_2). Table S5 and Table S6 present estimation errors of $\widehat{\mathbf{A}}$, $\widehat{\mathbf{B}}$, and $\widehat{\mathbf{S}}$ from ALSE and **Algorithm 2** tuned by sdCV and mCV, under different settings. Figure S5 displays the original data of the 10 economic indicators for the 10 countries from 1990 Q1 to 2019 Q4.

In higher-dimensional experiments with $(p_1, p_2) = (12, 15)$ and $(20, 20)$ for banded case, Figure S2 illustrates the box plots of the estimation errors $\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$ under various settings, comparing **Algorithm 1** and ALSE method in Chen et al. (2021). Table S2 further provides the average estimation errors for $\mathbf{A}$ and $\mathbf{B}$, demonstrating the convergence of their estimates. Figure S4 presents the box plots of the estimation errors $\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$ under different higher-dimensional settings with $(p_1, p_2) = (12, 15)$ and $(20, 20)$ in sparse case, comparing ALSE method in Chen et al. (2021) and **Algorithm 2** tuned by KSC, mCV and sdCV. The average estimation errors of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{S}$ for **Algorithm 2** tuned by KSC, mCV and sdCV under different higher-dimensional settings are presented in Table S4 - Table S8, respectively.

Following the reviewer's suggestion, we considered the estimation strategy using a banded initial estimator in the banded setting. Table S9 reports the estimation errors across various scenarios, comparing **Algorithm 1** with and with-

Table S1: The mean, median, and maximum of $\log_{10}(\|\widehat{\mathbf{A}}_1 - \widehat{\mathbf{A}}_2\|_F)$ and $\log_{10}(\|\widehat{\mathbf{B}}_1 - \widehat{\mathbf{B}}_2\|_F)$, where $(\widehat{\mathbf{A}}_1, \widehat{\mathbf{B}}_1)$ and $(\widehat{\mathbf{A}}_2, \widehat{\mathbf{B}}_2)$ are obtained by starting with the initial estimators $\widehat{\mathbf{B}}^{(0)}$ and $\widehat{\mathbf{A}}^{(0)}$ in **Algorithm 1**, respectively.

	$\log_{10}(\|\widehat{\mathbf{A}}_1 - \widehat{\mathbf{A}}_2\|_F)$			$\log_{10}(\|\widehat{\mathbf{B}}_1 - \widehat{\mathbf{B}}_2\|_F)$		
(p_1, p_2)	Mean	Median	Maximum	Mean	Median	Maximum
$(6, 4)$	-7.67	-7.75	-7.09	-6.64	-6.85	-6.39
$(8, 5)$	-7.57	-7.62	-7.19	-6.38	-6.43	-6.01
$(9, 6)$	-7.42	-7.52	-6.91	-6.23	-6.36	-6.00

out the banded initial estimator. The results show that both approaches yield identical estimation performance, confirming the robustness of our proposed method in **Algorithm 1**. Similar results were observed in the sparse case; however, we omit the details here for brevity.

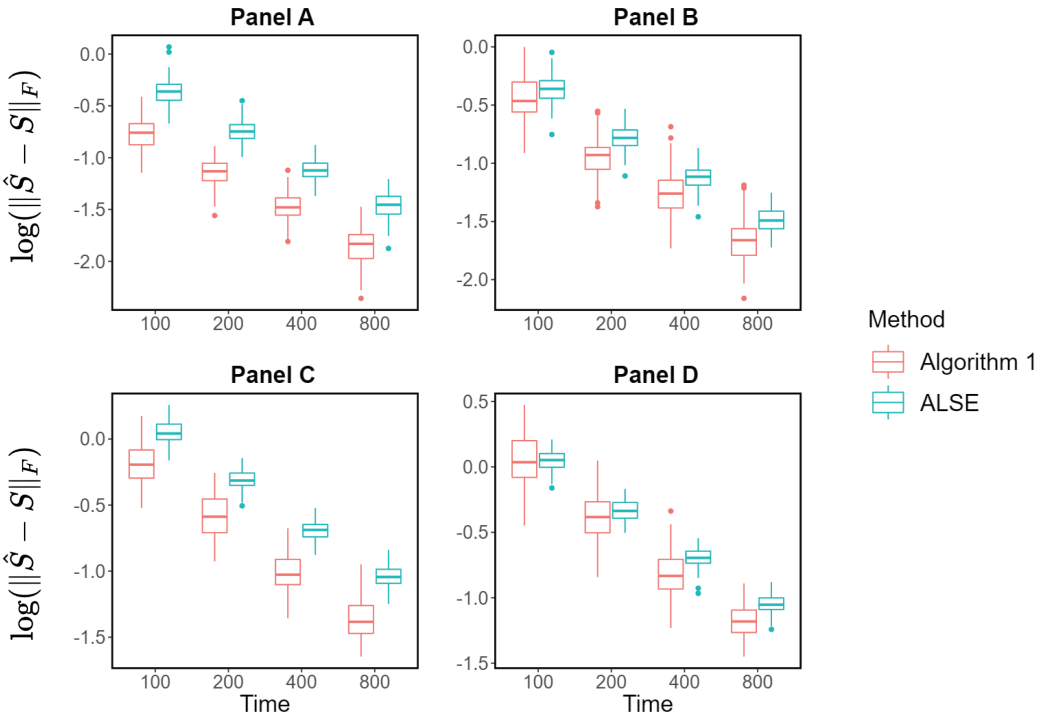

Figure S1: Error of Kronecker product $\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$ of ALSE and **Algorithm 1** under different settings of (p_1, p_2) and (k_1, k_2). **Panel A**: $(p_1, p_2) = (6, 4)$, $(k_1, k_2) = (1, 1)$; **Panel B**: $(p_1, p_2) = (6, 4)$, $(k_1, k_2) = (2, 1)$; **Panel C**: $(p_1, p_2) = (9, 6)$, $(k_1, k_2) = (2, 1)$; **Panel D**: $(p_1, p_2) = (9, 6)$, $(k_1, k_2) = (3, 2)$; Results are reported from 100 independent repetitions.

Table S2: The average estimation errors of the coefficient matrices by **Algorithm 1** and ALSE.

		Algorithm 1				ALSE			
(p_1, p_2)	ρ	T = 200	500	1000	2000	T = 200	500	1000	2000
				$\log(\|\widehat{\mathbf{A}} - \mathbf{A}\|_F)$					
$(6, 4)$	0.5	-1.960	-2.454	-2.842	-3.214	-1.830	-2.298	-2.637	-2.999
$(6, 4)$	0.8	-2.536	-3.022	-3.354	-3.677	-2.295	-2.777	-3.097	-3.462
$(9, 6)$	0.5	-1.191	-1.691	-2.236	-2.733	-1.215	-1.666	-2.009	-2.348
$(9, 6)$	0.8	-2.115	-2.499	-2.813	-3.089	-1.798	-2.267	-2.596	-2.949
				$\log(\|\widehat{\mathbf{B}} - \mathbf{B}\|_F)$					
$(6, 4)$	0.5	-1.701	-2.154	-2.474	-2.805	-1.401	-1.882	-2.208	-2.558
$(6, 4)$	0.8	-1.758	-2.232	-2.564	-2.945	-1.479	-1.961	-2.318	-2.696
$(9, 6)$	0.5	-1.314	-1.799	-2.154	-2.513	-0.860	-1.328	-1.704	-2.058
$(9, 6)$	0.8	-1.444	-1.967	-2.268	-2.615	-0.987	-1.477	-1.826	-2.175
				$\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$					
$(6, 4)$	0.5	-0.950	-1.434	-1.806	-2.166	-0.778	-1.252	-1.591	-1.946
$(6, 4)$	0.8	-1.139	-1.617	-1.953	-2.295	-0.892	-1.369	-1.705	-2.076
$(9, 6)$	0.5	-0.394	-0.885	-1.388	-1.855	-0.290	-0.753	-1.110	-1.454
$(9, 6)$	0.8	-0.735	-1.151	-1.464	-1.751	-0.381	-0.858	-1.195	-1.546

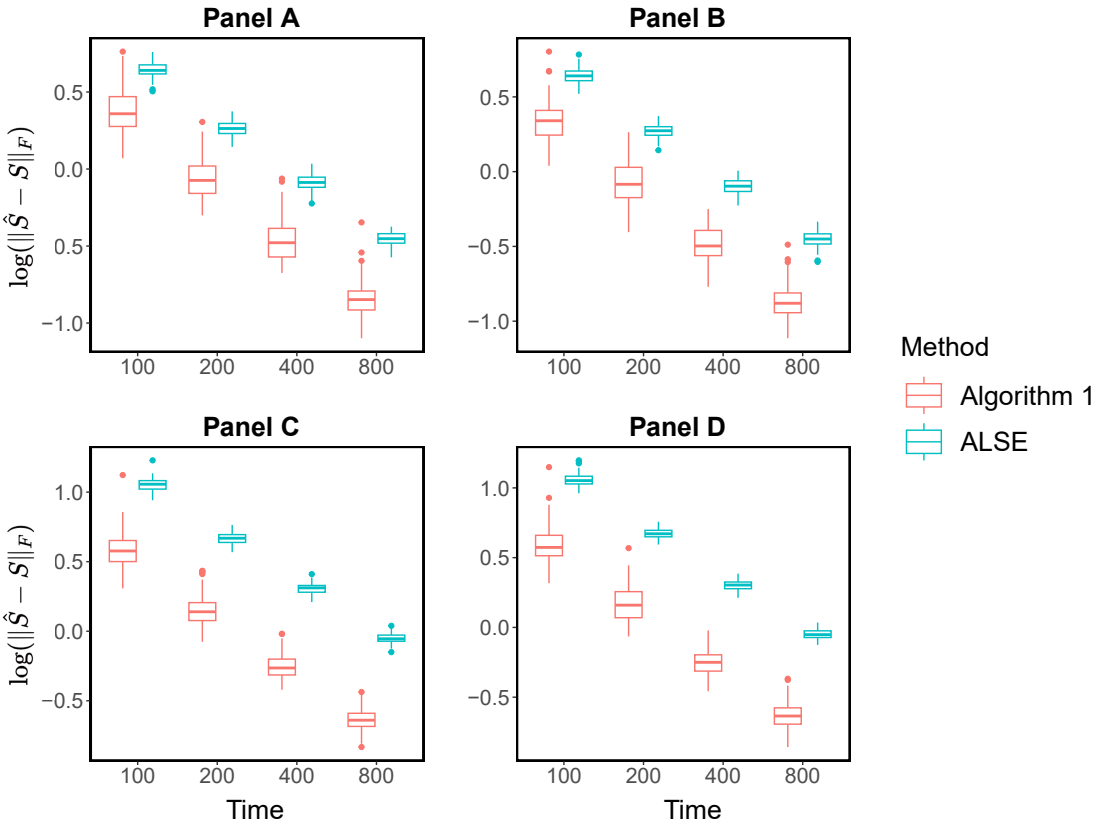

Figure S2: Log Frobenius norm error $\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$ of the Kronecker product estimation by ALSE and **Algorithm 1** under different settings of (p_1, p_2) and (k_1, k_2). **Panel A**: $(p_1, p_2) = (12, 15)$, $(k_1, k_2) = (2, 3)$; **Panel B**: $(p_1, p_2) = (12, 15)$, $(k_1, k_2) = (3, 2)$; **Panel C**: $(p_1, p_2) = (20, 20)$, $(k_1, k_2) = (2, 3)$; **Panel D**: $(p_1, p_2) = (20, 20)$, $(k_1, k_2) = (3, 2)$; Results are reported from 100 independent repetitions.

Table S3: The performance of **Algorithm 2** under different settings and different tuning methods: sdCV, mCV and KSC. E_{alse} and E_{lasso} denote the error of Kronecker product $\log(\|\widehat{\mathbf{S}}-\mathbf{S}\|_F)$ of ALSE and **Algorithm 2**, respectively.

			sdCV			mCV			KSC		
(p_1, p_2)	T	E_{alse}	cr($\widehat{\mathbf{A}}$)	cr($\widehat{\mathbf{B}}$)	E_{lasso}	cr($\widehat{\mathbf{A}}$)	cr($\widehat{\mathbf{B}}$)	E_{lasso}	cr($\widehat{\mathbf{A}}$)	cr($\widehat{\mathbf{B}}$)	E_{lasso}
$(6,4)$	100	-0.558	0.979	0.999	-0.324	0.691	0.725	-0.848	0.931	0.965	-0.797
	200	-0.906	0.987	0.999	-0.579	0.682	0.688	-1.211	0.930	0.951	-1.116
	400	-1.260	0.996	1	-0.751	0.706	0.716	-1.573	0.929	0.962	-1.525
	800	-1.616	0.998	1	-0.963	0.703	0.716	-1.920	0.936	0.966	-1.856
$(9,6)$	100	-0.247	0.93	0.983	0.001	0.663	0.674	-0.495	0.914	0.930	-0.287
	200	-0.602	0.956	0.990	-0.248	0.662	0.661	-0.856	0.915	0.944	-0.628
	400	-0.962	0.974	0.997	-0.463	0.671	0.663	-1.228	0.918	0.945	-0.990
	800	-1.305	0.981	0.999	-0.668	0.66	0.661	-1.558	0.913	0.937	-1.340

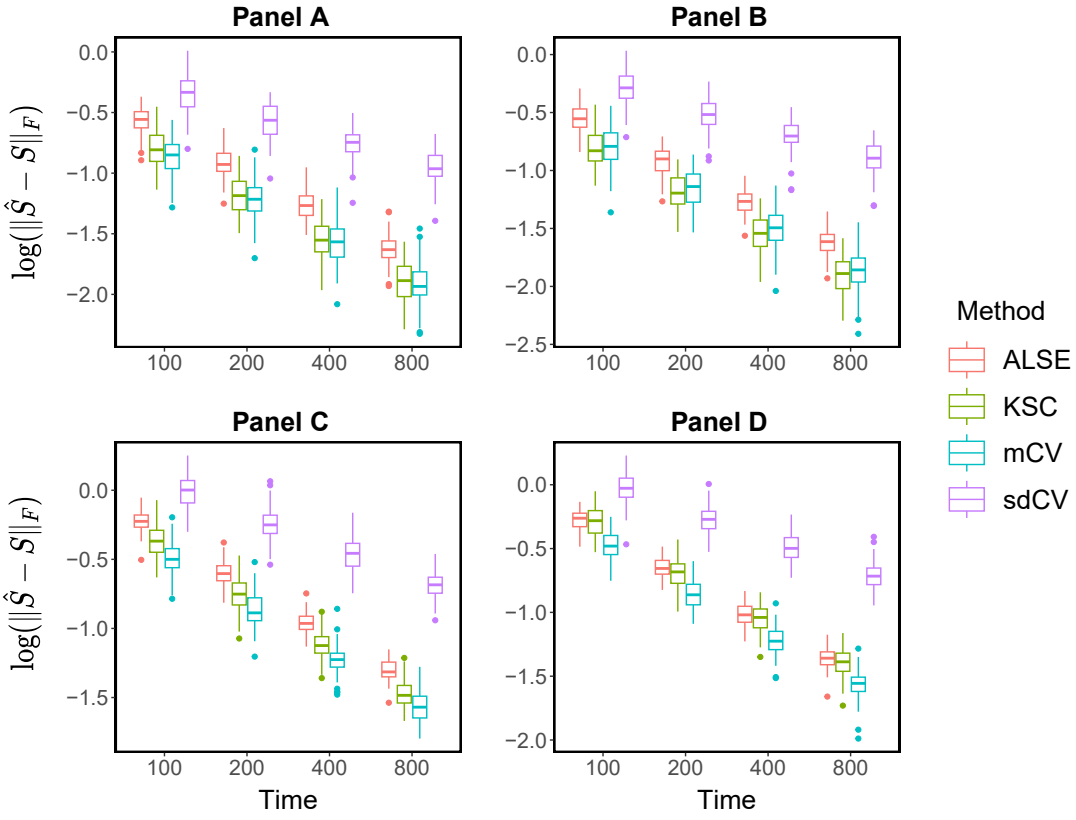

Figure S3: Estimation error $\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$ of ALSE and **Algorithm 2** with different tuning methods (sdCV, mCV and KSC) under different settings of (p_1, p_2) and (r_1, r_2). **Panel A**: $(p_1, p_2) = (6, 4)$, $(r_1, r_2) = (0.3, 0.3)$; **Panel B**: $(p_1, p_2) = (6, 4)$, $(r_1, r_2) = (0.4, 0.4)$; **Panel C**: $(p_1, p_2) = (9, 6)$, $(k_1, k_2) = (0.3, 0.3)$; **Panel D**: $(p_1, p_2) = (9, 6)$, $(k_1, k_2) = (0.4, 0.4)$. Results are reported from 100 independent repetitions.

Table S4: The estimation errors of the estimators obtained by ALSE and **Algorithm 2** with tuning parameters method KSC.

		Algorithm 2 tunned by KSC				ALSE			
(p_1, p_2)	ρ	T = 100	500	1000	2000	T = 100	500	1000	2000
				$\log(\|\widehat{\mathbf{A}} - \mathbf{A}\|_F)$					
$(6, 4)$	0.5	-1.603	-2.173	-2.553	-2.96	-1.465	-1.903	-2.243	-2.597
$(6, 4)$	0.9	-2.364	-2.853	-3.225	-3.571	-2.196	-2.645	-2.993	-3.344
$(9, 6)$	0.5	-1.753	-2.308	-2.710	-3.127	-1.546	-2.003	-2.355	-2.712
$(9, 6)$	0.9	-2.634	-3.124	-3.494	-3.893	-2.334	-2.793	-3.144	-3.495
				$\log(\|\widehat{\mathbf{B}} - \mathbf{B}\|_F)$					
$(6, 4)$	0.5	-1.319	-1.817	-2.162	-2.511	-1.265	-1.802	-2.152	-2.508
$(6, 4)$	0.9	-1.420	-1.902	-2.253	-2.612	-1.425	-1.921	-2.272	-2.631
$(9, 6)$	0.5	-0.913	-1.383	-1.737	-2.076	-0.880	-1.375	-1.730	-2.070
$(9, 6)$	0.9	-1.059	-1.521	-1.874	-2.219	-1.064	-1.538	-1.887	-2.231
				$\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$					
$(6, 4)$	0.5	-0.843	-1.389	-1.755	-2.134	-0.736	-1.216	-1.561	-1.918
$(6, 4)$	0.9	-0.990	-1.479	-1.840	-2.192	-0.891	-1.359	-1.706	-2.063
$(9, 6)$	0.5	-0.469	-0.989	-1.366	-1.742	-0.324	-0.798	-1.151	-1.505
$(9, 6)$	0.9	-0.697	-1.174	-1.534	-1.902	-0.523	-0.988	-1.338	-1.687

Table S5: The estimation errors of the estimators obtained by ALSE method in Chen et al. (2021) and **Algorithm 2** with tuning parameters method sdCV. The dimensions are $(p_1, p_2) = (6, 4)$ and $(9, 6)$, the non-zero proportions are set to be $(r_1, r_2) = (0.3, 0.3)$, and $\rho(\mathbf{A})\rho(\mathbf{B}) = 0.5$ and 0.9 are considered, and the sample size $T = 200, 500, 1000, 2000$. We report the means of three estimation errors using 100 replications.

		Algorithm 2 tunned by sdCV				ALSE			
(p_1, p_2)	ρ	T = 100	500	1000	2000	T = 100	500	1000	2000
				$\log(\|\widehat{\mathbf{A}} - \mathbf{A}\|_F)$					
$(6, 4)$	0.5	-1.337	-1.856	-2.188	-2.494	-1.465	-1.903	-2.243	-2.597
$(6, 4)$	0.9	-2.140	-2.514	-2.756	-2.962	-2.196	-2.645	-2.993	-3.344
$(9, 6)$	0.5	-1.612	-2.115	-2.478	-2.843	-1.546	-2.003	-2.355	-2.712
$(9, 6)$	0.9	-2.549	-2.940	-3.235	-3.539	-2.334	-2.793	-3.144	-3.495
				$\log(\|\widehat{\mathbf{B}} - \mathbf{B}\|_F)$					
$(6, 4)$	0.5	-0.443	-0.655	-0.829	-1.029	-1.265	-1.802	-2.152	-2.508
$(6, 4)$	0.9	-0.535	-0.746	-0.914	-1.113	-1.425	-1.921	-2.272	-2.631
$(9, 6)$	0.5	-0.201	-0.467	-0.661	-0.821	-0.880	-1.375	-1.730	-2.070
$(9, 6)$	0.9	-0.345	-0.612	-0.816	-0.976	-1.064	-1.538	-1.887	-2.231
				$\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$					
$(6, 4)$	0.5	-0.312	-0.571	-0.763	-0.973	-0.736	-1.216	-1.561	-1.918
$(6, 4)$	0.9	-0.398	-0.641	-0.820	-1.020	-0.891	-1.359	-1.706	-2.063
$(9, 6)$	0.5	-0.055	-0.364	-0.581	-0.765	-0.324	-0.798	-1.151	-1.505
$(9, 6)$	0.9	-0.223	-0.512	-0.729	-0.908	-0.523	-0.988	-1.338	-1.687

Table S6: The estimation errors of the estimators obtained by ALSE method in Chen et al. (2021) and **Algorithm 2** with tuning parameters method mCV. The dimensions are $(p_1, p_2) = (6, 4)$ and $(9, 6)$, the non-zero proportions are set to be $(r_1, r_2) = (0.3, 0.3)$, and $\rho(\mathbf{A})\rho(\mathbf{B}) = 0.5$ and 0.9 are considered, and the sample size $T = 200, 500, 1000, 2000$. We report the means of three estimation errors using 100 replications.

		Algorithm 2 tunned by mCV				ALSE			
(p_1, p_2)	ρ	T = 100	500	1000	2000	T = 100	500	1000	2000
				$\log(\|\widehat{\mathbf{A}} - \mathbf{A}\|_F)$					
$(6, 4)$	0.5	-1.659	-2.133	-2.483	-2.855	-1.465	-1.903	-2.243	-2.597
$(6, 4)$	0.9	-2.499	-2.961	-3.308	-3.648	-2.196	-2.645	-2.993	-3.344
$(9, 6)$	0.5	-1.757	-2.250	-2.616	-2.996	-1.546	-2.003	-2.355	-2.712
$(9, 6)$	0.9	-2.655	-3.116	-3.474	-3.856	-2.334	-2.793	-3.144	-3.495
				$\log(\|\widehat{\mathbf{B}} - \mathbf{B}\|_F)$					
$(6, 4)$	0.5	-1.584	-2.042	-2.352	-2.699	-1.265	-1.802	-2.152	-2.508
$(6, 4)$	0.9	-1.654	-2.094	-2.426	-2.801	-1.425	-1.921	-2.272	-2.631
$(9, 6)$	0.5	-1.121	-1.551	-1.902	-2.251	-0.880	-1.375	-1.730	-2.070
$(9, 6)$	0.9	-1.262	-1.688	-2.047	-2.399	-1.064	-1.538	-1.887	-2.231
				$\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$					
$(6, 4)$	0.5	-0.982	-1.454	-1.788	-2.152	-0.736	-1.216	-1.561	-1.918
$(6, 4)$	0.9	-1.165	-1.621	-1.956	-2.315	-0.891	-1.359	-1.706	-2.063
$(9, 6)$	0.5	-0.561	-1.03	-1.387	-1.753	-0.324	-0.798	-1.151	-1.505
$(9, 6)$	0.9	-0.801	-1.246	-1.603	-1.97	-0.523	-0.988	-1.338	-1.687

Table S7: The estimation errors of the estimators obtained by ALSE method in Chen et al. (2021) and **Algorithm 2** with tuning parameters method mCV. The dimensions are $(p_1, p_2) = (12, 15)$ and $(20, 20)$, the non-zero proportions are set to be $(r_1, r_2) = (0.3, 0.3)$, and $\rho(\mathbf{A})\rho(\mathbf{B}) = 0.5$ and 0.9 are considered, and the sample size $T = 200, 500, 1000, 2000$. We report the means of three estimation errors using 100 replications.

		Algorithm 2 tunned by mCV				ALSE			
(p_1, p_2)	ρ	T = 100	500	1000	2000	T = 100	500	1000	2000
				$\log(\|\widehat{\mathbf{A}} - \mathbf{A}\|_F)$					
$(12, 15)$	0.5	-2.229	-2.723	-3.079	-3.437	-1.976	-2.449	-2.799	-3.147
$(12, 15)$	0.9	-3.089	-3.569	-3.911	-4.267	-2.763	-3.232	-3.578	-3.924
$(20, 20)$	0.5	-2.263	-2.753	-3.106	-3.466	-1.963	-2.43	-2.775	-3.123
$(20, 20)$	0.9	-3.134	-3.615	-3.969	-4.316	-2.749	-3.218	-3.564	-3.91
				$\log(\|\widehat{\mathbf{B}} - \mathbf{B}\|_F)$					
$(12, 15)$	0.5	-0.284	-0.737	-1.072	-1.417	-0.003	-0.48	-0.825	-1.169
$(12, 15)$	0.9	-0.493	-0.955	-1.297	-1.642	-0.217	-0.689	-1.033	-1.373
$(20, 20)$	0.5	0.018	-0.441	-0.786	-1.127	0.281	-0.195	-0.548	-0.895
$(20, 20)$	0.9	-0.176	-0.633	-0.976	-1.32	0.078	-0.386	-0.738	-1.084
				$\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$					
$(12, 15)$	0.5	-0.037	-0.501	-0.842	-1.19	0.246	-0.231	-0.578	-0.924
$(12, 15)$	0.9	-0.263	-0.729	-1.07	-1.418	0.037	-0.434	-0.779	-1.122
$(20, 20)$	0.5	0.328	-0.142	-0.488	-0.836	0.621	0.148	-0.201	-0.549
$(20, 20)$	0.9	0.099	-0.366	-0.712	-1.057	0.419	-0.048	-0.397	-0.743

Table S8: The estimation errors of the estimators obtained by ALSE method in Chen et al. (2021) and **Algorithm 2** with tuning parameters method sdCV. The dimensions are $(p_1, p_2) = (12, 15)$ and $(20, 20)$, the non-zero proportions are set to be $(r_1, r_2) = (0.3, 0.3)$, and $\rho(\mathbf{A})\rho(\mathbf{B}) = 0.5$ and 0.9 are considered, and the sample size $T = 200, 500, 1000, 2000$. We report the means of three estimation errors using 100 replications.

		Algorithm 2 tunned by sdCV				ALSE			
(p_1, p_2)	ρ	T = 100	500	1000	2000	T = 100	500	1000	2000
				$\log(\|\widehat{\mathbf{A}} - \mathbf{A}\|_F)$					
$(12, 15)$	0.5	-2.219	-2.708	-3.03	-3.342	-1.976	-2.449	-2.799	-3.147
$(12, 15)$	0.9	-3.042	-3.441	-3.696	-3.958	-2.763	-3.232	-3.578	-3.924
$(20, 20)$	0.5	-2.33	-2.811	-3.163	-3.496	-1.963	-2.43	-2.775	-3.123
$(20, 20)$	0.9	-3.174	-3.591	-3.904	-4.18	-2.749	-3.218	-3.564	-3.91
				$\log(\|\widehat{\mathbf{B}} - \mathbf{B}\|_F)$					
$(12, 15)$	0.5	0.206	-0.058	-0.282	-0.487	-0.003	-0.48	-0.825	-1.169
$(12, 15)$	0.9	0.022	-0.257	-0.482	-0.687	-0.217	-0.689	-1.033	-1.373
$(20, 20)$	0.5	0.444	0.145	-0.07	-0.295	0.281	-0.195	-0.548	-0.895
$(20, 20)$	0.9	0.314	0.012	-0.197	-0.429	0.078	-0.386	-0.738	-1.084
				$\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$					
$(12, 15)$	0.5	0.303	0.009	-0.224	-0.438	0.246	-0.231	-0.578	-0.924
$(12, 15)$	0.9	0.121	-0.177	-0.404	-0.615	0.037	-0.434	-0.779	-1.122
$(20, 20)$	0.5	0.571	0.24	0.006	-0.232	0.621	0.148	-0.201	-0.549
$(20, 20)$	0.9	0.424	0.103	-0.121	-0.358	0.419	-0.048	-0.397	-0.743

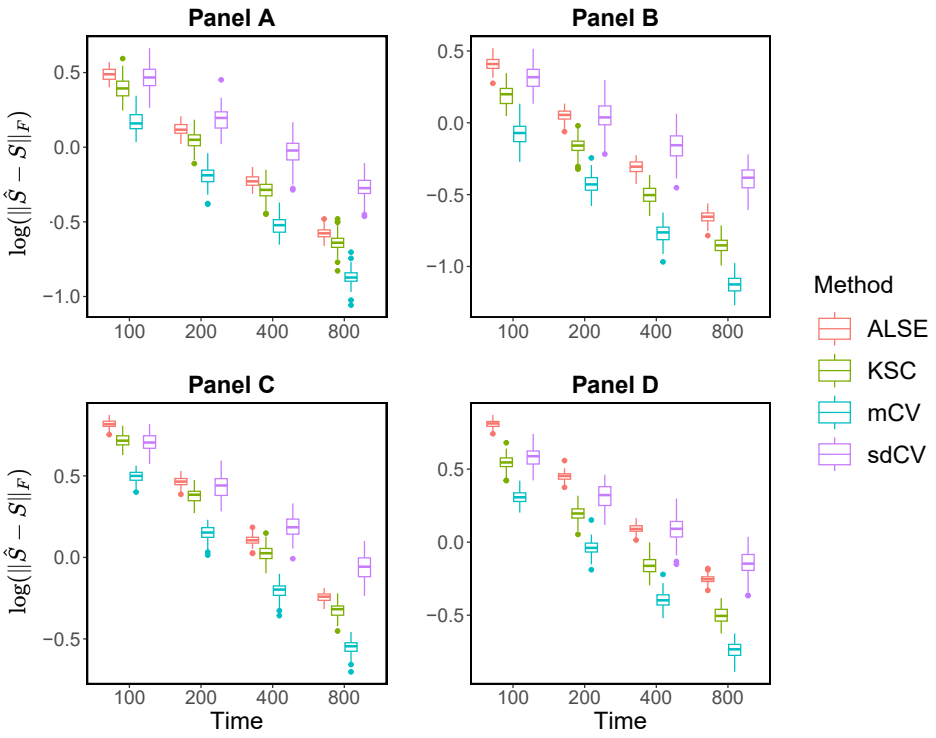

Figure S4: Estimation error $\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$ of ALSE and **Algorithm 2** using different tuning methods (sdCV, mCV and KSC) under various settings of (p_1, p_2) and (r_1, r_2). **Panel A**: $(p_1, p_2) = (12, 15)$, $(r_1, r_2) = (0.3, 0.3)$; **Panel B**: $(p_1, p_2) = (12, 15)$, $(r_1, r_2) = (0.2, 0.2)$; **Panel C**: $(p_1, p_2) = (20, 20)$, $(k_1, k_2) = (0.3, 0.3)$; **Panel D**: $(p_1, p_2) = (20, 20)$, $(k_1, k_2) = (0.2, 0.2)$. Results are reported from 100 independent repetitions.

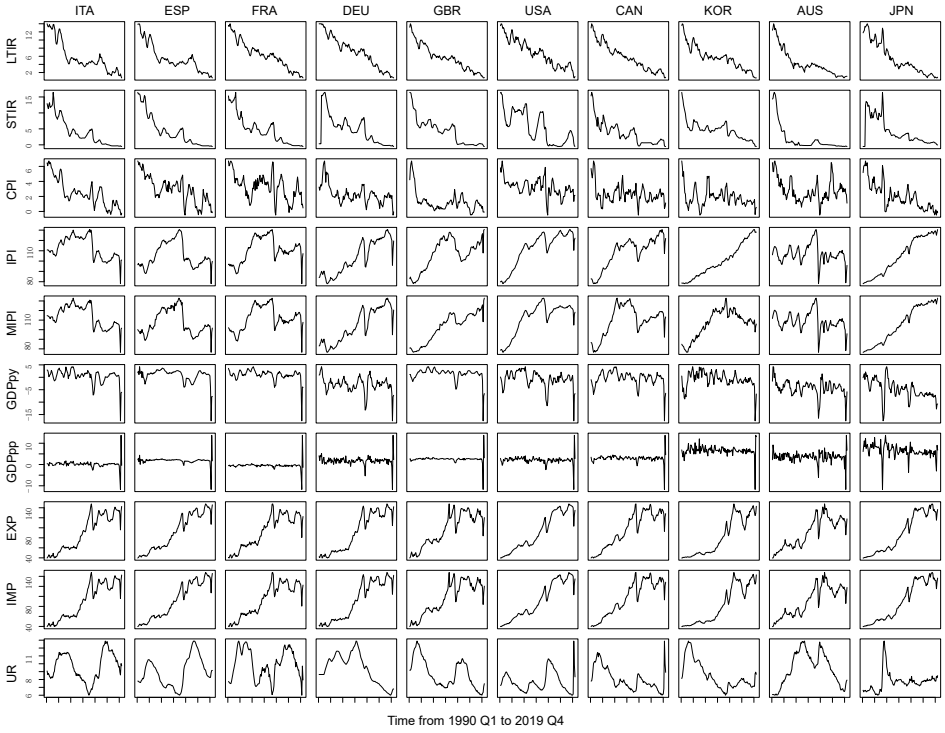

Figure S5: Time series plots of the 10 economic indicators for the 10 countries from 1990 Q1 to 2019 Q4.

Table S9: In the banded case, the estimation errors for **Algorithm 1** estimated by different initial estimator. Both estimation methods yielded identical estimation results across all simulation experiment settings. We report the means of three estimation errors using 100 replications.

	Algorithm 1				Algorithm 1 with banded initial value			
(p_1, p_2, k_1, k_2)	T = 200	500	1000	2000	T = 200	500	1000	2000
				$\log(\|\widehat{\mathbf{A}} - \mathbf{A}\|_F)$				
$(6, 4, 1, 1)$	-2.246	-2.752	-3.103	-3.445	-2.246	-2.752	-3.103	-3.445
$(9, 6, 2, 1)$	-2.532	-3.073	-3.418	-3.761	-2.532	-3.073	-3.418	-3.761
$(12, 15, 2, 3)$	-2.924	-3.354	-3.685	-4.065	-2.924	-3.354	-3.685	-4.065
				$\log(\|\widehat{\mathbf{B}} - \mathbf{B}\|_F)$				
$(6, 4, 1, 1)$	-1.792	-2.272	-2.589	-2.957	-1.792	-2.272	-2.589	-2.957
$(9, 6, 2, 1)$	-1.572	-2.075	-2.381	-2.769	-1.572	-2.075	-2.381	-2.769
$(12, 15, 2, 3)$	-0.486	-1.057	-1.404	-1.735	-0.486	-1.057	-1.404	-1.735
				$\log(\|\widehat{\mathbf{S}} - \mathbf{S}\|_F)$				
$(6, 4, 1, 1)$	-1.246	-1.738	-2.066	-2.422	-1.246	-1.738	-2.066	-2.422
$(9, 6, 2, 1)$	-0.885	-1.419	-1.752	-2.108	-0.885	-1.419	-1.752	-2.108
$(12, 15, 2, 3)$	-0.278	-0.799	-1.139	-1.490	-0.278	-0.799	-1.139	-1.490

S2 Proofs

In this section, we present proofs of main results in the paper. The idea for proving Theorem 1 is similar to that in Guo et al. (2016), but our setting is different since we consider matrix-variate time series. The proof for Theorem 3 is similar to that for proving the consistency of lasso estimator. We use C to denote a generic positive constant, which may change at different places.

Before going to prove the main results, we present some lemmas.

Lemma S1 (Merlevede, Peligrad and Rio, 2009)**.** *Let $(X_j)_{j\in\mathbb{Z}}$ be a sequence of centered real valued random variables. For any positive M, define $\phi_M(x) = (x \wedge M) \vee (-M)$ and*

$$V = \sup_{M\geq 1}\sup_{i>0}\left(\mathrm{Var}\left(\varphi_M\left(X_i\right)\right) + 2\sum_{j>i}\left|\mathrm{Cov}\left(\varphi_M\left(X_i\right),\varphi_M\left(X_j\right)\right)\right|\right).$$

Assume that (a) $(X_j)_{j\in\mathbb{Z}}$ is strongly mixing with strong mixing coefficients such that $\alpha(n) \leq \exp(-cn^{\gamma_1})$ and (b) $\sup_{k>0}\mathbb{P}(|X_k| > t) \leq \exp(1 - t^{\gamma_2})$, then V is finite and, for any $n \geq 4$, there exit positive constants C_1, C_2, C_3 and C_4 depending only on c, γ_1, and γ_2, such that, for any positive x,

$$\mathbb{P}\left(\sup_{j\leq n}|S_j| \geq x\right) \leq n\exp\left(-\frac{x^\gamma}{C_1}\right) + \exp\left(-\frac{x^2}{C_2(1+nV)}\right) + \exp\left(-\frac{x^2}{C_3 n}\exp\left(\frac{x^{\gamma(1-\gamma)}}{C_4(\log x)^\gamma}\right)\right),$$

where $\gamma = (1/\gamma_1 + 1/\gamma_2)^{-1}$, and $S_j = \sum_{i=1}^{j} X_i$.

Lemma S2. *Let conditions A1-A2, A4 and A6 hold. For any two sub-matrices $\widehat{\mathbf{Z}}_t$ and $\widehat{\mathbf{U}}_t$, each of which consisting of finite number of columns of $\widehat{\mathbf{Q}}_t'$ (or $\widehat{\mathbf{R}}_t$) with*

$\widehat{\mathbf{Z}}_t \neq \widehat{\mathbf{U}}_t$, *if* $\|\widehat{\mathbf{B}}^{(i-1)}-\mathbf{B}\|_F \to 0$(*or* $\|\widehat{\mathbf{A}}^{(i-1)}-\mathbf{A}\|_F \to 0$), *there exist positive constants* $\lambda_1 \leq \lambda_2$, *such that (1)* $\lambda_1 \leq \lambda_{\min}(T^{-1}p^{-1}\widehat{\mathbf{U}}'\widehat{\mathbf{U}}) \leq \lambda_{\max}(T^{-1}p^{-1}\widehat{\mathbf{U}}'\widehat{\mathbf{U}}) \leq \lambda_2$, *and (2)* $\lambda_1 \leq \lambda_{\min}(T^{-1}p^{-1}\widehat{\mathbf{Z}}(\mathbf{I}-\mathbf{H}_{\widehat{\mathbf{U}}})\widehat{\mathbf{Z}}') \leq \lambda_{\max}(T^{-1}p^{-1}\widehat{\mathbf{Z}}(\mathbf{I}-\mathbf{H}_{\widehat{\mathbf{U}}})\widehat{\mathbf{Z}}') \leq \lambda_2$, *where* $\widehat{\mathbf{U}} = (\widehat{\mathbf{U}}_1', \cdots, \widehat{\mathbf{U}}_T')$ *and* $\widehat{\mathbf{Z}} = (\widehat{\mathbf{Z}}_1', \cdots, \widehat{\mathbf{Z}}_T')$, *and* $p = p_2$ *or* p_1 *depending on whether we use the sub-columns of* $\widehat{\mathbf{Q}}_t'$ *or* $\widehat{\mathbf{R}}_t$ *correspondingly.*

Proof. In the following, $\widehat{\mathbf{B}}^{(i-1)}$ is written as $\widehat{\mathbf{B}}$ to simplify notations. We only prove the results for $\widehat{\mathbf{Q}}_t'$ since those for $\widehat{\mathbf{R}}_t$ can be carried out in a similar way.

Without loss of generality, we assume $\widehat{\mathbf{Z}}_t \in R^{p_2 \times s}$ and $\widehat{\mathbf{U}}_t \in R^{p_2 \times l}$ are two sub-matrices of $\widehat{\mathbf{Q}}_t'$ consisting of its s and l columns, respectively.

By the definition of $\widehat{\mathbf{U}}_t$ and $\widehat{\mathbf{Z}}_t$, there exist two matrices $\mathbf{Y}_{1,t}' \in R^{p_2 \times s}$ and $\mathbf{Y}_{2,t}' \in R^{p_2 \times l}$ consisting of the sub-columns of $\mathbf{Y}_t'$ such that $\widehat{\mathbf{U}}_t = \widehat{\mathbf{B}}\mathbf{Y}_{1,t}'$ and $\widehat{\mathbf{Z}}_t = \widehat{\mathbf{B}}\mathbf{Y}_{2,t}'$. Let $\mathbf{U}_t = \mathbf{B}\mathbf{Y}_{1,t}'$ and $\mathbf{Z}_t = \mathbf{B}\mathbf{Y}_{2,t}'$, we have

$$
\begin{aligned}
\frac{1}{Tp_2}\widehat{\mathbf{U}}'\widehat{\mathbf{U}} =& \frac{1}{Tp_2}\sum_{t=1}^{T}\widehat{\mathbf{U}}_t'\widehat{\mathbf{U}}_t \\
=& \frac{1}{Tp_2}\sum_{t=1}^{T}\mathbf{Y}_{1,t}\widehat{\mathbf{B}}'\widehat{\mathbf{B}}\mathbf{Y}_{1,t}' \\
=& \frac{1}{Tp_2}\sum_{t=1}^{T}\mathbf{Y}_{1,t}\mathbf{B}'\mathbf{B}\mathbf{Y}_{1,t}' + \frac{1}{Tp_2}\sum_{t=1}^{T}\mathbf{Y}_{1,t}\mathbf{B}'(\widehat{\mathbf{B}}-\mathbf{B})\mathbf{Y}_{1,t}' + \frac{1}{Tp_2}\sum_{t=1}^{T}\mathbf{Y}_{1,t}(\widehat{\mathbf{B}}-\mathbf{B})'\mathbf{B}\mathbf{Y}_{1,t}' \\
&+ \frac{1}{Tp_2}\sum_{t=1}^{T}\mathbf{Y}_{1,t}(\widehat{\mathbf{B}}-\mathbf{B})'(\widehat{\mathbf{B}}-\mathbf{B})\mathbf{Y}_{1,t}' \\
=& \Pi_1 + \Pi_2 + \Pi_3 + \Pi_4. \qquad \text{(S2.1)}
\end{aligned}
$$

Under the assumption that $\|\mathbf{B}\|_2$ is finite, we can show that each entry of $\mathbf{Y}_{1,t}\mathbf{B}'\mathbf{B}\mathbf{Y}_{1,t}'/p_2$

has finite variance. Under the α-mixing condition in Assumption A1, by a similar argument as (A.2) in the supplement of Gao and Tsay (2021, JASA), we can show that

$$\|\Pi_1 - p_2^{-1}E(\mathbf{U}_t'\mathbf{U}_t)\|_2 = \|\frac{1}{Tp_2}\sum_{t=1}^{T}\mathbf{Y}_{1,t}\mathbf{B}'\mathbf{B}\mathbf{Y}_{1,t}' - p_2^{-1}E(\mathbf{U}_t'\mathbf{U}_t)\|_2 = O_p(s/\sqrt{T}).$$

By a similar argument as above, we can show that

$$\|\Pi_2\|_2 = O_p(s\|\widehat{\mathbf{B}}-\mathbf{B}\|_2), \|\Pi_3\|_2 = O_p(s\|\widehat{\mathbf{B}}-\mathbf{B}\|_2), \text{and } \|\Pi_4\|_2 = O_p(s\|\widehat{\mathbf{B}}-\mathbf{B}\|_2^2).$$

Note that s is finite and $\|\widehat{\mathbf{B}} - \mathbf{B}\|_F = o_p(1)$, by (S2.1) and the asymptotic rates specified above, we obtain that

$$\|\frac{1}{Tp_2}\widehat{\mathbf{U}}\widehat{\mathbf{U}}' - p_2^{-1}E(\mathbf{U}_t'\mathbf{U}_t)\|_2 = o_p(1).$$

It follows from Assumption A4 and Weyl's inequality that, with probability tending to one,

$$\lambda_1 \leq \lambda_{\min}(T^{-1}p_2^{-1}\widehat{\mathbf{U}}\widehat{\mathbf{U}}') \leq \lambda_{\max}(T^{-1}p_2^{-1}\widehat{\mathbf{U}}\widehat{\mathbf{U}}') \leq \lambda_2,$$

which proves the statement in Lemma S2(1). The proof for Lemma S2(2) can be done by a similar argument as above, and we omit the details. This completes the proof. □

Lemma S3. *Let* $\widehat{\mathbf{B}}^{(i-1)}$ *be the latest estimator of* $\mathbf{B}$ *in* ***Algorithm 1*** *with* $\widehat{\mathbf{B}}^0 = \widehat{\mathbf{B}}_0$ *and* $i \geq 1$*, and* $\mathbf{x}_{k,t}$ *be* k*-th column of* $\widehat{\mathbf{Q}}_t'$*,* $\mathbf{x}_k = [\mathbf{x}_{k,1}', \cdots, \mathbf{x}_{k,T}']'$ *and* $\mathbf{e}_j =$

$[\mathbf{i}_j'\mathbf{E}_2, \cdots, \mathbf{i}_j'\mathbf{E}_{T+1}]'$, *for* $j, k = 1, 2, \cdots, p_1$. *For* $i > 1$, *let* $p = \max\{p_1, p_2\}$, *if* $p^5/T \to 0$, *(a) under conditions A1-A2 and A6, we have*

$$\sup_{1\le j,k\le p_1} |\mathbf{e}_j'\mathbf{x}_k| = O_p\{(Tp_2 \log p_1)^{1/2}\}$$

(b) following notations for model (2.5), under conditions A1-2 and A4-6, for any finite $k \ge k_1$, *we have,*

$$\sup_{1\le j\le p_1} |\frac{RSS_j(k)}{Tp_2\sigma_1^2} - 1| = O_p\{(T^{-1}p_2^{-1}\log p_1)^{1/2}\}$$

Proof. In the following, $\widehat{\mathbf{B}}^{(i-1)}$ is written as $\widehat{\mathbf{B}}$ to simplify notations, and let $\mathbf{\Delta} = (\mathbf{B} - \widehat{\mathbf{B}})' = (\Delta_{ij})$. Note that for $i > 1$, $\mathbf{\Delta}$ is banded.

Part (a). Since $\mathbf{x}_{k,t}$ be k-th column of $\widehat{\mathbf{Q}}_t'$, we have $\mathbf{x}_{k,t} = \widehat{\mathbf{Q}}_t'\mathbf{i}_k$. Accordingly, we define $\mathbf{u}_{k,t} = \mathbf{Q}_t'\mathbf{i}_k$ as k-th column of $\mathbf{Q}_t' = (q_{t,ij})$, with $\mathbf{Q}_t = \mathbf{Y}_t\mathbf{B}'$, and $\mathbf{u}_k = [\mathbf{u}_{k,1}', \cdots, \mathbf{u}_{k,T}']'$. Now, we have,

$$\sup_{1\le j,k\le p_1} |\mathbf{e}_j'\mathbf{x}_k| \le \sup_{1\le j,k\le p_1} |\mathbf{e}_j'\mathbf{u}_k| + \sup_{1\le j,k\le p_1} |\mathbf{e}_j'(\mathbf{x}_k - \mathbf{u}_k)|.$$

First, according to definition, we have $\mathbf{e}_j'\mathbf{u}_k = p_2^{1/2}\sum_{t=1}^T p_2^{-1/2}\mathbf{i}_j'\mathbf{E}_{t+1}\mathbf{u}_{k,t} = p_2^{1/2}\sum_{t=1}^T z_t$, where $z_t = p_2^{-1/2}\mathbf{e}_{j,t+1}'\mathbf{u}_{k,t}$ and $\mathbf{e}_{j,t+1} = \mathbf{E}_{t+1}'\mathbf{i}_j$. Under assumption A1(b) and A2(b), the following claim holds for z_t.

Claim 1. Under assumption A1(b) and A2(b), there exist some $\gamma > 0$, such that $\sup_{t>0}\mathbb{P}(|z_t| > x) \le C\exp(1 - C|x|^\gamma)$.

Combing Claim 1 and Lemma S1, we have $p_2^{-1/2}\sup_{1\le j,k\le p_1}|\mathbf{e}_j'\mathbf{u}_k| = O_p\{(T\log p_1)^{1/2}\}$, thus, $\sup_{1\le j,k\le p_1}|\mathbf{e}_j'\mathbf{u}_k| = O_p\{(Tp_2\log p_1)^{1/2}\}$.

Proof of Claim 1.

Under assumption A1(b) and A2(b), for any $s > 0$, we have:

$$\sup_{ijt} \mathbb{P}(|y_{ij,t}| > s) \leq \exp(1 - s^{\gamma_2}) \text{ and } \sup_{ijt} \mathbb{P}(|e_{ij,t}| > s) \leq \exp(1 - s^{\gamma_3}),$$

Under assumption A6, there exist a constant C, such that for elements in $\mathbf{B} = (b_{ij})$, we have $|b_{ij}| \leq C$.

$$\begin{aligned}
\sup_{ijt} \mathbb{P}(|q_{t,ij}| > s) &= \sup_{ijt} \mathbb{P}(\sum_{|k-j|\leq k_2} y_{ik,t} b_{kj} > s) \\
&\leq \sup_{ijt} \mathbb{P}(\sum_{|k-j|\leq k_2} y_{ik,t} > s/C) \\
&\leq (2k_2+1) \sup_{ikt} \mathbb{P}(y_{ik,t} > \frac{s}{C(2k_2+1)}) \\
&\leq \exp(1 - (\frac{s}{C(2k_2+1)})^{\gamma_2} + \log(2k_2+1))
\end{aligned}$$

Thus, there exist $c_1 > 0$, such that $\sup_{ijt} \mathbb{P}(|q_{t,ij}| > s) \leq \exp(1 - c_2(\frac{s}{c_1})^{\gamma_2})$.

$$\begin{aligned}
\sup_{t} \mathbb{P}(|z_t| > s) &= \sup_{t} \mathbb{P}(\sum_{i=1}^{p_2} |p_2^{-1/2} e_{ji,t+1} q_{t,ik}| > s) \\
&\leq p_2 \sup_{ijkt} \mathbb{P}(|e_{ji,t+1} q_{t,ik}| > s p_2^{-1/2}) \\
&\leq p_2 \{\sup_{ikt} \mathbb{P}(M|q_{t+1,ik}| > s p_2^{-1/2}) + \sup_{ijt} \mathbb{P}(|e_{ji,t+1}| > M)\} \\
&\leq 2p_2 \exp(1 - (s p_2^{-1/2}/c_1)^{\gamma}),
\end{aligned}$$

where $\gamma = \gamma_2\gamma_3/(\gamma_2 + \gamma_3)$ and $M = (s p_2^{-1/2}/c_1)^{\gamma_2/(\gamma_2+\gamma_3)}$. Thus, there exist $c_2 > 0$, such that $\sup_t \mathbb{P}(|z_t| > s) \leq c_2 \exp(1 - (\frac{s}{c_2})^{\gamma})$. This proves Claim 1.

Second, when conditions A1-A2 hold, we have,

$$\begin{aligned}\sup_{1\le j,k\le p_1}|\mathbf{e}_j'(\mathbf{x}_k-\mathbf{u}_k)| &= \sup_{1\le j,k\le p_1}|\sum_{t=1}^{T}\mathbf{i}_j\mathbf{E}_{t+1}'\boldsymbol{\Delta}\mathbf{Y}_t'\mathbf{i}_k| \\ &= \sup_{1\le j,k\le p_1}|\sum_{t=1}^{T}\sum_{1\le l,s\le p_2}e_{jl,t+1}\Delta_{ls}y_{kst}| \\ &\le (\sum_{1\le l,s\le p_2}|\Delta_{ls}|)\sup_{1\le j,k\le p_1}|\sum_{t=1}^{T}e_{jl,t+1}y_{kst}| \\ &= O_p(\sqrt{p_2}\|\boldsymbol{\Delta}\|_F\sqrt{T\log p_1}) \qquad \text{(S2.2)}\end{aligned}$$

Part (b). According to model (2.5), when $k \ge k_1$, we have $\text{RSS}_j(k) = \boldsymbol{f}_j'\boldsymbol{f}_j - \boldsymbol{f}_j'\mathbf{H}_{\mathbf{X}_{j,k}}\boldsymbol{f}_j$. By equation (2.4),

$$\boldsymbol{f}_j = [\mathbf{i}_j'\boldsymbol{\Delta}_1,\cdots,\mathbf{i}_j'\boldsymbol{\Delta}_T]' + \mathbf{e}_j = \boldsymbol{\delta}_j + \mathbf{e}_j, \qquad \text{(S2.3)}$$

where $\boldsymbol{\Delta}_t = \mathbf{A}\mathbf{Y}_t(\mathbf{B}-\widehat{\mathbf{B}})'$ for $t = 1, 2, \cdots, T$. Thus, $\boldsymbol{f}_j'\boldsymbol{f}_j = \boldsymbol{\delta}_j'\boldsymbol{\delta}_j + 2\boldsymbol{\delta}_j'\mathbf{e}_j + \mathbf{e}_j'\mathbf{e}_j$, and $\boldsymbol{f}_j'\mathbf{H}_{\mathbf{X}_{j,k}}\boldsymbol{f}_j = \boldsymbol{\delta}_j'\mathbf{H}_{\mathbf{X}_{j,k}}\boldsymbol{\delta}_j + 2\boldsymbol{\delta}_j'\mathbf{H}_{\mathbf{X}_{j,k}}\mathbf{e}_j + \mathbf{e}_j'\mathbf{H}_{\mathbf{X}_{j,k}}\mathbf{e}_j$

Firstly, following the proof of part (a), we get

$$\sup_{1\le j\le p_1}|\mathbf{e}_j'\mathbf{e}_j - p_2T\sigma_1^2| = O_p\{(Tp_2\log p_1)^{1/2}\}.$$

Next, we show, under conditions A1-A2, A4 and A6, that there holds following claims.

D1. $\sup_{1\le j\le p_1}\boldsymbol{\delta}_j'\boldsymbol{\delta}_j = O_p(T\|\boldsymbol{\Delta}\|_F^2 + \sqrt{T\log p_1}\|\boldsymbol{\Delta}\|_F^2)$

D2. $\sup_{1\le j,k\le p_1}|\mathbf{x}_k'\boldsymbol{\delta}_j| = O_p(T\sqrt{p_2}\|\boldsymbol{\Delta}\|_F)$.

D3. $\sup_{1\le j\le p_1}|\boldsymbol{\delta}_j'\mathbf{e}_j| = O_p(\|\boldsymbol{\Delta}\|_F\sqrt{Tp_2\log p_1})$

D4. $\sup_{1\le j\le p_1} \boldsymbol{\delta}_j' \mathbf{H}_{\mathbf{X}_{j,k}} \boldsymbol{\delta}_j = O_p(T\|\boldsymbol{\Delta}\|_F^2)$

D5. $\sup_{1\le j\le p_1} |\boldsymbol{\delta}_j' \mathbf{H}_{\mathbf{X}_{j,k}} \mathbf{e}_j| = O_p(\|\boldsymbol{\Delta}\|_F^2 \log p_1)$

D6. $\sup_{1\le j\le p_1} \mathbf{e}_j' \mathbf{H}_{\mathbf{X}_{j,k}} \mathbf{e}_j = O_p(\log p_1)$.

Proof of Claim D1.

$$\begin{aligned}
\boldsymbol{\delta}_j'\boldsymbol{\delta}_j &= \sum_{t=1}^{T} \mathbf{i}_j'\boldsymbol{\Delta}_t\boldsymbol{\Delta}_t'\mathbf{i}_j \\
&= \sum_{t=1}^{T} \mathbf{i}_j'\widetilde{\mathbf{Y}}_t\boldsymbol{\Delta}\boldsymbol{\Delta}'\widetilde{\mathbf{Y}}_t'\mathbf{i}_j \\
&= \sum_{t=1}^{T}\sum_{i=1}^{p_2}(\sum_{k=1}^{p_2}\widetilde{y}_{tjk}\Delta_{ki})(\sum_{l=1}^{p_2}\widetilde{y}_{tjl}\Delta_{li}) \\
&= \sum_{t=1}^{T}\sum_{i=1}^{p_2}\sum_{k=1}^{p_2}\sum_{l=1}^{p_2}\widetilde{y}_{tjk}\widetilde{y}_{tjl}\Delta_{ki}\Delta_{li}
\end{aligned}$$

where $\widetilde{\mathbf{Y}}_t = \mathbf{A}\mathbf{Y}_t = (\widetilde{y}_{tij}) \in \mathbb{R}^{p_1\times p_2}$. Since $\mathbf{A}$ is banded, under conditions A1 and A5, by Lemma S1, we have,

$$\sup_{1\le j\le p_1} \boldsymbol{\delta}_j'\boldsymbol{\delta}_j \le CT\|\boldsymbol{\Delta}\|_F^2 + O_p(\sqrt{T\log p_1})(\|\boldsymbol{\Delta}\|_F^2 + \sum_{i=1}^{p_2}\sum_{k=1}^{p_2}\sum_{l=1}^{p_2}|\Delta_{ki}\Delta_{li}|) \quad \text{(S2.4)}$$

The result follows by $\sum_{i=1}^{p_2}\sum_{k=1}^{p_2}\sum_{l=1}^{p_2}|\Delta_{ki}\Delta_{li}| \le C\|\boldsymbol{\Delta}\|_F^2$.

Proof of Claim D2.

$$\sup_{1\le j,k\le p_1} |\boldsymbol{\delta}_j'\mathbf{x}_k| \le \sup_{1\le j,k\le p_1} |\boldsymbol{\delta}_j'\mathbf{u}_k| + \sup_{1\le j,k\le p_1} |\boldsymbol{\delta}_j'(\mathbf{x}_k - \mathbf{u}_k)|.$$

where $\mathbf{u}_k$ is defined in Part (a).

First,

$$\begin{aligned}\boldsymbol{\delta}_j'\mathbf{u}_k = \sum_{t=1}^{T}\mathbf{i}_j'\boldsymbol{\Delta}_t\mathbf{u}_{k,t} &= \sum_{t=1}^{T}\mathbf{i}_j'\widetilde{\mathbf{Y}}_t\boldsymbol{\Delta}\mathbf{Q}_t'\mathbf{i}_k = \sum_{t=1}^{T}\sum_{l=1}^{p_2}\sum_{s=1}^{p_2}\widetilde{y}_{t,jl}\Delta_{ls}q_{t,sk} \\ &= \sum_{l=1}^{p_2}\sum_{s=1}^{p_2}\Delta_{ls}\sum_{t=1}^{T}\widetilde{y}_{t,jl}q_{t,sk}.\end{aligned}$$

By assumption A1, $\sup_{1\le j,k\le p_1}|\boldsymbol{\delta}_j'\mathbf{u}_k| = O_p(T\sqrt{p_2}\|\boldsymbol{\Delta}\|_F)+O_p(p_2\|\boldsymbol{\Delta}\|_F\sqrt{Tp_2\log p_1}) = O_p(T\sqrt{p_2}\|\boldsymbol{\Delta}\|_F)$.

Second,

$$\begin{aligned}|\boldsymbol{\delta}_j'(\mathbf{x}_k-\mathbf{u}_k)| &= |\sum_{t=1}^{T}\mathbf{i}_j'\boldsymbol{\Delta}_t(\mathbf{x}_{k,t}-\mathbf{u}_{k,t})| \\ &= |\sum_{t=1}^{T}\mathbf{i}_j'\widetilde{\mathbf{Y}}_t\boldsymbol{\Delta}\boldsymbol{\Delta}'\mathbf{Y}_t\mathbf{i}_k| \\ &= |\sum_{t=1}^{T}\sum_{l=1}^{p_2}\sum_{s=1}^{p_2}\sum_{i=1}^{p_2}\widetilde{y}_{t,jl}\Delta_{li}\Delta_{si}y_{sk,t}| \\ &= |\sum_{l=1}^{p_2}\sum_{s=1}^{p_2}\sum_{i=1}^{p_2}\Delta_{li}\Delta_{si}\sum_{t=1}^{T}\widetilde{y}_{t,jl}y_{sk,t}|.\end{aligned}$$

By assumption A1, $\sup_{1\le j,k\le p_1}|\boldsymbol{\delta}_j'(\mathbf{x}_k-\mathbf{u}_k)| = O_p(T\|\boldsymbol{\Delta}\|_F^2)+O_p(\|\boldsymbol{\Delta}\|_F^2\sqrt{T\log p_1})$.

Proof of Claim D3.

$$\begin{aligned}\boldsymbol{\delta}_j'\mathbf{e}_j = \sum_{t=1}^{T}\mathbf{i}_j'\boldsymbol{\Delta}_t\mathbf{e}_{j,t+1} &= \sum_{t=1}^{T}\mathbf{i}_j'\widetilde{\mathbf{Y}}_t\boldsymbol{\Delta}\mathbf{E}_{t+1}'\mathbf{i}_k = \sum_{t=1}^{T}\sum_{l=1}^{p_2}\sum_{s=1}^{p_2}\widetilde{y}_{t,jl}\Delta_{ls}e_{sk,t+1} \\ &= \sum_{l=1}^{p_2}\sum_{s=1}^{p_2}\Delta_{ls}\sum_{t=1}^{T}\widetilde{y}_{t,jl}e_{sk,t+1}.\end{aligned}$$

By assumption A1, $\sup_{1\le j,k\le p_1}\boldsymbol{\delta}_j'\mathbf{e}_j = O_p(\|\boldsymbol{\Delta}\|_F\sqrt{Tp_2\log p_1})$.

Proof of Claim D4. This claim follows by Lemma S2 and Claim D2.

Proof of Claim D5. This claim follows by Lemma S2, Claim D2 and Claim D3.

Proof of Claim D6. This claim follows by Lemma S2 and part(a).

Combing claims D1-D6, we get the conclusion. □

Lemma S4. *Let* $\mathbf{x}_{k,t}$ *be* k*-th column of* $\widehat{\mathbf{R}}_t$, $\mathbf{x}_k = [\mathbf{x}'_{k,1}, \cdots, \mathbf{x}'_{k,T}]'$ *and* $\mathbf{e}_j = [\mathbf{i}'_j\mathbf{E}'_2, \cdots, \mathbf{i}'_j\mathbf{E}'_{T+1}]'$, *for* $j, k = 1, 2, \cdots, p_2$. *Let* $\widehat{\mathbf{A}}^{(i-1)}$ *be the latest estimator of* $\mathbf{A}$ *in* ***Algorithm 1****, and define* $\eta = \|\widehat{\mathbf{A}}^{(i-1)} - \mathbf{A}\|_F^2$, *if* $\eta = O_p(\sqrt{\frac{\log p_2}{Tp_1}})$, *(a) Under conditions A1-A2, we have*

$$\sup_{1\le j,k\le p_2} |\mathbf{e}'_j\mathbf{x}_k| = O_p\{(Tp_1 \log p_2)^{1/2}\}.$$

(b) Following notations for model (2.6), under conditions A1-A2, A4, and A6, for any finite $k \ge k_2$, *we have,*

$$\sup_{1\le j\le p_2} |\frac{RSS_j(k)}{Tp_1\sigma_2^2} - 1| = O_p\{(T^{-1}p_1^{-1}\log p_2)^{1/2}\}.$$

Proof. The proof is similar to that of Lemma A.3, thus omitted. □

Lemma S5. *Let conditions A1-A2, A4, and A6 hold. Let* $\widehat{\mathbf{B}}^{(i-1)}$ *be the latest estimator of* $\mathbf{B}$ *in* ***Algorithm 1****, and* $\widehat{\mathbf{U}}_t(\mathbf{U}_t)$ *be any* $p_2 \times s(s \le p_1)$ *sub-matrices, consisting of finite number of columns of* $\widehat{\mathbf{Q}}'_t(\mathbf{Q}'_t)$, $\widehat{\Sigma} = T^{-1}p_2^{-1}\widehat{\mathbf{U}}'\widehat{\mathbf{U}}$, *and* $\Sigma = \mathbb{E}p_2^{-1}\mathbf{U}'_t\mathbf{U}_t$. *Define* $\eta = \|\widehat{\mathbf{B}}^{(i-1)} - \mathbf{B}\|_F^2$, *if* $\eta = O_p(p_2^{-1})$, *we have,*

$$\sup_{1\le j,k\le s} |\widehat{\Sigma}_{jk} - \Sigma_{jk}| = O_p\{\sqrt{T^{-1}p_2^{-1}\log p_1}\},$$

where $\widehat{\mathbf{U}} = (\widehat{\mathbf{U}}_1', \cdots, \widehat{\mathbf{U}}_T')$.

Proof. In the following, $\widehat{\mathbf{B}}^{(i-1)}$ is written as $\widehat{\mathbf{B}}$ to simplify notations. By the definition of $\widehat{\mathbf{U}}_t$ and $\mathbf{U}_t$, there exist $\mathbf{Y}_{1,t}' \in R^{p_2 \times s}$ consisting of the sub-columns of $\mathbf{Y}_t'$ such that $\widehat{\mathbf{U}}_t = \widehat{\mathbf{B}}\mathbf{Y}_{1,t}'$ and $\mathbf{U}_t = \mathbf{B}\mathbf{Y}_{1,t}'$. Hence,

$$\begin{aligned}
\frac{1}{Tp_2}\widehat{\mathbf{U}}'\widehat{\mathbf{U}} =& \frac{1}{Tp_2}\sum_{t=1}^{T}\widehat{\mathbf{U}}_t'\widehat{\mathbf{U}}_t \\
=& \frac{1}{Tp_2}\sum_{t=1}^{T}\mathbf{Y}_{1,t}\widehat{\mathbf{B}}'\widehat{\mathbf{B}}\mathbf{Y}_{1,t}' \\
=& \frac{1}{Tp_2}\sum_{t=1}^{T}\mathbf{Y}_{1,t}\mathbf{B}'\mathbf{B}\mathbf{Y}_{1,t}' + \frac{1}{Tp_2}\sum_{t=1}^{T}\mathbf{Y}_{1,t}\mathbf{B}'(\widehat{\mathbf{B}}-\mathbf{B})\mathbf{Y}_{1,t}' + \frac{1}{Tp_2}\sum_{t=1}^{T}\mathbf{Y}_{1,t}(\widehat{\mathbf{B}}-\mathbf{B})'\mathbf{B}\mathbf{Y}_{1,t}' \\
&+ \frac{1}{Tp_2}\sum_{t=1}^{T}\mathbf{Y}_{1,t}(\widehat{\mathbf{B}}-\mathbf{B})'(\widehat{\mathbf{B}}-\mathbf{B})\mathbf{Y}_{1,t}' \\
=& \Pi_1 + \Pi_2 + \Pi_3 + \Pi_4. \qquad \text{(S2.5)}
\end{aligned}$$

Now,

$$\sup_{j,k}|\widehat{\Sigma}_{jk} - \Sigma_{jk}| \leq \sup_{j,k}|\Pi_{jk,1} - \Sigma_{jk}| + \sup_{jk}|\Pi_{j,k,2}| + \sup_{j,k}|\Pi_{jk,3}| + \sup_{j,k}|\Pi_{jk,4}|. \qquad \text{(S2.6)}$$

Under the assumption that $\|\mathbf{B}\|_2$ is finite, we can show that each entry of $\mathbf{Y}_{1,t}\mathbf{B}\mathbf{B}'\mathbf{Y}_{1,t}'/p_2$ has finite variance. Under the $\alpha-$mixing condition in Assumption A1, by Lemma S1, we can show that

$$\sup_{j,k}|\Pi_{jk,1} - \Sigma_{jk}| = O_p(\sqrt{T^{-1}p_2^{-1}\log s}).$$

By a similar argument as above, we can show that

$$\sup_{j,k} |\Pi_{jk,i}| = O_p(\|\widehat{\mathbf{B}} - \mathbf{B}\|_F \sqrt{T^{-1} \log s}), i = 2, 3,$$

and

$$\sup_{j,k} |\Pi_{jk,4}| = O_p(\|\widehat{\mathbf{B}} - \mathbf{B}\|_F^2 \sqrt{T^{-1} \log s}).$$

Combing all these results, and the result follows. □

Now, we are ready to prove main results.

Proof of Proposition 1. Let $\boldsymbol{W}_{t-1} = \begin{pmatrix} \boldsymbol{G}_{t-1}^1 \\ \boldsymbol{G}_{t-1}^2 \end{pmatrix}$, where $\boldsymbol{G}_{t-1}^1 = (\mathbf{Y}_{t-1}\mathbf{B}') \otimes \mathbf{I}_{p_1} = (g_{ij,t-1}^1)$ is the first p_1^2 rows of $\boldsymbol{W}_{t-1}$, and $\boldsymbol{G}_{t-1}^2 = \mathbf{I}_{p_2} \otimes (\mathbf{Y}_{t-1}'\mathbf{A}') = (g_{ij,t-1}^2)$ is the last p_2^2 rows of $\boldsymbol{W}_{t-1}$. Following Chen et al. (2021), there exists a matrix $\mathbf{H}$ such that

$$\boldsymbol{H} \begin{pmatrix} \operatorname{vec}\left(\widehat{\boldsymbol{A}}_0 - \boldsymbol{A}\right) \\ \operatorname{vec}\left(\widehat{\boldsymbol{B}}_0' - \boldsymbol{B}'\right) \end{pmatrix} = \frac{1}{T} \sum_t \boldsymbol{W}_{t-1} \operatorname{vec}(\boldsymbol{E}_t) + o_P\left(T^{-1/2}\right).$$

Then

$$\begin{aligned} \|\widehat{\mathbf{A}}_0 - \mathbf{A}\|_F^2 + \|\widehat{\mathbf{B}}_0 - \mathbf{B}\|_F^2 &= \|\operatorname{vec}\left(\widehat{\mathbf{A}}_0 - \mathbf{A}\right)\|_2^2 + \|\operatorname{vec}\left(\widehat{\mathbf{B}}_0 - \mathbf{B}\right)\|_2^2 \\ &\leq \frac{1}{\lambda_h^2} \|\frac{1}{T} \sum_t \boldsymbol{G}_{t-1}^1 \operatorname{vec}(\mathbf{E}_t)\|_2^2 + \frac{1}{\lambda_h^2} \|\frac{1}{T} \sum_t \boldsymbol{G}_{t-1}^2 \operatorname{vec}(\mathbf{E}_t)\|_2^2 + o_p(\frac{1}{T}), \end{aligned}$$

where $1 < \lambda_h < \infty$ is the smallest eigenvalue of $\mathbf{H}$. Note that there are only p_2

non-zero elements in k-th row of $\boldsymbol{G}_{t-1}^1$, thus,

$$\begin{aligned}\|\frac{1}{T}\sum_t \boldsymbol{G}_{t-1}^1 \operatorname{vec}(\mathbf{E}_t)\|_2^2 &= \sum_{k=1}^{p_1^2}\|\frac{1}{T}\sum_t \mathbf{i}'_k \boldsymbol{G}_{t-1}^1 \operatorname{vec}(\mathbf{E}_t)\|_2^2 \\ &= \sum_{k=1}^{p_1^2}\|\frac{1}{T}\sum_t\sum_j g_{kj,t-1}^1 e_{jt}\|_2^2 \\ &\le \sum_{k=1}^{p_1^2}\sum_j\|\frac{1}{T}\sum_t g_{kj,t-1}^1 e_{jt}\|_2^2 = O_p(\frac{p_1^2 p_2}{T}),\end{aligned}$$

where $e_{j,t}$ is the j-th element of $\operatorname{vec}(\mathbf{E}_t)$.

Similarly, there are only p_1 non-zero elements in k-th row of $\boldsymbol{G}_{t-1}^2$, thus,

$$\|\frac{1}{T}\sum_t \boldsymbol{G}_{t-1}^2 \operatorname{vec}(\mathbf{E}_t)\|_2^2 = O_p(\frac{p_2^2 p_1}{T}).$$

This completes the proof. □

Proof of Theorem 3.1. We only prove the results for $\widehat{\mathbf{A}}^{(i)}$, since the proof of that for $\widehat{\mathbf{B}}^{(i)}$ is similar. First, we show $\|\widehat{\mathbf{A}}^{(i)} - \mathbf{A}\|_F^2 \to 0$.

By Theorem 2, we may just consider the convergence of $\widehat{\mathbf{A}}^{(i)}$ on event $\{\widehat{k}_1^{(i)} = k_1\}$. To simplify notations, we write $\widehat{\boldsymbol{\beta}}_{j,k_1}$ and $\boldsymbol{\beta}_{j,k_1}$ as $\widehat{\boldsymbol{\beta}}_j$ and $\boldsymbol{\beta}_j$, respectively. Now, according to model (2.5) or model (S2.9), we have,

$$\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j = (\mathbf{X}'_{j,k_1}\mathbf{X}_{j,k_1})^{-1}\mathbf{X}'_{j,k_1}\boldsymbol{f}_j.$$

Thus, by Lemma S2,

$$\begin{aligned}\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|_2^2 &\le \frac{C}{T^2p_2^2}\|\mathbf{X}'_{j,k_1}\boldsymbol{\delta}_j\|_2^2 + \frac{C}{T^2p_2^2}\|\mathbf{X}'_{j,k_1}\boldsymbol{e}_j\|_2^2 \\ &\le \frac{C}{T^2p_2^2}\|\mathbf{X}'_{j,k_1}\boldsymbol{\delta}_j\|_2^2 + \frac{C}{T^2p_2^2}\sup_{1\le j,k\le p_1}\|\mathbf{x}'_k\boldsymbol{e}_j\|_2^2,\end{aligned}$$

where $\mathbf{x}_k = [\mathbf{x}'_{k,1}, \cdots, \mathbf{x}'_{k,T}]'$ with $\mathbf{x}_{k,t}$ being k-th column of $\widehat{\mathbf{Q}}'_t$.

Let $\boldsymbol{\Delta} = \widehat{\mathbf{B}}^{(i-1)} - \mathbf{B}$, by Lemma S3(a), we have,

$$\begin{aligned}
\|\widehat{\mathbf{A}}^{(i)} - \mathbf{A}\|_F^2 &= \sum_{j=1}^{p_1} \|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|_2^2 = \frac{C}{T^2 p_2^2} \sum_{j=1}^{p_1} \|\mathbf{X}'_{j,k_1} \boldsymbol{\delta}_j\|_2^2 + O_p(\frac{p_1 \log p_1}{T p_2}) \\
&= \frac{C}{T^2 p_2^2} \| \sum_{t=1}^{T} \mathbf{X}_{j,t}^{k_1'} \boldsymbol{\Delta}'_t \|_F^2 + O_p(\frac{p_1 \log p_1}{T p_2}) \\
&= \frac{C}{T^2 p_2^2} \| \sum_{t=1}^{T} \mathbf{X}_{j,t}^{k_1'} \boldsymbol{\Delta} \mathbf{Y}'_t \mathbf{A}' \|_F^2 + O_p(\frac{p_1 \log p_1}{T p_2}) \\
&\overset{(i)}{=} \frac{C}{T^2 p_2^2} \| \sum_{t=1}^{T} (\mathbf{A}\mathbf{Y}_t) \otimes \mathbf{X}_{j,t}^{k_1'} \|_F^2 \|\text{vec}(\boldsymbol{\Delta})\|_2^2 + O_p(\frac{p_1 \log p_1}{T p_2}) \\
&\overset{(ii)}{=} C \|\boldsymbol{\Delta}\|_F^2 \| \frac{1}{T p_2} \sum_{t=1}^{T} (\mathbf{A}\mathbf{Y}_t) \otimes \mathbf{X}_{j,t}^{k_1'} \|_F^2 + O_p(\frac{p_1 \log p_1}{T p_2}) \\
&\overset{(iii)}{=} O_p(\|\boldsymbol{\Delta}\|_F^2 + \frac{p_1 \log p_1}{T p_2}).
\end{aligned}$$

where (i) follows by $\text{vec}(\mathbf{CZD}) = (\mathbf{D}'\mathbf{C})\text{vec}(\mathbf{Z})$; (ii) follows by $(\mathbf{CF}) \otimes (\mathbf{DG}) = (\mathbf{C} \otimes \mathbf{D})(\mathbf{F} \otimes \mathbf{G})$ for any matrix $\mathbf{C} \in \mathbb{R}^{m \times n}$, $\mathbf{D} \in \mathbb{R}^{p \times q}$, $\mathbf{F} \in \mathbb{R}^{n \times k}$, $\mathbf{G} \in \mathbb{R}^{q \times l}$, and $\mathbf{Z} \in \mathbb{R}^{n \times p}$, and (iii) follows by assumption A8. The conclusion is obtained by Proposition 1.

Similarly, we can prove that $\|\widehat{\mathbf{B}}^{(i)} - \mathbf{B}\|_F^2 = O_p(\|\boldsymbol{\Delta}\|_F^2)$.

Second, we show the convergence rate of $\|\widehat{\mathbf{A}}^{(i)} - \mathbf{A}\|_2^2$. Recall that for any matrix $\mathbf{M}$, it holds $\|\mathbf{M}\|_2^2 \leq \|\mathbf{M}\|_1 \|\mathbf{M}\|_\infty$. Hence,

$$\|\widehat{\mathbf{A}}^{(i)} - \mathbf{A}\|_2 \leq \sqrt{\|\widehat{\mathbf{A}}^{(i)} - \mathbf{A}\|_1 \|\widehat{\mathbf{A}}^{(i)} - \mathbf{A}\|_\infty} \leq (2k_1 + 1) \sup_{i,j} |\widehat{\beta}_{ij} - \beta_{ij}|,$$

where $\widehat{\beta}_{ij}$ and β_{ij} is the i-th element of $\widehat{\boldsymbol{\beta}}_j$ and $\boldsymbol{\beta}_j$, respectively.

By Lemma S2,

$$\begin{aligned} |\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j| &\leq \frac{C}{Tp_2}|\mathbf{X}'_{j,k_1}\boldsymbol{\delta}_j| + \frac{C}{Tp_2}|\mathbf{X}'_{j,k_1}\boldsymbol{e}_j| \\ &\leq \frac{C}{Tp_2}\sup_{1\leq j,k\leq p_1}|\mathbf{x}'_k\boldsymbol{\delta}_j| + \frac{C}{Tp_2}\sup_{1\leq j,k\leq p_1}|\mathbf{x}'_k\boldsymbol{e}_j|. \end{aligned}$$

Now, by Lemma S3(a), and Claim D2 in the proof of Lemma S3(b),

$$\sup_{i,j}|\widehat{\beta}_{ij} - \beta_{ij}| \leq |\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j| = O_p(\|\boldsymbol{\Delta}\|_F) + O_p(\sqrt{\frac{\log p_1}{Tp_2}}). \tag{S2.7}$$

Thus, $\|\widehat{\mathbf{A}}^{(i)} - \mathbf{A}\|_2 = O_p(\|\boldsymbol{\Delta}\|_F) + O_p(\sqrt{\frac{\log p_1}{Tp_2}}) = O_p(\|\boldsymbol{\Delta}\|_F)$ by Proposition 1. Similarly, we can prove that $\|\widehat{\mathbf{B}}^{(i)} - \mathbf{B}\|_2 = O_p(\|\boldsymbol{\Delta}\|_F)$

□

Proof of Theorem 3.2. We only prove the conclusions for $\widehat{\mathbf{A}}^{(i)}$ and $\widehat{k}_1^{(i)}$, since the proof of those for $\widehat{\mathbf{B}}^{(i)}$ and $\widehat{k}_2^{(i)}$ is similar.

Given the estimator $\widehat{\mathbf{B}}^{(i-1)}$ of $\mathbf{B}$, let $\widehat{\mathbf{Q}}_t = \mathbf{Y}_t\widehat{\mathbf{B}}^{(i-1)}$ and $\mathbf{Q}_t = \mathbf{Y}_t\mathbf{B}$, we have,

$$\mathbf{Y}_t = \mathbf{A}\widehat{\mathbf{Q}}_{t-1} + \mathbf{F}_t, \mathbf{F}_t = \mathbf{A}(\mathbf{Q}_{t-1} - \widehat{\mathbf{Q}}_{t-1}) + \mathbf{E}_t. \tag{S2.8}$$

Assuming $k = k_1$ is the true bandwidth, model (2.5) is written as

$$\mathbf{v}_j = \mathbf{X}_{j,k_1}\boldsymbol{\beta}_{j,k_1} + \boldsymbol{f}_j, j = 1, 2, \cdots, p_1, \tag{S2.9}$$

where $\boldsymbol{f}_j = [\mathbf{i}'_j\boldsymbol{\Delta}_1, \cdots, \mathbf{i}'_j\boldsymbol{\Delta}_T]' + \mathbf{e}_j = \boldsymbol{\delta}_j + \mathbf{e}_j$ by equation (S2.8), with $\boldsymbol{\Delta}_t = \mathbf{A}\mathbf{Y}_t(\mathbf{B} - \widehat{\mathbf{B}}^{(i-1)})$ for $t = 1, 2, \cdots, T$. In addition, $\text{RSS}_j(k_1) = \boldsymbol{f}'_j(\mathbf{I} - \mathbf{H}_{k_1})\boldsymbol{f}_j$ where $\mathbf{H}_{\mathbf{X}_{j,k}}$ is simply denoted as $\mathbf{H}_k$ for any $k > 0$ to simplify notations.

Note that $P(\widehat{k}_1^{(i)} \neq k_1) \leq P(\widehat{k}_1^{(i)} > k_1) + P(\widehat{k}_1^{(i)} < k_1)$, thus, we should prove that $P(\widehat{k}_1^{(i)} < k_1) \to 0$ and $P(\widehat{k}_1^{(i)} > k_1) \to 0$.

Firstly, we prove $P(\widehat{k}_1^{(i)} < k_1) \to 0$.

Denote $\widehat{k}_j = \arg\min_{1\leq k\leq K} \text{BIC}_j(k)$, and $\widehat{k}_1^{(i)} = \max_j \widehat{k}_j$. Note that $P(\widehat{k}_1^{(i)} < k_1) \leq P(\widehat{k}_j < k_1) \leq P(\min_{1\leq k<k_1} \text{BIC}_j(k) < \text{BIC}_j(k_1))$ for any $j \in \{1, 2, \cdots, p_1\}$. Thus, we only need to prove for any $j \in \{1, 2, \cdots, p_1\}$, $P(\min_{1\leq k<k_1} \text{BIC}_j(k) < \text{BIC}_j(k_1)) \to 0$.

For $k < k_1$, model (2.5) can be written as $\mathbf{v}_j = \mathbf{X}_{j,k}\boldsymbol{\beta}_{1j} + \mathbf{U}_j\boldsymbol{\beta}_{2j} + \boldsymbol{f}_j$ and $\mathbf{X}_{j,k_1} = [\mathbf{X}_{j,k}, \mathbf{U}_j]$ according to nesting property of banded matrix. Furthermore,

$$\begin{aligned}
\text{RSS}_j(k) &= \mathbf{v}_j'(\mathbf{I} - \mathbf{H}_k)\mathbf{v}_j \\
&= (\mathbf{X}_{j,k}\boldsymbol{\beta}_{1j} + \mathbf{U}_j\boldsymbol{\beta}_{2j} + \boldsymbol{f}_j)'(\mathbf{I} - \mathbf{H}_k)(\mathbf{X}_{j,k}\boldsymbol{\beta}_{1j} + \mathbf{U}_j\boldsymbol{\beta}_{2j} + \boldsymbol{f}_j) \\
&= \boldsymbol{f}_j'(\mathbf{I} - \mathbf{H}_k)\boldsymbol{f}_j + \boldsymbol{\beta}_{2j}'\mathbf{U}_j'(\mathbf{I} - \mathbf{H}_k)\mathbf{U}_j\boldsymbol{\beta}_{2j} + 2\boldsymbol{\beta}_{2j}'\mathbf{U}_j'(\mathbf{I} - \mathbf{H}_k)\boldsymbol{f}_j.
\end{aligned}$$

$$\begin{aligned}
\text{RSS}_j(k) - \text{RSS}_j(k_1) &= \boldsymbol{\beta}_{2j}'\mathbf{U}_j'(\mathbf{I} - \mathbf{H}_k)\mathbf{U}_j\boldsymbol{\beta}_{2j} + 2\boldsymbol{\beta}_{2j}'\mathbf{U}_j'(\mathbf{I} - \mathbf{H}_k)\boldsymbol{f}_j + \boldsymbol{f}_j'(\mathbf{H}_{k_1} - \mathbf{H}_k)\boldsymbol{f}_j \\
&= R_1 + R_2 + R_3.
\end{aligned}$$

By Lemma S3, there exists $\delta > 0$ such that $\text{RSS}_j(k_1) < (1+\delta)Tp_2\sigma_1^2$. In addition, by Lemma S2, there exists $r > \lambda_1\{(1+\delta)\sigma_1^2\}^{-1}$, such that $\lambda_{\min}\{\mathbf{U}_j'(\mathbf{I} - \mathbf{H}_k)\mathbf{U}_j\} > r(1+\delta)Tp_2\sigma_1^2$. Also, we have, $R_2 = CTp_2\|\widehat{\mathbf{B}}^{(i-1)} - \mathbf{B}\|_F + O_p(\|\widehat{\mathbf{B}}^{(i-1)} - \mathbf{B}\|_F\sqrt{Tp_2^2\log p_1})$, and $R_3 = o_p(1)$.

Hence, we have $\text{RSS}_j(k) - \text{RSS}_j(k_1) > r\text{RSS}_j(k_1)(a_{j,j-k_1}^2 + a_{j,j+k_1}^2)$. Thus,

$\log \text{RSS}_j(k) - \log \text{RSS}_j(k_1) > \log\{1+r(a^2_{j,j-k_1}+a^2_{j,j+k_1})\} > \min\{0.5r(a^2_{j,j-k_1}+ a^2_{j,j+k_1}), \log 2\}$. The second inequality follows from $\log(1+x) > \min\{0.5x, \log 2\}$ for any $x > 0$.

Thus, under conditions A5, it follows with probability going to 1, $\min_{1\le k<k_1} \text{BIC}_j(k) - \text{BIC}_j(k_1) > \min\{0.5r(a^2_{j,j-k_1} + a^2_{j,j+k_1}), \log 2\} - \frac{C_{M_2}}{M_2} k_1 \log(p_1 \vee M_2) > 0$. That is, we have $P(\widehat{k}_1^{(i)} \le k_1) \to 0$.

Second, we prove $P(\widehat{k}_1^{(i)} > k_1) \to 0$.

For $k > k_1$, model (2.5) can be written as $\mathbf{v}_j = \mathbf{X}_{j,k}\boldsymbol{\beta}_{j,k} + \boldsymbol{f}_j = \mathbf{X}_{j,k_1}\boldsymbol{\beta}_{1j} + \mathbf{U}_j\boldsymbol{\beta}_{2j} + \boldsymbol{f}_j$.

By some algebra, we have $\text{RSS}_j(k) = \text{RSS}_j(k_1) - \|\widetilde{\mathbf{U}}_j\widehat{\boldsymbol{\beta}}_{2j}\|^2$,where $\widetilde{\mathbf{U}}_j = (\mathbf{I} - \mathbf{H}_{k_1})\mathbf{U}_j$, and $\widehat{\boldsymbol{\beta}}_{2j} = (\widetilde{\mathbf{U}}_j'\widetilde{\mathbf{U}}_j)^{-1}\widetilde{\mathbf{U}}_j'\boldsymbol{f}_j$. Under condition A4, by Lemma S2, there exits $r_0, \delta > 0$, such that,

$$\begin{aligned} \text{RSS}_j(k_1) - \text{RSS}_j(k) &= \boldsymbol{f}_j'\widetilde{\mathbf{U}}_j(\widetilde{\mathbf{U}}_j'\widetilde{\mathbf{U}}_j)^{-1}\widetilde{\mathbf{U}}_j'\boldsymbol{f}_j \\ &\le r_0(1-\delta)\sigma_1^2(\tau_j(k) - \tau_j(k_1)) \sup_{i,s\le p_1} |T^{-1/2}p_2^{-1/2}\boldsymbol{f}_i'(\mathbf{I} - \mathbf{H}_{k_1})\mathbf{x}_s|^2. \end{aligned}$$

where $\mathbf{x}_s = [\mathbf{x}_{s,1}', \cdots, \mathbf{x}_{s,T}']'$ with $\mathbf{x}_{s,t}$ being the s-th column of $\widehat{\mathbf{Q}}_t$.

According to the definition of $\mathbf{X}_{j,t}^{k_1}$, there exist $\mathbf{Y}_{1,t}$ such that $\mathbf{X}_{j,t}^{k_1} = \mathbf{Y}_{1,t}(\widehat{\mathbf{B}}^{i-1})'$, and $\mathbf{X}_{j,k_1} = [\mathbf{X}_{j,1}^{k_1}; \cdots, \mathbf{X}_{j,T}^{k_1}]$. Accordingly, we define $\mathbf{U}_{j,t} = \mathbf{Y}_{1,t}\mathbf{B}'$. Let

$\widehat{\Sigma} = \frac{1}{Tp_2}\mathbf{X}'_{j,k_1}\mathbf{X}_{j,k_1}$, and $\Sigma = p_2^{-1}\mathbb{E}\mathbf{U}'_{j,t}\mathbf{U}_{j,t}$, we have,

$$\begin{aligned}\boldsymbol{f}_i'(\mathbf{I}-\mathbf{H}_{k_1})\mathbf{x}_s &= \boldsymbol{f}_i'(\mathbf{I}-\frac{1}{Tp_2}\mathbf{X}_{j,k_1}\Sigma^{-1}\mathbf{X}'_{j,k_1})\mathbf{x}_s + \frac{1}{Tp_2}\boldsymbol{f}_i'\mathbf{X}_{j,k_1}(\widehat{\Sigma}^{-1}-\Sigma^{-1})\mathbf{X}'_{j,k_1}\mathbf{x}_s \\ &= \Pi_1+\Pi_2.\end{aligned}$$

Similar to (A.2), we have,

$$\sup_{i,s\le p_1}\Pi_1 = O_p(Tp_2\|\widehat{\mathbf{B}}^{(i-1)}-\mathbf{B}\|_F) + O_p(\sqrt{Tp_2\log p_1}), \tag{S2.10}$$

and, by Lemma S5,

$$\sup_{i,s\le p_1}\Pi_2 = O_p\{(T\log p_1)^{1/2}p_2^{3/2}\|\widehat{\mathbf{B}}^{(i-1)}-\mathbf{B}\|_F\} + O_p(p_2\log p_1). \tag{S2.11}$$

Combing equation (S2.10) and equation (S2.11), we get $\sup_{i,s\le p_1}|T^{-1/2}p_2^{-1/2}\boldsymbol{f}_i'(\mathbf{I}-\mathbf{H}_{k_1})\mathbf{x}_s|^2 \le O_p\{\log(M_2\vee p_1)\}$. Thus,

$$\begin{aligned}\text{RSS}_j(k_1)-\text{RSS}_j(k) &\le C_{M_2}(1-\delta)\sigma_1^2(\tau_j(k)-\tau_j(k_1))\log(M_2\vee p_1) \\ &\le \text{RSS}_j(k)(\tau_j(k)-\tau_j(k_1))M_2^{-1}C_{M_2}\log(M_2\vee p_1),\end{aligned}$$

where $M_2 = p_2T$, and $C_{M_2} = \log\log(M_2)$, and the second inequality follows from part (b) of Lemma S3.

Finally, using the fact that $\log(1+x)\le x$ for any $x>0$, we have,

$$\begin{aligned}\text{BIC}_j(k)-\text{BIC}_j(k_1) &= \log\text{RSS}_j(k)-\log\text{RSS}_j(k_1)+\frac{C_{M_2}}{M_2}(\tau_j(k)-\tau_j(k_1))\log(M_2\vee p_1) \\ &\ge -(\text{RSS}_j(k_1)-\text{RSS}_j(k))/\text{RSS}_j(k)+\frac{C_{M_2}}{M_2}(\tau_j(k)-\tau_j(k_1))\log(M_2\vee p_1) \\ &\ge -(\tau_j(k)-\tau_j(k_1))\frac{C_{M_2}}{M_2}\log(M_2\vee p_1)+\frac{C_{M_2}}{M_2}(\tau_j(k)-\tau_j(k_1))\log(M_2\vee p_1) \\ &= 0.\end{aligned}$$

This proves $P(\widehat{k}_1^{(i)} > k_1) \to 0$. □

Proof of Theorem 3.3. We only prove the result for $\mathbf{A}^{(i)}$, since the proof of that for $\mathbf{B}^{(i)}$ is similar. Recall that, given $\mathbf{B}^{(i-1)}$, $\mathbf{A}^{(i)}$ is the minimizer of

$$\frac{1}{T}\sum_{t=2}^{T+1}\|\text{vec}(\mathbf{Y}_t) - (\mathbf{B}^{(i-1)}\mathbf{Y}_{t-1}')\otimes\mathbf{I}_{p_1})\text{vec}(\mathbf{A})\|_2^2 + \lambda\|\text{vec}(\mathbf{A})\|_1. \quad \text{(S2.12)}$$

Writing $\mathbf{u}_t = \text{vec}(\mathbf{Y}_t), \widehat{\mathbf{Z}}_{t-1} = (\mathbf{B}^{(i-1)}\mathbf{Y}_{t-1}')\otimes\mathbf{I}_{p_1}, \mathbf{Z}_{t-1} = (\mathbf{B}\mathbf{Y}_{t-1}')\otimes\mathbf{I}_{p_1}, \boldsymbol{\beta} = \text{vec}(\mathbf{A})$, it is simplified as

$$\frac{1}{T}\sum_{t=2}^{T+1}\|\mathbf{u}_t - \widehat{\mathbf{Z}}_{t-1}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1. \quad \text{(S2.13)}$$

Define $\boldsymbol{\beta}^{(i)} = \text{vec}(\mathbf{A}^{(i)})$, by the basic inequality in Lemma 6.1 of Buhlmann and Van De Geer (2011),

$$\frac{1}{T}\sum_{t=2}^{T+1}\|\mathbf{u}_t - \widehat{\mathbf{Z}}_{t-1}\boldsymbol{\beta}^{(i)}\|_2^2 + \lambda\|\boldsymbol{\beta}^{(i)}\|_1 \le \frac{1}{T}\sum_{t=2}^{T+1}\|\mathbf{u}_t - \widehat{\mathbf{Z}}_{t-1}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1. \quad \text{(S2.14)}$$

which imply

$$\frac{1}{T}\sum_{t=2}^{T+1}\|\widehat{\mathbf{Z}}_{t-1}\Delta_i\|_2^2 \le \frac{2}{T}\sum_{t=2}^{T+1}\Delta_i'\widehat{\mathbf{Z}}_{t-1}'((\mathbf{Z}_{t-1} - \widehat{\mathbf{Z}}_{t-1})\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t) + \lambda\|\Delta_i\|_1, \quad \text{(S2.15)}$$

where $\boldsymbol{\varepsilon}_t = \text{vec}(\mathbf{E}_t)$ and $\Delta_i = \boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}$.

First, by assumption A1-2 and A6, $\Pi_1 = \frac{2}{T}\sum_{t=2}^{T+1}(\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta})'\widehat{\mathbf{Z}}_{t-1}'(\mathbf{Z}_{t-1} - \widehat{\mathbf{Z}}_{t-1})\boldsymbol{\beta} = O_p(\|\boldsymbol{\beta} - \boldsymbol{\beta}^{(i)}\|_1\|\mathbf{B} - \mathbf{B}^{(i-1)}\|_2) \le O_p(\eta^{1/2}\|\boldsymbol{\beta} - \boldsymbol{\beta}^{(i)}\|_1)$, where $\eta = \|\mathbf{B} - \mathbf{B}^{(i-1)}\|_F^2$.

Second, by assumption A1-2 and A6, following Lemma S1,

$$\Pi_2 = \frac{2}{T}\sum_{t=2}^{T+1}(\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta})'\widehat{\mathbf{Z}}_{t-1}'\boldsymbol{\varepsilon}_t \le \|\boldsymbol{\beta} - \boldsymbol{\beta}^{(i)}\|_1 \sup_{1\le j\le p_2}\frac{2}{T}\sum_{t=2}^{T+1}\boldsymbol{\varepsilon}_t'\widehat{\mathbf{Z}}_{t-1}^j = O_p(\|\boldsymbol{\beta} - \boldsymbol{\beta}^{(i)}\|_1\sqrt{\frac{p_2\log p_2}{T}}),$$

where $\widehat{\mathbf{Z}}_{t-1}^{j}$ is j-th column of $\widehat{\mathbf{Z}}_{t-1}$.

Thus, for $\lambda_C > C\sqrt{\frac{p_1 p_2^2}{T}}$, by assumption A7 and (S2.15),

$$\kappa\|\Delta_i\|_2^2 \leq 3\lambda_C\|\Delta_i\|_1 = 3\lambda_C(\|\Delta_{i,S_0}\|_1+\|\Delta_{i,S_0^c}\|_1) \leq 12\lambda_C\|\Delta_{i,S_0}\|_1 \leq 12\lambda_C\sqrt{s_0}\|\Delta_i\|_2,$$

where $s_0 = |S_0|$. That is, $\|\mathbf{A}^{(i)} - \mathbf{A}\|_2 \leq \|\mathbf{A}^{(i)} - \mathbf{A}\|_F \leq 12\lambda_C\sqrt{s_0}/\kappa$, this completes the proof. □

Bibliography

Chen, R., Xiao, H., and Yang, D. (2021). Autoregressive models for matrix-valued time series. *Journal of Econometrics*, **222(1)**, 539–560.

Gao, Z. and Tsay, R. S. (2022). Modeling high-dimensional time series: a factor model with dynamically dependent factors and diverging eigenvalues. *Journal of the American Statistical Association*, **117(539)**, 1398-1414.

Guo, S., Wang, Y. and Yao, Q. (2016). High dimensional and banded vector autoregression. *Biometrika,* **103(4)**, 889–903.

Meinshausen N. and Buhlmann P. (2010). Stability selection. *Journal of the Royal Statistical Society, Series B*, **72**, 414-473.