Estimation and Selection of Spatial Weight Matrix in a Spatial Lag Model

Supplementary materials

Clifford Lam*[1] and Pedro CL Souza†[2]

[1]Department of Statistics, London School of Economics and Political Science
[2]Department of Economics, University of Warwick

*Clifford Lam is Associate Professor, Department of Statistics, London School of Economics. Email: C.Lam2@lse.ac.uk

†Pedro CL Souza is Assistant Professor, Department of Economics, University of Warwick.

1 Simulations

We conduct the following four additional Monte-Carlo exercises to demonstrate the robustness of the estimator in alternative scenarios. We consider cases with lower sparsity, time-varying $\mathbf{A}^*$, and explore how the correlations between covariates $\mathbf{X}_t$, instruments $\mathbf{B}_t$ and disturbance term $\boldsymbol{\epsilon}_t$ affect results in Table 1. For that, we focus on the $N = 25$, $T = 100$ and "no expert knowledge" case. Results for various N, T and models are available upon request.

We specifically implemented the following four cases, which we present along with the findings.

(a) Increase the number of non-zero elements of $\mathbf{A}^*$ from 5% to 25%. Results can be found in Table 1. We find that for moderate levels of sparsity the performance of the estimator is maintained, at least up to 10% sparse. This corresponds to 60 non-zero paramenters with $N = 25$. As the number of non-zero elements increases (and sparsity decreases), sensitivity is reduced as parameters decrease and are shrunk to zero, while specificity is still above 90%.

(b) Introduce time-varying $\mathbf{A}^*$, in the following way. A proportion $(1 - p)$ of non-zero elements of $\mathbf{A}^*$ are taken as fixed over time, and a proportion p is reshuffled at every period. We consider $p \in \{0\%, 10\%, 20\%, 30\%, 40\%\}$. Results are in Table 2. Specificity and sensitivity performance indicadors are calculated with respect to the stable portion of $\mathbf{A}^*$. We find that performance is robust even if $\mathbf{A}^*$ is moderately changing over time.

(c) We increase the correlation between $\mathbf{X}_t$ and $\boldsymbol{\epsilon}_t$. Baseline data is generated according to $X_t = \nu_t + c \cdot \boldsymbol{\epsilon}_t$ where, $c = 0.5$, ν_t and $\boldsymbol{\epsilon}_t$ are independent and drawn from a standard normal distribution, which corresponds to a correlation of .447 between $\mathbf{X}_t$ and $\boldsymbol{\epsilon}_t$. We take $c = \{0.625, 0.75, 0.875, 1\}$, respectively corresponding to correlations of .53, .6, .658 and $\frac{1}{\sqrt{2}} = .707$. We find that performance decreases only slightly as c increases.

(d) We descrease the correlation between $\mathbf{X}_t$ and $\mathbf{B}_t$. This is the case where instruments are allowed to be weaker. In baseline simulations, we constructed the instruments as $\mathbf{B}_t = \mathbf{X}_t + \nu_t$, where $\mathbf{X}_t$ and ν_t are independent standard normals. So the baseline covariance between $\mathbf{X}_t$ and $\mathbf{B}_t$ is $\frac{1}{\sqrt{2}}$. We then multiply this correlation by 0.8, 0.6, 0.4 and 0.2. As expected, the performance deteriorates as the correlation decreases. Yet, with correlation $0.2\frac{1}{\sqrt{2}} \approx 0.141$, we obtain 99% of specificity and 68% of sensitivity, although bias in $\boldsymbol{\beta}$ increases.

Table 1: Sparsity

	$s = 5\%$	10%	15%	20%	25%
$\mathbf{A}^*$ Specificity	.998	.984	.958	.929	.902
	(.002)	(.007)	(.009)	(.013)	(.014)
$\mathbf{A}^*$ Sensitivity	1.000	.964	.837	.693	.562
	(.004)	(.026)	(.044)	(.047)	(.046)
$\mathbf{A}^*$ bias	-.021	-.003	-.006	-.008	-.009
	(.000)	(.001)	(.001)	(.001)	(.001)
Lasso L1	.033	.020	.024	.027	.030
	(.002)	(.001)	(.001)	(.002)	(.002)
AdaLasso L1	.022	.006	.010	.014	.017
	(.000)	(.001)	(.001)	(.001)	(.001)
Sparsity	.948	.889	.839	.805	.786
	(.002)	(.006)	(.009)	(.013)	(.015)
$\boldsymbol{\beta}$ bias	.015	.015	.016	.017	.017
	(.008)	(.008)	(.008)	(.009)	(.009)
$\boldsymbol{\delta}^*$ Specificity	1.000	1.000	1.000	1.000	1.000
	(.000)	(.000)	(.000)	(.000)	(.000)
$\boldsymbol{\delta}^*$ Sensitivity	-	-	-	-	-
$\boldsymbol{\delta}^*$ Bias	.000	.000	.000	.000	.000
	(.000)	(.000)	(.000)	(.000)	(.000)

Notes: Simulated results for $N = 25$, $T = 100$, "No knowledge case"and 1,000 iterations. Sparsity s of $\mathbf{A}^*$ varies from 5% to 25%. Each column refers to the results for a given sparsity of $\mathbf{A}^*$. Specificity (Sensitivity) refers to the proportion of true zeros (non-zeros) that are estimated as zeros (non-zeros). Lasso L1 and AdaLasso L1 refer to the L_1 norm of the vectorized sparse deviation matrix of the LASSO and adaptive LASSO steps, respectively. Standard error across iterations Penalization parameters are chosen by BIC.

Table 2: Time-varying $\mathbf{A}^*$

	$p = 0\%$	10%	20%	30%	40%
$\mathbf{A}^*$ Specificity	.998	.977	.966	.960	.947
	(.002)	(.007)	(.008)	(.009)	(.010)
$\mathbf{A}^*$ Sensitivity	1.000	.999	.999	.998	.999
	(.004)	(.005)	(.007)	(.010)	(.009)
$\mathbf{A}^*$ bias	-.021	-.005	-.005	-.005	-.003
	(.000)	(.000)	(.001)	(.001)	(.001)
Lasso L1	.033	.025	.028	.030	.031
	(.002)	(.001)	(.002)	(.002)	(.002)
AdaLasso L1	.022	.006	.008	.008	.007
	.000	(.001)	(.001)	(.001)	(.001)
Sparsity	.948	.933	.927	.926	.919
	(.002)	(.006)	(.008)	(.009)	(.009)
$\boldsymbol{\beta}$ bias	.015	.016	.017	.019	.018
	(.008)	(.008)	(.010)	(.010)	(.011)
$\boldsymbol{\delta}^*$ Specificity	1.000	1.000	1.000	1.000	1.000
	(.000)	(.000)	(.000)	(.000)	(.000)
$\boldsymbol{\delta}^*$ Sensitivity	-	-	-	-	-
$\boldsymbol{\delta}^*$ Bias	.000	.000	.000	.000	.000
	(.000)	(.000)	(.000)	(.000)	(.000)

Notes: Simulated results for $N = 25$, $T = 100$, "No knowledge case"and 1,000 iterations. True $\mathbf{A}^*$ varies over time. A proportion $(1 - p)$ of non-zero elements of $\mathbf{A}^*$ are taken as fixed over time, and a proportion p is reshuffled at every period. Specificity (Sensitivity) refers to the proportion of true zeros (non-zeros) that are estimated as zeros (non-zeros), and refer to the time-invariant portion of $\mathbf{A}^*$. Lasso L1 and AdaLasso L1 refer to the L_1 norm of the vectorized sparse deviation matrix of the LASSO and adaptive LASSO steps, respectively. Standard error across iterations Penalization parameters are chosen by BIC.

Table 3: Correlation between $\mathbf{X}_t$ and $\boldsymbol{\epsilon}_t$

	$c = .5$	.625	.75	.875	1
$\mathbf{A}^*$ Specificity	.998	.987	.986	.986	.986
	(.002)	(.008)	(.009)	(.008)	(.008)
$\mathbf{A}^*$ Sensitivity	1.000	.889	.893	.899	.903
	(.004)	(.050)	(.055)	(.052)	(.053)
$\mathbf{A}^*$ bias	-.021	-.003	-.003	-.003	-.003
	(.000)	(.001)	(.001)	(.001)	(.001)
Lasso L1	.033	.027	.027	.027	.027
	(.002)	(.003)	(.003)	(.003)	(.003)
AdaLasso L1	.022	.006	.006	.006	.006
	(.000)	(.001)	(.001)	(.001)	(.001)
Sparsity	.948	.943	.942	.941	.941
	(.002)	(.008)	(.009)	(.009)	(.009)
$\boldsymbol{\beta}$ bias	.015	.042	.042	.042	.044
	(.008)	(.023)	(.023)	(.022)	(.022)
$\boldsymbol{\delta}^*$ Specificity	1.000	1.000	1.000	1.000	1.000
	(.000)	(.000)	(.000)	(.000)	(.000)
$\boldsymbol{\delta}^*$ Sensitivity	-	-	-	-	-
$\boldsymbol{\delta}^*$ Bias	.000	.000	.000	.000	.000
	(.000)	(.000)	(.000)	(.000)	(.000)

Notes: Simulated results for $N = 25$, $T = 100$, "No knowledge case"and 1,000 iterations. Each column j of the matrix of covariates $\mathbf{X}_t$ is generated as $\mathbf{X}_{jt} = \nu_t + c \cdot \boldsymbol{\epsilon}_t$. Columns of the table refers to different values of c. Specificity (Sensitivity) refers to the proportion of true zeros (non-zeros) that are estimated as zeros (non-zeros). Lasso L1 and AdaLasso L1 refer to the L_1 norm of the vectorized sparse deviation matrix of the LASSO and adaptive LASSO steps, respectively. Standard error across iterations Penalization parameters are chosen by BIC.

Table 4: Correlation between $\mathbf{X}_t$ and $\mathbf{B}_t$

	$\sigma_{xz} = \frac{1}{\sqrt{2}}$	$0.8\frac{1}{\sqrt{2}}$	$0.6\frac{1}{\sqrt{2}}$	$0.4\frac{1}{\sqrt{2}}$	$0.2\frac{1}{\sqrt{2}}$
$\mathbf{A}^*$ Specificity	.998	.991	.990	.990	.990
	(.002)	(.004)	(.004)	(.004)	(.004)
$\mathbf{A}^*$ Sensitivity	1.000	.826	.777	.731	.687
	(.004)	(.059)	(.065)	(.065)	(.076)
$\mathbf{A}^*$ bias	-.021	-.005	-.006	-.007	-.008
	(.000)	(.001)	(.001)	(.002)	(.002)
Lasso L1	.033	.026	.027	.027	.028
	(.002)	(.002)	(.002)	(.002)	(.002)
AdaLasso L1	.022	.008	.009	.010	.011
	(.000)	(.001)	(.001)	(.001)	(.002)
Sparsity	.948	.950	.952	.954	.957
	(.002)	(.005)	(.005)	(.005)	(.005)
$\boldsymbol{\beta}$ bias	.015	.052	.067	.107	.219
	(.008)	(.028)	(.038)	(.060)	(.127)
$\boldsymbol{\delta}^*$ Specificity	1.000	1.000	1.000	1.000	1.000
	(.000)	(.000)	(.000)	(.000)	(.000)
$\boldsymbol{\delta}^*$ Sensitivity	-	-	-	-	-
$\boldsymbol{\delta}^*$ Bias	.000	.000	.000	.000	.000
	(.000)	(.000)	(.000)	(.000)	(.000)

Notes: Simulated results under various combinations of N and T for 1,000 iterations. "No knowledge case" refers to the "No expert knowledge case", where expert matrices are not used in the estimated model, and the true network is defined by the sparse deviation only. In the "Partial knowledge" case, the true matrix is a combination of two expert matrices and a sparse deviation. There are no sparse deviations in true matrix of the "Full knowledge" case, but it is included in the estimated model. Specificity (Sensitivity) refers to the proportion of true zeros (non-zeros) that are estimated as zeros (non-zeros). Lasso L1 and AdaLasso L1 refer to the L_1 norm of the vectorized sparse deviation matrix of the LASSO and adaptive LASSO steps, respectively. Standard error across iterations Penalization parameters are chosen by BIC.

2 Proof of Theorem S.1, 1, 2, 3, 4, 5 and 6.

For proving Theorem S.1 below, we present an important Lemma which is a combination of Theorems 2(ii) and 2(iii) of Liu et al. (2013).

Lemma 1 *For a zero mean time series process $\mathbf{x}_t = \mathbf{f}(\mathcal{F}_t)$ defined in (3.11) with dependence measure $\theta^x_{t,d,i}$ defined in (3.12), assume $\Theta^x_{m,a} \le Cm^{-\alpha}$ as in Assumption M4. Then there exists constants C_1, C_2 and C_3 independent of v, T and the index j such that*

$$P\left(\left|\frac{1}{T}\sum_{t=1}^{T} x_{jt}\right| > v\right) \le \frac{C_1 T^{w(\frac{1}{2}-\widetilde{\alpha})}}{(Tv)^w} + C_2 \exp(-C_3 T^{\widetilde{\beta}} v^2),$$

where $\widetilde{\alpha} = \alpha \wedge (1/2 - 1/w)$, and $\widetilde{\beta} = (3 + 2\widetilde{\alpha}w)/(1 + w)$.

Furthermore, assume another zero mean time series process $\{\mathbf{z}_t\}$ (can be the same process $\{\mathbf{x}_t\}$) with both $\Theta^x_{m,2w}, \Theta^z_{m,2w} \le Cm^{-\alpha}$, as in Assumption M4. Then provided $\max_j \left\|x_{jt}\right\|_{2w}$, $\max_j \left\|z_{jt}\right\|_{2w} \le c_0 < \infty$ where c_0 is a constant, the above Nagaev-type inequality holds for the product process $\{x_{jt}z_{t\ell} - E(x_{jt}z_{t\ell})\}$.

This lemma concerns with the tail probability of the average of a general time series process as defined in (3.11), and is an important foundation for all the theorems in Section 3.4. Note that if $\alpha > 1/2 - 1/w$, then $w(1/2 - \widetilde{\alpha}) = \widetilde{\beta} = 1$, simplifying the form of the inequality. This is what we assumed in the theorems presented in Section 3.4. However, this can be relaxed at the expense of more complicated rates in those theorems, which we chose not to pursue for the sake of simplicity in presentation.

To be able to present the proofs smoothly, we present another lemma first.

Lemma 2 *For any $N \times N$ matrix $\mathbf{H} = (\mathbf{h}_1, \ldots, \mathbf{h}_N)^{\mathrm{T}}$ and any $N \times K$ matrix $\mathbf{M}$, define*

$$\mathbf{V}_{\mathbf{H}} = \begin{pmatrix} \mathbf{I}_K \otimes \mathbf{h}_1 \\ \vdots \\ \mathbf{I}_K \otimes \mathbf{h}_N \end{pmatrix}.$$

Then we have

$$\mathbf{HM} = (\mathbf{I}_N \otimes \mathrm{vec}^{\mathrm{T}}(\mathbf{M}))\mathbf{V}_{\mathbf{H}}.$$

Proof of Lemma 2. It is straight forward to verify that the (i, j)th entry of the RHS is indeed the same as LHS. □

Denote $B_{t,ij}$ and $X_{t,ij}$ the (i, j) entry of $\mathbf{B}_t$ and $\mathbf{X}_t$ respectively, and define $\mathcal{M} = \bigcap_{i=1}^{7} \mathcal{A}_i$, where

$$\begin{aligned}
\mathcal{A}_1 &= \left\{ \max_{1\le i,k\le N} \max_{1\le j,\ell\le K} \left| \frac{1}{T}\sum_{t=1}^{T}[B_{t,ij}X_{t,k\ell} - E(B_{t,ij}X_{t,k\ell})] \right| < c_T \right\}, \\
\mathcal{A}_2 &= \left\{ \max_{1\le i,k\le N} \max_{1\le j\le K} \left| \frac{1}{T}\sum_{t=1}^{T} B_{t,ij}\epsilon_{kt} \right| < c_T \right\}, \\
\mathcal{A}_3 &= \left\{ \max_{1\le k\le K} \left| \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{N} B_{t,sk}\epsilon_{st} \right| < c_T N^{\frac{1}{2}+\frac{1}{2w}} \right\}, \\
\mathcal{A}_4 &= \left\{ \max_{1\le i\le N} \max_{1\le j\le K} \left| \bar{B}_{\cdot,ij} - E(B_{t,ij}) \right| < c_T \right\}, \\
\mathcal{A}_5 &= \left\{ \max_{1\le j\le N} |\bar{\epsilon}_{j,\cdot}| < c_T \right\}, \\
\mathcal{A}_6 &= \left\{ \max_{1\le i\le N} \max_{1\le j\le K} |\bar{X}_{\cdot,ij}| < c_T \right\}, \\
\mathcal{A}_7 &= \left\{ \max_{1\le k\le K} \left| \sum_{s=1}^{N} \bar{B}_{\cdot,sk}\bar{\epsilon}_{s,\cdot} \right| < 2^{1/2} c_T N^{1/2} \log^{1/2}(T\vee N) S_\epsilon(\max_{i,j}|E(B_{t,ij})| + c_T) \right\},
\end{aligned} \tag{S.1}$$

with $\bar{B}_{\cdot,sk} = T^{-1}\sum_{t=1}^{T} B_{t,sk}$, $\bar{X}_{\cdot,ij} = T^{-1}\sum_{t=1}^{T} X_{t,ij}$ and $\bar{\epsilon}_{s,\cdot} = T^{-1}\sum_{t=1}^{T}\epsilon_{st}$.

Our theorems presented in the paper are actually describing properties of estimators on the set $\mathcal{M} = \bigcap_{i=1}^{7}\mathcal{A}_i$. It turns out that $P(\mathcal{M}) \to 1$ as $T, N \to \infty$, as shown in the following theorem. Hence in the proof of the remaining theorems, it is sufficient to prove the corresponding properties on the set $\mathcal{M}$.

Theorem S.1 *Let all the assumptions in Section 3.3 hold (M2, R5, R8 or M2', R5', R8'). Suppose $\alpha > 1/2 - 1/w$ in Assumption M4, and suppose for the application of the Nagaev-type inequality in Lemma 1 for the processes in $\mathcal{A}_1$ to $\mathcal{A}_6$, the constants C_1, C_2 and C_3 are the same. Then with $d \ge \sqrt{3/C_3}$ where d is the constant defined in $c_T = dT^{-1/2}\log^{1/2}(T\vee N)$, we have*

$$P(\mathcal{M}) \ge 1 - 8C_1K^2\left(\frac{C_3}{3}\right)^{w/2}\frac{N^2}{T^{w/2-1}\log^{w/2}(T\vee N)} - \frac{8C_2K^2N^2}{T^3\vee N^3} - \frac{2K}{T\vee N}.$$

It approaches 1 if we assume further that $N = o(T^{w/4-1/2}\log^{w/4}(T))$.

With Assumption M3, we can show that for any fixed $w > 0$, we have $\|B_{t,jk}\|_{2w}$, $\|X_{t,jk}\|_{2w}$ and $\|\epsilon_{jt}\|_{2w} < \infty$, so that Lemma 1 in the supplementary material can be applied, allowing the probability bound in the above theorem to hold. And, there are many examples with $\Theta_{m,2w} \le Cm^{-\alpha}$ where only the constant C is dependent on w. See for example the stationary linear process Example 2.2 in Chen et al. (2013). Therefore, we can set w to be large enough from the beginning so that $N = o(T^{w/4-1/2}\log^{w/4}(T))$ is satisfied, ensuring $P(\mathcal{M}) \to 1$. Note also that we can actually allow $\alpha \le 1/2 - 1/w$ at the expense of more complicated rates and longer proofs of the theorems presented in the paper.

Proof of Theorem S.1. Our aim is to apply Lemma 1 on the processes defined in the sets $\mathcal{A}_1$ to $\mathcal{A}_7$ (see Section 3.4 for their definitions). To this end, we first show that with tail Assumption

M3, a process $\{\mathbf{z}_t\}$ has $\max_j \|z_{jt}\|_{2w} < \infty$ for any $w > 0$. Indeed, by the Fubini's Theorem,

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

so that $\|z_{jt}\|_{2w} \le \mu_{2w} < \infty$ for any $w > 0$. Together with Assumption M4, Lemma 1 can be applied for the processes $\{B_{t,ij}X_{t,k\ell} - E(B_{t,ij}X_{t,k\ell})\}$, $\{B_{t,ij}\epsilon_{t,k}\}$, $\{B_{t,ij} - \mu_{b,ij}\}$, $\{\epsilon_{jt}\}$ and $\{X_{t,ij}\}$ in the sets $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_4, \mathcal{A}_5$ and $\mathcal{A}_6$ respectively. Since we assumed $\alpha > 1/2 - 1/w$, we have $w(1/2 - \widetilde{\alpha}) = \widetilde{\beta} = 1$ in Lemma 1. By the union sum inequality, we then have

$$\begin{aligned} P(\mathcal{A}_1^c) &\le \sum_{\substack{1\le i,k\le N\\ 1\le j,\ell\le K}} P\left(\left|T^{-1}\sum_{t=1}^T B_{t,ij}X_{t,k\ell} - E(B_{t,ij}X_{t,k\ell})\right| \ge c_T\right) \\ &\le N^2K^2\left(\frac{C_1T}{(Tc_T)^w} + C_2\exp(-C_3Tc_T^2)\right) \\ &\le C_1K^2\left(\frac{C_3}{3}\right)^{w/2}\frac{N^2}{T^{w/2-1}\log^{w/2}(T\vee N)} + \frac{C_2K^2N^2}{T^3\vee N^3}. \end{aligned} \tag{S.3}$$

Similarly, we have

$$\begin{aligned} P(\mathcal{A}_2^c) &\le C_1K\left(\frac{C_3}{3}\right)^{w/2}\frac{N^2}{T^{w/2-1}\log^{w/2}(T\vee N)} + \frac{C_2KN^2}{T^3\vee N^3}, \\ P(\mathcal{A}_4^c) &\le C_1K\left(\frac{C_3}{3}\right)^{w/2}\frac{N}{T^{w/2-1}\log^{w/2}(T\vee N)} + \frac{C_2KN}{T^3\vee N^3}, \\ P(\mathcal{A}_5^c) &\le C_1\left(\frac{C_3}{3}\right)^{w/2}\frac{N}{T^{w/2-1}\log^{w/2}(T\vee N)} + \frac{C_2N}{T^3\vee N^3}, \\ P(\mathcal{A}_6^c) &\le C_1K\left(\frac{C_3}{3}\right)^{w/2}\frac{N}{T^{w/2-1}\log^{w/2}(T\vee N)} + \frac{C_2KN}{T^3\vee N^3}. \end{aligned} \tag{S.4}$$

To find an upper bound for $P(\mathcal{A}_3^c)$, define $\mathbf{B}_{t,k}$ to be the kth column of $\mathbf{B}_t$. If we can show that

$$\max_{1\le k\le K}\left\|N^{-\frac{1}{2}-\frac{1}{2w}}\boldsymbol{\epsilon}_t^{\mathrm{T}}\mathbf{B}_{t,k}\right\|_{2w} < \infty, \tag{S.5}$$

$$\Theta_{m,2w} = \sum_{t=m}^\infty \max_{1\le k\le K}\left\|N^{-\frac{1}{2}-\frac{1}{2w}}(\boldsymbol{\epsilon}_t^{\mathrm{T}}\mathbf{B}_{t,k} - \boldsymbol{\epsilon}_t'^{\mathrm{T}}\mathbf{B}'_{t,k})\right\|_{2w} \le am^{-\alpha}, \tag{S.6}$$

for some $a > 0$ and all $m \ge 1$, then we can apply Lemma 1 for $\mathcal{A}_3$ to obtain

$$\begin{aligned} P(\mathcal{A}_3^c) &\le \sum_{k=1}^K P\left(\left|T^{-1}\sum_{t=1}^T N^{-\frac{1}{2}-\frac{1}{2w}}\boldsymbol{\epsilon}_t^{\mathrm{T}}\mathbf{B}_{t,k}\right| \ge c_T\right) \\ &\le C_1\left(\frac{C_3}{3}\right)^{w/2}\frac{K}{T^{w/2-1}\log^{w/2}(T\vee N)} + \frac{C_2K}{T^3\vee N^3}. \end{aligned} \tag{S.7}$$

To show (S.5), write

$$\sum_{j=1}^{N} B_{t,jk}\epsilon_{jt} = \mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\epsilon}_t = \mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2}\boldsymbol{\epsilon}_t^* = \sum_{j=1}^{N}(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_j\epsilon_{jt}^*,$$

where $\{\boldsymbol{\epsilon}_t^*\}$ is as in Assumption R2. Then by the independence of $\{\mathbf{B}_t\}$ and $\{\boldsymbol{\epsilon}_t\}$ (thus $\{\boldsymbol{\epsilon}_t^*\}$) assumed in M3,

$$\begin{aligned}
&E\big((\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_j\epsilon_{jt}^*|(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_s, \epsilon_{st}^*,\ s\le j-1\big)\\
&= E\big((\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_j|(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_s,\ s\le j-1\big)\cdot E(\epsilon_{jt}^*|\epsilon_{st}^*,\ s\le j-1) = 0,
\end{aligned}$$

since $\{\epsilon_{jt}^*\}_{1\le j\le N}$ is a martingale difference. Hence $\{(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_j\epsilon_{jt}\}_{1\le j\le N}$ is a martingale difference. By Lemma 2.1 of Li (2003), Assumptions M3, R2 and (S.2), we then have,

$$\begin{aligned}
&E\Big|N^{-\frac{1}{2}-\frac{1}{2w}}\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\epsilon}_t\Big|^{2w} = E\Big|N^{-\frac{1}{2}-\frac{1}{2w}}\sum_{j=1}^{N}(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_j\epsilon_{jt}^*\Big|^{2w}\\
&\le N^{-2}(36w)^{2w}(1+(2w-1)^{-1})^{w}\sum_{j=1}^{N}E|(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_j\epsilon_{jt}^*|^{2w}\\
&= N^{-2}(36w)^{2w}(1+(2w-1)^{-1})^{w}\sum_{j=1}^{N}E|(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_j|^{2w}E|\epsilon_{jt}^*|^{2w}\\
&\le N^{-2}(36w\mu_{2w})^{2w}(1+(2w-1)^{-1})^{w}\sum_{j=1}^{N}E\Big|\max_{1\le j\le N}|B_{t,jk}|\Big|^{2w}\big\|\boldsymbol{\Sigma}_{\epsilon}^{1/2}\big\|_{\infty}^{2w}\\
&\le N^{-2}(36w\mu_{2w}S_{\epsilon})^{2w}(1+(2w-1)^{-1})^{w}\sum_{j=1}^{N}N\max_{1\le j\le N}E|B_{t,jk}|^{2w}\\
&\le (36w\mu_{2w}^2S_{\epsilon})^{2w}(1+(2w-1)^{-1})^{w} < \infty,
\end{aligned}$$

so that $\max_{1\le k\le K}\big\|N^{-\frac{1}{2}-\frac{1}{2w}}\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\epsilon}_t\big\|_{2w} < \infty$, which is (S.5).

To show (S.6), observe that

$$\begin{aligned}
\Theta_{m,2w} &\le \sum_{t=m}^{\infty}\max_{1\le k\le K}N^{-\frac{1}{2}-\frac{1}{2w}}\Big[\big\|\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2}(\boldsymbol{\epsilon}_t^*-\boldsymbol{\epsilon}_t'^*)\big\|_{2w}\\
&\qquad + \big\|(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2}-\mathbf{B}_{t,k}'^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})\boldsymbol{\epsilon}_t'^*\big\|_{2w}\Big]\\
&\le \sum_{t=m}^{\infty}\max_{1\le k\le K}N^{-\frac{1}{2}-\frac{1}{2w}}\Bigg[\Big\|\sum_{j=1}^{N}(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_j(\epsilon_{jt}^*-\epsilon_{jt}'^*)\Big\|_{2w}\\
&\qquad + \Big\|\sum_{j=1}^{N}(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2}-\mathbf{B}_{t,k}'^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_j\epsilon_{jt}'^*\Big\|_{2w}\Bigg].
\end{aligned}$$

The terms $\{(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_j(\epsilon_{jt}^*-\epsilon_{jt}'^*)\}_j$ and $\{(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2}-\mathbf{B}_{t,k}'^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_j\epsilon_{jt}'^*\}_j$ can be shown to be martingale differences with respect to the filtration

$$\mathcal{F}_j = \sigma(\epsilon_{st}^*, \epsilon_{st}'^*, (\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_s, (\mathbf{B}_{t,k}'^{\mathrm{T}}\boldsymbol{\Sigma}_{\epsilon}^{1/2})_s,\ s\le j),$$

using similar arguments as before. Hence we can use Lemma 2.1 of Li (2003), Assumptions M3, R2, M4 and (S.2) to show that

$$\begin{aligned}
&\Big\|N^{-\frac{1}{2}-\frac{1}{2w}}\sum_{j=1}^{N}(\mathbf{B}_{t,k}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon}^{1/2})_j(\epsilon_{jt}^{*}-\epsilon_{jt}^{\prime *})\Big\|_{2w}\\
&\leq 36w(1+(2w-1)^{-1})^{1/2}\\
&\cdot\left[N^{-2}\sum_{j=1}^{N}E\Big|\max_{1\leq j\leq N}|B_{t,jk}|\Big|^{2w}\big\|\mathbf{\Sigma}_{\epsilon}^{1/2}\big\|_{\infty}^{2w}(\theta_{t,2w,j}^{\epsilon^*})^{2w}\right]^{1/2w}\\
&\leq 36w\mu_w S_{\epsilon}(1+(2w-1)^{-1})^{1/2}\max_{1\leq j\leq N}\theta_{t,2w,j}^{\epsilon^*}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\Big\|N^{-\frac{1}{2}-\frac{1}{2w}}\sum_{j=1}^{N}(\mathbf{B}_{t,k}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon}^{1/2}-\mathbf{B}_{t,k}^{\prime\mathrm{T}}\mathbf{\Sigma}_{\epsilon}^{1/2})_j\epsilon_{jt}^{\prime *}\Big\|_{2w}\\
&\leq 36w\mu_w S_{\epsilon}(1+(2w-1)^{-1})^{1/2}\max_{1\leq j\leq NK}\theta_{t,2w,j}^{b}.
\end{aligned}$$

Hence combining and using Assumption M4, we have

$$\begin{aligned}
\Theta_{m,2w}&\leq 36w\mu_w S_{\epsilon}(1+(2w-1)^{-1})^{1/2}(\Theta_{m,2w}^{\epsilon^*}+\Theta_{m,2w}^{b})\\
&\leq 72Cw\mu_w S_{\epsilon}(1+(2w-1)^{-1})^{1/2}m^{-\alpha},
\end{aligned}$$

which is (S.6).

Finally, to find an upper bound for $P(\mathcal{A}_7^c)$, write

$$\sum_{j=1}^{N}\bar{B}_{\cdot,jk}\bar{\epsilon}_{j,\cdot}=\sum_{j=1}^{N}(\bar{\mathbf{B}}_{\cdot,k}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon}^{1/2})_j\bar{\epsilon}_{j,\cdot}^{*},$$

where $\bar{\mathbf{B}}_{\cdot,k}$ is the sample mean of $\{\mathbf{B}_{t,k}\}$, and similarly for $\bar{\epsilon}_{j,\cdot}$. By the independence of $\{\mathbf{B}_t\}$ and $\{\boldsymbol{\epsilon}_t\}$,

$$\begin{aligned}
&E\big((\bar{\mathbf{B}}_{\cdot,k}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon}^{1/2})_j\bar{\epsilon}_{j,\cdot}^{*}|(\mathbf{B}_{t,k}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon}^{1/2})_s,\epsilon_{st}^{*},\ t=1,\ldots,T,\ s\leq j-1\big)\\
&=E\big((\bar{\mathbf{B}}_{\cdot,k}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon}^{1/2})_j|(\mathbf{B}_{t,k}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon}^{1/2})_s,\ t=1,\ldots,T,\ s\leq j-1\big)\\
&\quad\cdot E(\bar{\epsilon}_{j,\cdot}^{*}|\epsilon_{st}^{*},\ t=1,\ldots,T,\ s\leq j-1)=0,
\end{aligned}$$

since $\{\epsilon_{jt}^{*}\}_{1\leq j\leq N}$ is a martingale difference. Hence $\{(\bar{\mathbf{B}}_{\cdot,k}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon}^{1/2})_j\bar{\epsilon}_{j,\cdot}\}_{1\leq j\leq N}$ is a martingale difference. Moreover, on $\mathcal{A}_4\cap\mathcal{A}_5$, it is easy to show that $\max_j|(\bar{\mathbf{B}}_{\cdot,k}^{\mathrm{T}}\mathbf{\Sigma}_{\epsilon}^{1/2})_j\bar{\epsilon}_{j,\cdot}^{*}|\leq c_T(\mu_{b,\max}+c_T)S_{\epsilon}$.

Hence, we can apply the Azuma's inequality to get

$$
\begin{aligned}
P(\mathcal{A}_7^c) &\leq P(\mathcal{A}_7^c \cap \mathcal{A}_4 \cap \mathcal{A}_5) + P(\mathcal{A}_4^c \cup \mathcal{A}_5^c), \\
&\leq KP\bigg(\Big|\sum_{j=1}^{N} (\bar{\mathbf{B}}_{\cdot,k}^{\mathrm{T}} \mathbf{\Sigma}_\epsilon^{1/2})_j \bar{\epsilon}_{j,\cdot}^* \Big| \qquad\qquad (\text{S.8}) \\
&\qquad\qquad \geq 2^{1/2} c_T N^{1/2} (\mu_{b,\max} + c_T) S_\epsilon \log^{1/2}(T \vee N), \mathcal{A}_4 \cap \mathcal{A}_5 \bigg) \\
&\quad + P(\mathcal{A}_4^c) + P(\mathcal{A}_5^c) \\
&\leq 2K \exp\left(\frac{-2c_T^2 N (\mu_{b,\max} + c_T)^2 S_\epsilon^2 \log(T \vee N)}{2N c_T^2 (\mu_{b,\max} + c_T)^2 S_\epsilon^2} \right) + P(\mathcal{A}_4^c) + P(\mathcal{A}_5^c) \\
&= \frac{2K}{T \vee N} + P(\mathcal{A}_4^c) + P(\mathcal{A}_5^c). \qquad\qquad (\text{S.9})
\end{aligned}
$$

Combining (S.3), (S.4), (S.7) and (S.9), and using $P(\mathcal{M}) \geq 1 - \sum_{j=1}^{7} P(\mathcal{A}_j^c)$, we can arrive at the result as stated in the theorem. □

For the remaining theorems, as argued before, we shall just prove the corresponding results on the set $\mathcal{M}$, and the result in Theorem S.1 does the rest.

Proof of Theorem 1. From (3.4), using $\mathbf{y}^v = \mathbf{\Pi}^{*\otimes}(\mathbf{1}_T \otimes \boldsymbol{\mu}^* + \mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}^v)$ where we define $\mathbf{\Pi}^{*\otimes} = \big(\mathbf{I}_{TN} - \mathbf{A}^{*\otimes} - \sum_{i=1}^{M} \delta_i^* \mathbf{W}_{0i}^{\otimes}\big)^{-1}$, we can easily show that

$$
\boldsymbol{\beta}(\boldsymbol{\theta}^*) = \boldsymbol{\beta}^* + (\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\boldsymbol{\epsilon}^v.
$$

Hence we have

$$
\begin{aligned}
\boldsymbol{\beta}(\widetilde{\boldsymbol{\theta}}) &= (\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\Big((\mathbf{\Pi}^{*\otimes})^{-1} + (\mathbf{A}^{*\otimes} - \widetilde{\mathbf{A}}^{\otimes}) + \sum_{i=1}^{M}(\delta_i^* - \widetilde{\delta}_i)\mathbf{W}_{0i}^{\otimes}\Big) \\
&\quad \cdot \mathbf{\Pi}^{*\otimes}(\mathbf{1}_T \otimes \boldsymbol{\mu}^* + \mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}^v) \\
&= \boldsymbol{\beta}(\boldsymbol{\theta}^*) + (\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\Big((\mathbf{A}^{*\otimes} - \widetilde{\mathbf{A}}^{\otimes}) + \sum_{i=1}^{M}(\delta_i^* - \widetilde{\delta}_i)\mathbf{W}_{0i}^{\otimes}\Big)\mathbf{\Pi}^{*\otimes}(\mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}^v).
\end{aligned}
$$

From this, we can decompose $\boldsymbol{\beta}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{\beta}^* = \sum_{j=1}^{4} I_j$, where

$$
\begin{aligned}
I_1 &= (E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t))^{-1}(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t) - T^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\mathbf{X})(\boldsymbol{\beta}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{\beta}^*), \\
I_2 &= (E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t))^{-1}T^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\boldsymbol{\epsilon}^v, \\
I_3 &= (E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t))^{-1}T^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\Big((\mathbf{A}^{*\otimes} - \widetilde{\mathbf{A}}^{\otimes}) + \sum_{i=1}^{M}(\delta_i^* - \widetilde{\delta}_i)\mathbf{W}_{0i}^{\otimes}\Big)\mathbf{\Pi}^{*\otimes}\mathbf{X}\boldsymbol{\beta}^*, \\
I_4 &= (E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t))^{-1}T^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\Big((\mathbf{A}^{*\otimes} - \widetilde{\mathbf{A}}^{\otimes}) + \sum_{i=1}^{M}(\delta_i^* - \widetilde{\delta}_i)\mathbf{W}_{0i}^{\otimes}\Big)\mathbf{\Pi}^{*\otimes}\boldsymbol{\epsilon}^v.
\end{aligned}
$$

To bound the above, by Assumption R3, $\sigma_K(E(\mathbf{X}^{\mathrm{T}}\mathbf{B}_t)) \geq Nu$, where $u > 0$ is a constant. This implies that

$$
\begin{aligned}
&\lambda_{\min}(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t)) = \sigma_K^2(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)) \geq N^2u^2, \text{ so that} \\
&\big\|\big(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t)\big)^{-1}\big\|_1 \leq \frac{K^{1/2}}{\lambda_{\min}(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t))} \leq \frac{K^{1/2}}{N^2u^2}. \qquad (\text{S.10})
\end{aligned}
$$

Then defining $\mathbf{U} = \mathbf{I}_N \otimes T^{-1}\sum_{t=1}^T \text{vec}(\mathbf{B}_t - \bar{\mathbf{B}})\text{vec}^{\mathrm{T}}(\mathbf{X}_t)$ and $\mathbf{U}_0 = \mathbf{I}_N \otimes E(\mathbf{b}_t\mathbf{x}_t^{\mathrm{T}})$, we can write $T^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v = \mathbf{V}_{\mathbf{I}_N}^{\mathrm{T}}\mathbf{U}\mathbf{V}_{\mathbf{I}_N}$ and $E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t) = \mathbf{V}_{\mathbf{I}_N}^{\mathrm{T}}(\mathbf{I}_N \otimes E(\mathbf{b}_t\mathbf{x}_t^{\mathrm{T}}))\mathbf{V}_{\mathbf{I}_N}$ by Lemma 2.

We then have $\|\boldsymbol{\beta}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{\beta}^*\|_1 \leq \sum_{i=j}^4 \|I_j\|_1$, where on the set $\mathcal{M}$,

$$\begin{aligned}
\|I_1\|_1 &\leq \big\|\big(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t)\big)^{-1}\big\|_1\big\|(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t) - T^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\mathbf{X})(\boldsymbol{\beta}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{\beta}^*)\big\|_1 \\
&\leq \frac{K^{1/2}}{N^2u^2}\Big\{\big\|\mathbf{V}_{\mathbf{I}_N}^{\mathrm{T}}(\mathbf{U}_0 - \mathbf{U})^{\mathrm{T}}\mathbf{V}_{\mathbf{I}_N}\mathbf{V}_{\mathbf{I}_N}^{\mathrm{T}}\mathbf{U}_0\big\|_1 \\
&\qquad + \big\|\mathbf{V}_{\mathbf{I}_N}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{V}_{\mathbf{I}_N}\mathbf{V}_{\mathbf{I}_N}^{\mathrm{T}}(\mathbf{U}_0 - \mathbf{U})\big\|_1\Big\}\big\|\mathbf{V}_{\mathbf{I}_N}(\boldsymbol{\beta}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{\beta}^*)\big\|_1 \\
&\leq \frac{K^{1/2}}{N^2u^2}\Big\{K\|\mathbf{U}_0 - \mathbf{U}\|_{\max} \cdot N \cdot K\|\mathbf{U}_0\|_{\max} + \Big(K\big\|\mathbf{V}_{\mathbf{I}_N}^{\mathrm{T}}(\mathbf{U} - \mathbf{U}_0)^{\mathrm{T}}\mathbf{V}_{\mathbf{I}_N}\big\|_{\max} \\
&\qquad + K\big\|\mathbf{V}_{\mathbf{I}_N}^{\mathrm{T}}\mathbf{U}_0^{\mathrm{T}}\mathbf{V}_{\mathbf{I}_N}\big\|_{\max}\Big) \cdot K\|\mathbf{U}_0 - \mathbf{U}\|_{\max}\Big\} \cdot N\|\boldsymbol{\beta}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{\beta}^*\|_1 \\
&\leq K^{1/2}(2c_T\sigma_{bx}(1 + \mu_{b,\max} + c_T) + c_T^2(1 + \mu_{b,\max} + c_T)^2)\|\boldsymbol{\beta}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{\beta}^*\|_1 \\
&= O(c_T\|\boldsymbol{\beta}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{\beta}^*\|_1),
\end{aligned}$$

where the second last line used Assumption R3, that the entries in $\mathbf{U}_0$ are uniformly bounded away from infinity (by $\sigma_{bx} < \infty$, say). To bound $\|I_2\|_1$, on $\mathcal{M}$,

$$\begin{aligned}
\|I_2\|_1 &\leq \frac{K^{1/2}}{N^2u^2}\|T^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\|_1\|T^{-1}\mathbf{B}^{v\mathrm{T}}\boldsymbol{\epsilon}^v\|_1 \\
&\leq \frac{K^{1/2}}{N^2u^2}\big\|\mathbf{V}_{\mathbf{I}_N}^{\mathrm{T}}(\mathbf{U} - \mathbf{U}_0)\mathbf{V}_{\mathbf{I}_N} + \mathbf{V}_{\mathbf{I}_N}^{\mathrm{T}}\mathbf{U}_0^{\mathrm{T}}\mathbf{V}_{\mathbf{I}_N}\big\|_1 \\
&\quad \cdot (K \cdot c_TN^{\frac{1}{2}+\frac{1}{2w}} + \sqrt{2}Kc_TN^{1/2}\log^{1/2}(T \vee N)S_\epsilon(\mu_{b,\max} + c_T)) \\
&= O(c_TN^{-\frac{1}{2}+\frac{1}{2w}}).
\end{aligned}$$

To bound $\|I_3\|_1$, we denote $\mathbf{W}_j$, $\mathbf{B}_{t,j}$ and $\mathbf{X}_{t,j}$ the jth column of $\mathbf{W}, \mathbf{B}_t$ and $\mathbf{X}_t$ respectively. Also, define $\boldsymbol{\pi}_j^{\mathrm{T}}$ to be the jth row of $\boldsymbol{\Pi}$. Then on $\mathcal{M}$, writing $\mathbf{W} = \mathbf{A} + \sum_{i=1}^M \delta_i\mathbf{W}_{0i}$,

$$\begin{aligned}
\|I_3\|_1 &\leq \frac{K^{1/2}}{N^2u^2}\|T^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\|_1\Big\|T^{-1}\sum_{t=1}^T(\mathbf{B}_t - \bar{\mathbf{B}})^{\mathrm{T}}(\mathbf{W}^* - \widetilde{\mathbf{W}})\boldsymbol{\Pi}^*\mathbf{X}_t\Big\|_1\|\boldsymbol{\beta}^*\|_1 \\
&\leq \frac{K^{1/2}\|\boldsymbol{\beta}^*\|_1}{N^2u^2} \cdot O(N) \cdot \Big(K\max_{1\leq r\leq K}\Big|\sum_{j=1}^N(\mathbf{W}_j^* - \widetilde{\mathbf{W}}_j)^{\mathrm{T}}T^{-1}\sum_{t=1}^T(\mathbf{B}_{t,r} - \bar{\mathbf{B}}_{\cdot,r})\mathbf{X}_{t,r}^{\mathrm{T}}\boldsymbol{\pi}_j^*\Big|\Big) \\
&\leq O(N^{-1}) \cdot \sum_{j=1}^N(\sigma_{bx} + c_T(1 + \mu_{b,\max} + c_T))\|\mathbf{W}_j^* - \widetilde{\mathbf{W}}_j\|_1\|\boldsymbol{\pi}^*\|_1 \\
&\leq O(N^{-1})\Big(\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1 + cN\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|_1\Big) \\
&= O(\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|_1 + N^{-1}\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1),
\end{aligned}$$

where the constants η and c are from Assumptions M1 and R1 respectively. Similarly, on $\mathcal{M}$,

$$\|I_4\|_1 = O(c_T\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|_1 + c_TN^{-1}\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1).$$

From the above, combining with $P(\mathcal{M}) \to 1$ from Theorem S.1, we have

$$\|\boldsymbol{\beta}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{\beta}^*\|_1 = O_p(c_TN^{-\frac{1}{2}+\frac{1}{2w}} + \|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|_1 + N^{-1}\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1). \tag{S.11}$$

It remains to find the order of $\|\widetilde{\boldsymbol\delta}-\boldsymbol\delta\|_1$. Firstly, from (3.5),

$$\begin{aligned}\mathbf{B}^{\mathrm T}\mathbf{Z}\widetilde{\boldsymbol\xi}-\mathbf{B}^{\mathrm T}\mathbf{y}+\mathbf{K}(\mathbf{I}_{TN}-\widetilde{\mathbf{A}}^{\otimes})\mathbf{y}^v &= \mathbf{B}^{\mathrm T}\mathbf{Z}(\widetilde{\boldsymbol\xi}-\boldsymbol\xi^*)-\mathbf{B}^{\mathrm T}\boldsymbol\epsilon-\mathbf{B}^{\mathrm T}\mathbf{Z}\mathbf{V}_0\boldsymbol\delta^*-\mathbf{B}^{\mathrm T}\mathbf{X}_{\boldsymbol\beta^*}\mathrm{vec}(\mathbf{I}_N)\\ &\quad+\mathbf{K}(\mathbf{I}_{TN}-\mathbf{A}^{*\otimes})\mathbf{y}^v+\mathbf{K}(\mathbf{A}^{*\otimes}-\widetilde{\mathbf{A}}^{\otimes})\mathbf{y}^v,\ \text{with}\\ \mathbf{K}(\mathbf{I}_{TN}-\mathbf{A}^{*\otimes})\mathbf{y}^v &= \mathbf{H}\boldsymbol\delta^*+\mathbf{K}\mathbf{X}\boldsymbol\beta^*+\mathbf{K}\boldsymbol\epsilon^v\\ &= \mathbf{H}\boldsymbol\delta^*+T^{-1/2}N^{-a/2}\sum_{t=1}^T\mathbf{X}_t\otimes(\mathbf{B}_t-\bar{\mathbf{B}})\boldsymbol\gamma\boldsymbol\beta^*+\mathbf{K}\boldsymbol\epsilon^v\\ &= \mathbf{H}\boldsymbol\delta^*+\mathbf{B}^{\mathrm T}\mathbf{X}_{\boldsymbol\beta^*}\mathrm{vec}(\mathbf{I}_N)+\mathbf{K}\boldsymbol\epsilon^v.\end{aligned}$$

Hence

$$\begin{aligned}\mathbf{B}^{\mathrm T}\mathbf{Z}\widetilde{\boldsymbol\xi}-\mathbf{B}^{\mathrm T}\mathbf{y}+\mathbf{K}(\mathbf{I}_{TN}-\widetilde{\mathbf{A}}^{\otimes})\mathbf{y}^v &= \mathbf{B}^{\mathrm T}\mathbf{Z}(\widetilde{\boldsymbol\xi}-\boldsymbol\xi^*)-\mathbf{B}^{\mathrm T}\boldsymbol\epsilon+(\mathbf{H}-\mathbf{B}^{\mathrm T}\mathbf{Z}\mathbf{V}_0)\boldsymbol\delta^*\\ &\quad+\mathbf{K}[(\mathbf{A}^{*\otimes}-\widetilde{\mathbf{A}}^{\otimes})\mathbf{y}^v+\boldsymbol\epsilon^v].\end{aligned}$$

With this, defining $\widetilde{\mathbf{H}}=\mathbf{K}[(\mathbf{A}^{*\otimes}-\widetilde{\mathbf{A}}^{\otimes})\mathbf{y}^v+\boldsymbol\epsilon^v])$, we can decompose

$$\begin{aligned}&\widetilde{\boldsymbol\delta}-\boldsymbol\delta^*=D_1+D_2,\quad\text{where}\\ &D_1=[(\mathbf{H}-\mathbf{B}^{\mathrm T}\mathbf{Z}\mathbf{V}_0)^{\mathrm T}(\mathbf{H}-\mathbf{B}^{\mathrm T}\mathbf{Z}\mathbf{V}_0)]^{-1}(\mathbf{H}-\mathbf{B}^{\mathrm T}\mathbf{Z}\mathbf{V}_0)^{\mathrm T}(\mathbf{B}^{\mathrm T}\mathbf{Z}(\widetilde{\boldsymbol\xi}-\boldsymbol\xi^*)+\widetilde{\mathbf{H}}),\\ &D_2=-[(\mathbf{H}-\mathbf{B}^{\mathrm T}\mathbf{Z}\mathbf{V}_0)^{\mathrm T}(\mathbf{H}-\mathbf{B}^{\mathrm T}\mathbf{Z}\mathbf{V}_0)]^{-1}(\mathbf{H}-\mathbf{B}^{\mathrm T}\mathbf{Z}\mathbf{V}_0)^{\mathrm T}\mathbf{B}^{\mathrm T}\boldsymbol\epsilon.\end{aligned}$$

To proceed, we first find the order of $\|D_1\|_1$ on $\mathcal{M}$, and then find the order of the elements of D_2 by showing that it is asymptotically normal.

In order to do so, we define for $i=1,\ldots,M$,

$$\begin{aligned}&\mathbf{A}_1=T^{-1}\sum_{t=1}^T\mathbf{X}_t\otimes(\mathbf{B}_t-\bar{\mathbf{B}})\boldsymbol\gamma, && \mathbf{A}_1^0=E(\mathbf{X}_t\otimes\mathbf{B}_t\boldsymbol\gamma),\\ &\mathbf{A}_2=(\mathbf{V}_{\mathbf{I}_N}^{\mathrm T}\mathbf{U}^{\mathrm T}\mathbf{V}_{\mathbf{I}_N}\mathbf{V}_{\mathbf{I}_N}^{\mathrm T}\mathbf{U}\mathbf{V}_{\mathbf{I}_N})^{-1}, && \mathbf{A}_2^0=(\mathbf{V}_{\mathbf{I}_N}^{\mathrm T}\mathbf{U}_0^{\mathrm T}\mathbf{V}_{\mathbf{I}_N}\mathbf{V}_{\mathbf{I}_N}^{\mathrm T}\mathbf{U}_0\mathbf{V}_{\mathbf{I}_N})^{-1},\\ &\mathbf{A}_3=\mathbf{V}_{\mathbf{I}_N}^{\mathrm T}\mathbf{U}^{\mathrm T}\mathbf{V}_{\mathbf{I}_N}, && \mathbf{A}_3^0=\mathbf{V}_{\mathbf{I}_N}^{\mathrm T}\mathbf{U}_0^{\mathrm T}\mathbf{V}_{\mathbf{I}_N},\\ &\mathbf{A}_{4i}=\mathbf{V}_{\mathbf{W}_{0i}^{\mathrm T}}^{\mathrm T}\mathbf{U}\mathbf{V}_{\boldsymbol\Pi^*}\boldsymbol\beta^*, && \mathbf{A}_{4i}^0=\mathbf{V}_{\mathbf{W}_{0i}^{\mathrm T}}^{\mathrm T}\mathbf{U}_0\mathbf{V}_{\boldsymbol\Pi^*}\boldsymbol\beta^*,\\ &\mathbf{A}_{5i}=\mathbf{V}_{\mathbf{W}_{0i}^{\mathrm T}}^{\mathrm T}\Big(\mathbf{I}_N\otimes T^{-1}\sum_{t=1}^T\mathrm{vec}(\mathbf{B}_t-\bar{\mathbf{B}})\boldsymbol\epsilon_t^{\mathrm T}\Big)\mathrm{vec}(\boldsymbol\Pi^{*\mathrm T}).\end{aligned}\tag{S.12}$$

We find all related rates of the above first. On $\mathcal{M}$, it is immediate that

$$\|\mathbf{A}_1-\mathbf{A}_1^0\|_{\max}=O(c_T).\tag{S.13}$$

At the same time, using Assumption R5, on $\mathcal{M}$,

$$\|\mathbf{A}_1\|_1\le\|\mathbf{A}_1^0\|_1+\|\mathbf{A}_1-\mathbf{A}_1^0\|_1=O(N^{1+a}+c_TN^2)=(N^{1+a}).\tag{S.14}$$

Also, on $\mathcal{M}$,

$$\|\mathbf{A}_3^0\|_1\le K\|\mathbf{V}_{\mathbf{I}_N}^{\mathrm T}\mathbf{U}_0^{\mathrm T}\mathbf{V}_{\mathbf{I}_N}\|_{\max}=O(N),\quad\|\mathbf{A}_3-\mathbf{A}_3^0\|_1=O(c_TN),\quad\|\mathbf{A}_3\|_1=O(N).\tag{S.15}$$

Hence writing $\mathbf{A}_2=(\mathbf{A}_3\mathbf{A}_3^{\mathrm T})^{-1}$,

$$\|\mathbf{A}_2^0\|_1\le\frac{K^{1/2}}{\lambda_{\min}(\mathbf{A}_3^0\mathbf{A}_3^0)}\le\frac{K^{1/2}}{N^2u}=O(N^{-2}).\tag{S.16}$$

Also, since $\mathbf{A}_2 - \mathbf{A}_2^0 = (\mathbf{A}_2 - \mathbf{A}_2^0)((\mathbf{A}_2^0)^{-1} - \mathbf{A}_2^{-1})\mathbf{A}_2^0 + \mathbf{A}_2^0((\mathbf{A}_2^0)^{-1} - \mathbf{A}_2^{-1})\mathbf{A}_2^0$, and on $\mathcal{M}$,

$$\begin{aligned}\big\|(\mathbf{A}_2^0)^{-1} - \mathbf{A}_2^{-1}\big\|_1 &= \big\|\mathbf{A}_3^0\mathbf{A}_3^{0\mathrm{T}} - \mathbf{A}_3\mathbf{A}_3^{\mathrm{T}}\big\|_1 \le \big\|\mathbf{A}_3^0 - \mathbf{A}_3\big\|_1\big\|\mathbf{A}_3^{0\mathrm{T}}\big\|_1 + \big\|\mathbf{A}_3\big\|_1\big\|\mathbf{A}_3^{0\mathrm{T}} - \mathbf{A}_3^{\mathrm{T}}\big\|_1 \\ &= O(c_T N^2).\end{aligned}$$

Hence we have on $\mathcal{M}$,

$$\big\|\mathbf{A}_2 - \mathbf{A}_2^0\big\|_1 \le \frac{\big\|(\mathbf{A}_2^0)^{-1} - \mathbf{A}_2^{-1}\big\|_1\big\|\mathbf{A}_2^0\big\|_1^2}{1 - \big\|(\mathbf{A}_2^0)^{-1} - \mathbf{A}_2^{-1}\big\|_1\big\|\mathbf{A}_2^0\big\|_1} = O\Big(\frac{c_T N^2 \cdot N^{-4}}{1 - c_T N^2 \cdot N^{-2}}\Big) = O(c_T N^{-2}). \tag{S.17}$$

To bound $\big\|\mathbf{A}_{4i}\big\|_1$, note that

$$\begin{aligned}\big\|\mathbf{A}_{4i}^0\big\|_1 &\le K\big\|\boldsymbol{\beta}^*\big\|_1\big\|\mathbf{V}_{\mathbf{W}_{0i}^{\mathrm{T}}}^{\mathrm{T}}\mathbf{U}_0\mathbf{V}_{\boldsymbol{\Pi}^*}\big\|_{\max} = K\big\|\boldsymbol{\beta}^*\big\|_1 \max_{1\le k,m\le K}\Big|\sum_{\ell=1}^N \mathbf{W}_{0i,\ell}^{\mathrm{T}} E(\mathbf{X}_{t,k}\mathbf{B}_{t,m}^{\mathrm{T}})\boldsymbol{\pi}_\ell^*\Big| \\ &= O(\big\|\mathbf{W}_{0i}\big\|_1\big\|\boldsymbol{\Pi}^*\big\|_\infty \cdot N) = O(N),\end{aligned} \tag{S.18}$$

where the last line used Assumptions M1 and M3. Similarly, on $\mathcal{M}$,

$$\big\|\mathbf{A}_{4i} - \mathbf{A}_{4i}^0\big\|_1 = O(c_T N), \quad \big\|\mathbf{A}_{4i}\big\|_1 = O(N). \tag{S.19}$$

To bound $\big\|\mathbf{A}_{5i}\big\|_1$, note that on $\mathcal{M}$, an element in $\mathbf{A}_{5i}$ is bounded by

$$\Big|\sum_{\ell=1}^N \mathbf{W}_{0i,\ell}^{\mathrm{T}} T^{-1}\sum_{t=1}^T(\mathbf{B}_{t,k} - \bar{\mathbf{B}}_k)\boldsymbol{\epsilon}_t^{\mathrm{T}}\boldsymbol{\pi}_\ell^*\Big| = O(c_T N),$$

so that on $\mathcal{M}$,

$$\big\|\mathbf{A}_{5i}\big\|_1 = O(c_T N). \tag{S.20}$$

With these rates, we now focus on the order of $\big\|D_1\big\|_1$ first. We decompose $D_1 = F_1 + F_2 + F_3$, where

$$\begin{aligned}F_1 &= [(\mathbf{H}_{20} - \mathbf{H}_{10})^{\mathrm{T}}(\mathbf{H}_{20} - \mathbf{H}_{10})]^{-1}\Big\{(\mathbf{H}_{20} - \mathbf{H}_{10})^{\mathrm{T}}(\mathbf{H}_{20} - \mathbf{H}_{10}) \\ &\quad - T^{-1}N^a(\mathbf{H} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0)^{\mathrm{T}}(\mathbf{H} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0)\Big\}D_1, \\ F_2 &= [(\mathbf{H}_{20} - \mathbf{H}_{10})^{\mathrm{T}}(\mathbf{H}_{20} - \mathbf{H}_{10})]^{-1}\Big(T^{-1/2}N^{a/2}\mathbf{H} - \mathbf{H}_{20} - T^{-1/2}N^{a/2}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0 + \mathbf{H}_{10}\Big)^{\mathrm{T}} \\ &\quad \cdot T^{-1/2}N^{a/2}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*) + \widetilde{\mathbf{H}}), \\ F_3 &= [(\mathbf{H}_{20} - \mathbf{H}_{10})^{\mathrm{T}}(\mathbf{H}_{20} - \mathbf{H}_{10})]^{-1}(\mathbf{H}_{20} - \mathbf{H}_{10})^{\mathrm{T}} \cdot T^{-1/2}N^{a/2}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*) + \widetilde{\mathbf{H}}),\end{aligned}$$

with the definition of $\mathbf{H}_{10}$ and $\mathbf{H}_{20}$ being

$$\begin{aligned}\mathbf{H}_{10} &= \Big(\mathbf{I}_N \otimes (\mathbf{I}_N \otimes \boldsymbol{\gamma}^{\mathrm{T}})E(\mathrm{vec}(\mathbf{B}_t^{\mathrm{T}})\mathrm{vec}(\mathbf{X}_t^{\mathrm{T}})^{\mathrm{T}})(\mathbf{I}_N \otimes \boldsymbol{\beta}^*)\boldsymbol{\Pi}^{*\mathrm{T}}\Big)\mathbf{V}_0, \\ \mathbf{H}_{20} &= E(\mathbf{X}_t \otimes \mathbf{B}_t\boldsymbol{\gamma})\Big(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t)\Big)^{-1}E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)\Big(\mathbf{V}_{\mathbf{W}_{01}^{\mathrm{T}}}^{\mathrm{T}} \cdots \mathbf{V}_{\mathbf{W}_{0M}^{\mathrm{T}}}^{\mathrm{T}}\Big)(\mathbf{I}_M \otimes \mathbf{U}_0\mathbf{V}_{\boldsymbol{\Pi}^*}\boldsymbol{\beta}^*) \\ &= A_1^0A_2^0A_3^0(A_{41}, \ldots, A_{4M}),\end{aligned}$$

where we used the definitions in (S.12).

To bound the L_1 norm of the F_1 to F_3, we first observe that by Assumption R3, we have

$$\sigma_M^2(\mathbf{H}_{10}) \geq \sigma_M^2(\mathbf{V}_0)\sigma_N^2((\mathbf{I}_N \otimes \boldsymbol{\gamma}^{\mathrm{T}})E(\mathrm{vec}(\mathbf{B}_t^{\mathrm{T}})\mathrm{vec}(\mathbf{X}_t^{\mathrm{T}})^{\mathrm{T}})(\mathbf{I}_N \otimes \boldsymbol{\beta}^*)\boldsymbol{\Pi}^{*\mathrm{T}})$$
$$\geq CN \cdot N^a = CN^{1+a},$$

where $C > 0$ is a generic constant. Also, by Assumptions R4 and R5, the rate of $\lambda_{\min}(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}^{\mathrm{T}}\mathbf{X}_t))$ in (S.10) and the rate in $\mathbf{A}_{4i}$ derived in (S.18) which is also true elementwise, we have

$$\sigma_M(\mathbf{H}_{20}) \geq \sigma_K(\mathbf{A}_1^0)\sigma_K(\mathbf{A}_2^0)\sigma_K(\mathbf{A}_3^0)\sigma_{\min}(\mathbf{A}_{41}, \ldots, \mathbf{A}_{4M})$$
$$\geq \frac{CN^{1+a} \cdot N \cdot N}{\lambda_{\max}(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}^{\mathrm{T}}\mathbf{X}_t))} \geq CN^{1+a}.$$

Hence $\mathbf{H}_{20}$ has the smallest singular value of order larger than that for $\mathbf{H}_{10}$, and so we have

$$\sigma_M^2(\mathbf{H}_{20} - \mathbf{H}_{10}) \geq uN^{1+a}. \tag{S.21}$$

where $u > 0$ is a generic constant like C. Hence

$$\left\|[(\mathbf{H}_{20} - \mathbf{H}_{10})^{\mathrm{T}}(\mathbf{H}_{20} - \mathbf{H}_{10})]^{-1}\right\|_1 \leq \frac{M^{1/2}}{\lambda_{\min}((\mathbf{H}_{20} - \mathbf{H}_{10})^{\mathrm{T}}(\mathbf{H}_{20} - \mathbf{H}_{10}))} \leq \frac{M^{1/2}}{N^{1+a}u}. \tag{S.22}$$

Then we have

$$\begin{aligned} \left\|F_1\right\|_1 \leq \frac{M^{3/2}}{N^{1+a}u}\Big\{&\left\|\mathbf{H}_{20} - \mathbf{H}_{10}\right\|_1\Big(\left\|T^{-1/2}N^{a/2}\mathbf{H} - \mathbf{H}_{20}\right\|_{\max} + \left\|T^{-1/2}N^{a/2}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0 - \mathbf{H}_{10}\right\|_{\max}\Big) \\ &+ \left\|T^{-1/2}N^{a/2}(\mathbf{H} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0)\right\|_{\max} \\ &\cdot \Big(\left\|T^{-1/2}N^{a/2}\mathbf{H} - \mathbf{H}_{20}\right\|_1 + \left\|T^{-1/2}N^{a/2}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0 - \mathbf{H}_{10}\right\|_1\Big)\Big\}\left\|D_1\right\|_1, \end{aligned} \tag{S.23}$$

$$\begin{aligned} \left\|F_2\right\|_1 \leq \frac{M^{3/2}}{N^{1+a}u}&\Big(\left\|T^{-1/2}N^{a/2}\mathbf{H} - \mathbf{H}_{20}\right\|_{\max} + \left\|T^{-1/2}N^{a/2}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0 - \mathbf{H}_{10}\right\|_{\max}\Big) \\ &\cdot \Big(\left\|T^{-1/2}N^{a/2}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\right\|_1\left\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\right\|_1 + \left\|T^{-1/2}N^{a/2}\widetilde{\mathbf{H}}\right\|_1\Big), \end{aligned} \tag{S.24}$$

$$\left\|F_3\right\|_1 \leq \frac{M^{3/2}}{N^{1+a}u}\left\|\mathbf{H}_{20} - \mathbf{H}_{10}\right\|_{\max}\Big(\left\|T^{-1/2}N^{a/2}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\right\|_1\left\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\right\|_1 + \left\|T^{-1/2}N^{a/2}\widetilde{\mathbf{H}}\right\|_1\Big). \tag{S.25}$$

To bound the above, observe that the max-norm of $T^{-1/2}N^{a/2}\mathbf{H} - \mathbf{H}_{20}$ can be bounded by

$$\begin{aligned} &\left\|T^{-1/2}N^{a/2}\mathbf{H} - \mathbf{H}_{20}\right\|_{\max} = \max_{1\leq i\leq M}\left\|\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3(\mathbf{A}_{4i} + \mathbf{A}_{5i}) - \mathbf{A}_1\mathbf{A}_2\mathbf{A}_3\mathbf{A}_{4i}\right\|_{\max} \\ &\leq \max_{1\leq i\leq M}\left\|\mathbf{A}_1\right\|_{\max}\left\|\mathbf{A}_2\right\|_1\left\|\mathbf{A}_3\right\|_1\left\|\mathbf{A}_{5i}\right\|_1 \\ &\quad + \max_{1\leq i\leq M}\Big\{\left\|\mathbf{A}_1\right\|_{\max}\left\|\mathbf{A}_2\mathbf{A}_3\mathbf{A}_{4i} - \mathbf{A}_2^0\mathbf{A}_3^0\mathbf{A}_{4i}^0\right\|_1 + \left\|\mathbf{A}_1 - \mathbf{A}_1^0\right\|_{\max}\left\|\mathbf{A}_2^0\mathbf{A}_3^0\mathbf{A}_{4i}^0\right\|_1\Big\}, \text{ with} \end{aligned} \tag{S.26}$$

$$\begin{aligned} \left\|\mathbf{A}_2\mathbf{A}_3\mathbf{A}_{4i} - \mathbf{A}_2^0\mathbf{A}_3^0\mathbf{A}_{4i}^0\right\|_1 \leq \max_{1\leq i\leq M}&\Big(\left\|\mathbf{A}_2\right\|_1\left\|\mathbf{A}_3 - \mathbf{A}_3^0\right\|_1\left\|\mathbf{A}_{4i}\right\|_1 \\ &+ \left\|\mathbf{A}_2\right\|_1\left\|\mathbf{A}_3^0\right\|_1\left\|\mathbf{A}_{4i} - \mathbf{A}_{4i}^0\right\|_1 + \left\|\mathbf{A}_2 - \mathbf{A}_2^0\right\|_1\left\|\mathbf{A}_3^0\right\|_1\left\|A_{4i}\right\|_1^0\Big). \end{aligned}$$

With (S.13) to (S.20), (S.26) is, on $\mathcal{M}$,

$$\left\|T^{-1/2}N^{a/2}\mathbf{H} - \mathbf{H}_{20}\right\|_{\max} = O(c_T). \text{ Also, } \left\|T^{-1/2}N^{a/2}\mathbf{H} - \mathbf{H}_{20}\right\|_1 = O(c_T N^2). \tag{S.27}$$

Also, defining $\mathbf{L} = T^{-1}\sum_{t=1}^{T}\text{vec}((\mathbf{B}_t - \bar{\mathbf{B}})^{\mathrm{T}})\text{vec}^{\mathrm{T}}(\mathbf{X}_t^{\mathrm{T}})$,

$$\begin{aligned}
&\big\|T^{-1/2}N^{a/2}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\big\|_1 = \Big\|T^{-1}\sum_{t=1}^{T}(\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\mathbf{y}_t^{\mathrm{T}}\Big\|_1 \\
&\le \big\|(\mathbf{I}_N \otimes \boldsymbol{\gamma}^{\mathrm{T}})\mathbf{L}(\mathbf{I}_N \otimes \boldsymbol{\beta}^*)\boldsymbol{\Pi}^{*\mathrm{T}}\big\|_1 + \Big\|(\mathbf{I}_N \otimes \boldsymbol{\gamma}^{\mathrm{T}})T^{-1}\sum_{t=1}^{T}\text{vec}((\mathbf{B}_t - \bar{\mathbf{B}})^{\mathrm{T}})\boldsymbol{\epsilon}_t^{\mathrm{T}}(\mathbf{I}_N \otimes \boldsymbol{\beta}^*)\boldsymbol{\Pi}^{*\mathrm{T}}\Big\|_1 \\
&= O(c_T N + N^a + c_T N) = O(N^a). \qquad \text{(S.28)}
\end{aligned}$$

Next, defining $\mathbf{L}_0 = E(\text{vec}(\mathbf{B}_t^{\mathrm{T}})\text{vec}(\mathbf{X}_t^{\mathrm{T}})^{\mathrm{T}})$ and $\mathbf{w}_{0i,j}^{\mathrm{T}}$ the jth row of $\mathbf{W}_{0i}$, we have on $\mathcal{M}$ that

$$\begin{aligned}
&\big\|T^{-1/2}N^{a/2}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0 - \mathbf{H}_{10}\big\|_{\max} \\
&= \max_{\substack{1\le i\le M\\ 1\le j\le N}}\Big\|\Big(T^{-1}\sum_{t=1}^{T}(\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\mathbf{y}_t^{\mathrm{T}} - (\mathbf{I}_N \otimes \boldsymbol{\gamma}^{\mathrm{T}})\mathbf{L}_0(\mathbf{I}_N \otimes \boldsymbol{\beta}^*)\boldsymbol{\Pi}^{*\mathrm{T}}\Big)\mathbf{w}_{0i,j}\Big\|_{\max} \\
&\le \max_{\substack{1\le i\le M\\ 1\le j\le N}}\Big\{\big\|\mathbf{I}_N \otimes \boldsymbol{\gamma}^{\mathrm{T}}\big\|_\infty\big\|\mathbf{L} - \mathbf{L}_0\big\|_{\max}\big\|\mathbf{I}_N \otimes \boldsymbol{\beta}^*\big\|_1\big\|\boldsymbol{\Pi}^{*\mathrm{T}}\big\|_1\big\|\mathbf{w}_{0i,j}\big\|_1 \\
&\quad + \big\|\mathbf{I}_N \otimes \boldsymbol{\gamma}^{\mathrm{T}}\big\|_\infty\Big\|T^{-1}\sum_{t=1}^{T}\text{vec}((\mathbf{B}_t - \bar{\mathbf{B}})^{\mathrm{T}})\boldsymbol{\epsilon}_t^{\mathrm{T}}\Big\|_{\max}\big\|\boldsymbol{\Pi}^{*\mathrm{T}}\big\|_1\big\|\mathbf{w}_{0i,j}\big\|_1\Big\} = O(c_T). \qquad \text{(S.29)}
\end{aligned}$$

The above also implies

$$\big\|T^{-1/2}N^{a/2}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0 - \mathbf{H}_{10}\big\|_1 = O(c_T N^2). \qquad \text{(S.30)}$$

Finally, decompose

$$\begin{aligned}
T^{-1/2}N^{a/2}\widetilde{\mathbf{H}} &= \mathbf{A}_1\mathbf{A}_2\mathbf{A}_3\Big(T^{-1}\sum_{t=1}^{T}(\mathbf{B}_t - \bar{\mathbf{B}})^{\mathrm{T}}(\widetilde{\mathbf{A}} - \mathbf{A}^*)\mathbf{y}_t - T^{-1}\sum_{t=1}^{T}(\mathbf{B}_t - \bar{\mathbf{B}})^{\mathrm{T}}\boldsymbol{\epsilon}_t\Big) \\
&= \mathbf{A}_1\mathbf{A}_2\mathbf{A}_3\Big(\mathbf{V}_{(\widetilde{\mathbf{A}}-\mathbf{A}^*)^{\mathrm{T}}}^{\mathrm{T}}\Big(\mathbf{I}_N \otimes T^{-1}\sum_{t=1}^{T}\text{vec}(\mathbf{B}_t - \bar{\mathbf{B}})\text{vec}(\mathbf{X}_t)^{\mathrm{T}}\Big)\mathbf{V}_{\boldsymbol{\Pi}^*}\boldsymbol{\beta}^* \\
&\quad + \mathbf{V}_{(\widetilde{\mathbf{A}}-\mathbf{A}^*)^{\mathrm{T}}}^{\mathrm{T}}\Big(\mathbf{I}_N \otimes T^{-1}\sum_{t=1}^{T}\text{vec}(\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\epsilon}_t^{\mathrm{T}}\Big)\text{vec}(\boldsymbol{\Pi}^{*\mathrm{T}}) \\
&\quad - T^{-1}\sum_{t=1}^{T}(\mathbf{B}_t - \bar{\mathbf{B}})^{\mathrm{T}}\boldsymbol{\epsilon}_t\Big),
\end{aligned}$$

so that on $\mathcal{M}$,

$$\begin{aligned}
\big\|T^{-1/2}N^{a/2}\widetilde{\mathbf{H}}\big\|_1 &= O\Big(N^{1+a}N^{-2}N\Big\{\big\|\boldsymbol{\beta}^*\big\|_1 \cdot K \cdot \big\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\big\|_1 \cdot \frac{1}{1-\eta} \\
&\quad + K \cdot \frac{c_T}{1-\eta}\big\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\big\|_1 + c_T N^{\frac{1}{2}+\frac{1}{2w}}\Big\}\Big) \\
&= O(N^a\big\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\big\|_1 + c_T N^{\frac{1}{2}+\frac{1}{2w}+a}). \qquad \text{(S.31)}
\end{aligned}$$

Collecting all the inequalities proved from (S.22) to (S.31), we can then conclude that on $\mathcal{M}$,

$$\begin{aligned}\|D_1\|_1 &\le \frac{M^{3/2}}{N^{1+a}u}\big(o(c_TN^2)+o(1)\cdot o(c_TN^2+c_TN^2)\big)\|D_1\|_1\\&+\frac{M^{3/2}}{N^{1+a}u}\Big(O(c_T)\cdot O(N^a\|\widetilde{\boldsymbol{\xi}}-\boldsymbol{\xi}^*\|_1+c_TN^{\frac{1}{2}+\frac{1}{2w}+a})\Big)\\&+\frac{M^{3/2}}{N^{1+a}u}O(N^a\|\widetilde{\boldsymbol{\xi}}-\boldsymbol{\xi}^*\|_1+c_TN^{\frac{1}{2}+\frac{1}{2w}+a})\\&=O(N^{-1}\|\widetilde{\boldsymbol{\xi}}-\boldsymbol{\xi}^*\|_1+c_TN^{-\frac{1}{2}+\frac{1}{2w}}).\end{aligned}$$

For the rate of $\|D_2\|_1$, we refer the readers to the proof of asymptotic normality of $\widehat{\boldsymbol{\delta}}_H-\boldsymbol{\delta}_H$ in the proof of Theorem 5 for the asymptotic normality of D_2 (the proof is exactly the same except that here there is no restriction to the set H), and we just state (proof omitted) here that

$$T^{1/2}(\mathbf{R}_2\boldsymbol{\Sigma}\mathbf{R}_2^{\mathrm{T}})^{-1/2}D_2\xrightarrow{\mathcal{D}}N(\mathbf{0},\mathbf{I}_M),$$

where $\mathbf{R}_2=[(\mathbf{H}_{10}-\mathbf{H}_{20})^{\mathrm{T}}(\mathbf{H}_{10}-\mathbf{H}_{20})]^{-1}(\mathbf{H}_{10}-\mathbf{H}_{20})^{\mathrm{T}}$, and $\boldsymbol{\Sigma}$ is as defined in Theorem 3. By Assumption R6, we can conclude that all eigenvalues of Σ are of order N^b. Hence we have

$$\begin{aligned}\lambda_{\max}(\mathbf{R}_2\boldsymbol{\Sigma}\mathbf{R}_2^{\mathrm{T}})&\le\lambda_{\max}(\boldsymbol{\Sigma})\lambda_{\max}([(\mathbf{H}_{10}-\mathbf{H}_{20})^{\mathrm{T}}(\mathbf{H}_{10}-\mathbf{H}_{20})]^{-1})\\&\le\frac{\lambda_{\max}(\boldsymbol{\Sigma})}{\sigma_M^2(\mathbf{H}_{10}-\mathbf{H}_{20})}=O(N^{-1-a+b}),\end{aligned}$$

which can also be derived as the order for the lower bound of $\lambda_{\min}(\mathbf{R}_2\boldsymbol{\Sigma}\mathbf{R}_2)$. Hence we have $\|D_2\|_1=O_p(T^{-1/2}N^{-(1+a-b)/2})$. It means that

$$\begin{aligned}\|\widetilde{\boldsymbol{\delta}}-\boldsymbol{\delta}^*\|_1&=O_p(\|D_1\|_1+\|D_2\|_1)\\&=O_p(N^{-1}\|\widetilde{\boldsymbol{\xi}}-\boldsymbol{\xi}^*\|_1+c_TN^{-\frac{1}{2}+\frac{1}{2w}}+T^{-1/2}N^{-(1+a-b)/2})\\&=O_p(N^{-1}\|\widetilde{\boldsymbol{\xi}}-\boldsymbol{\xi}^*\|_1+c_TN^{-\frac{1}{2}+\frac{1}{2w}}).\end{aligned}$$

It is clear then from (S.11) that

$$\|\boldsymbol{\beta}(\widetilde{\boldsymbol{\theta}})-\boldsymbol{\beta}^*\|=O_p(c_TN^{-\frac{1}{2}+\frac{1}{2w}}+N^{-1}\|\widetilde{\boldsymbol{\xi}}-\boldsymbol{\xi}^*\|_1).$$

If $M=0$, then $\widetilde{\boldsymbol{\delta}}\equiv\boldsymbol{\delta}^*\equiv\mathbf{0}$, and so the above bound still holds from (S.11). This completes the proof of the Theorem. □

Proof of Theorem 2. Assume $M>0$ first. Since $\widetilde{\boldsymbol{\theta}}=(\widetilde{\boldsymbol{\xi}}^{\mathrm{T}},\widetilde{\boldsymbol{\delta}}^{\mathrm{T}})^{\mathrm{T}}$ is the LASSO solution for (2.7), we must have

$$\begin{aligned}&\frac{1}{2T}\|\mathbf{B}^{\mathrm{T}}\mathbf{y}-\mathbf{B}^{\mathrm{T}}\mathbf{Z}\widetilde{\boldsymbol{\xi}}-\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0\widetilde{\boldsymbol{\delta}}-\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\widetilde{\boldsymbol{\beta}}}\mathrm{vec}(\mathbf{I}_N)\|^2\\&\le\frac{1}{2T}\|\mathbf{B}^{\mathrm{T}}\mathbf{y}-\mathbf{B}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\xi}^*-\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0\boldsymbol{\delta}^*-\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\theta}^*)}\mathrm{vec}(\mathbf{I}_N)\|^2+\lambda_T(\|\boldsymbol{\xi}^*\|_1-\|\widetilde{\boldsymbol{\xi}}\|_1).\end{aligned}$$

But $\mathbf{B}^{\mathrm{T}}\mathbf{y}=\mathbf{B}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\xi}^*+\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0\boldsymbol{\delta}^*+\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N)+\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}$, so that the above becomes

$$\begin{aligned}&\frac{1}{2T}\|\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\boldsymbol{\xi}^*-\widetilde{\boldsymbol{\xi}})+\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0(\boldsymbol{\delta}^*-\widetilde{\boldsymbol{\delta}})+\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}^*-\widetilde{\boldsymbol{\beta}}}\mathrm{vec}(\mathbf{I}_N)+\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}\|^2\\&\le\frac{1}{2T}\|\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}^*-\boldsymbol{\beta}(\boldsymbol{\theta}^*)}\mathrm{vec}(\mathbf{I}_N)+\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}\|^2+\lambda_T(\|\boldsymbol{\xi}^*\|_1-\|\widetilde{\boldsymbol{\xi}}\|_1).\end{aligned}$$

Rearranging terms of the above and eliminating $\|\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}\|^2$ on both sides,

$$
\begin{aligned}
&\frac{1}{2T}\big\|\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\boldsymbol{\xi}^* - \widetilde{\boldsymbol{\xi}}) + \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0(\boldsymbol{\delta}^* - \widetilde{\boldsymbol{\delta}}) + \mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\theta}^*)-\widetilde{\boldsymbol{\beta}}}\mathrm{vec}(\mathbf{I}_N)\big\|^2 \\
&\qquad \le I_1 + \cdots + I_6 + \lambda_T(\|\boldsymbol{\xi}^*\|_1 - \|\widetilde{\boldsymbol{\xi}}\|_1), \quad \text{where} \\
&\quad I_1 = \frac{1}{T}\boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*),\ I_2 = \frac{1}{T}\boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*), \\
&\quad I_3 = \frac{1}{T}\boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N),\ I_4 = \frac{1}{T}\boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}^*-\boldsymbol{\beta}(\boldsymbol{\theta}^*)}\mathrm{vec}(\mathbf{I}_N), \\
&\quad I_5 = \frac{1}{T}\mathrm{vec}^{\mathrm{T}}(\mathbf{I}_N)\mathbf{X}^{\mathrm{T}}_{\boldsymbol{\beta}(\boldsymbol{\theta}^*)-\boldsymbol{\beta}^*}\mathbf{B}\mathbf{B}^{\mathrm{T}}\Big(\mathbf{Z}(\boldsymbol{\xi}^* - \widetilde{\boldsymbol{\xi}}) + \mathbf{Z}\mathbf{V}_0(\boldsymbol{\delta}^* - \widetilde{\boldsymbol{\delta}})\Big), \\
&\quad I_6 = \frac{1}{T}\mathrm{vec}^{\mathrm{T}}(\mathbf{I}_N)\mathbf{X}^{\mathrm{T}}_{\boldsymbol{\beta}(\boldsymbol{\theta}^*)-\boldsymbol{\beta}^*}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\theta}^*)-\widetilde{\boldsymbol{\beta}}}\mathrm{vec}(\mathbf{I}_N).
\end{aligned}
$$

Now by (S.28), on $\mathcal{M}$,

$$
\begin{aligned}
|I_1| &\le \|T^{-1}\boldsymbol{\epsilon}^{\mathrm{T}}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})\|_{\max} \cdot N^{-a}\|T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})^{\mathrm{T}}\mathbf{Z}\|_1 \cdot \|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1 \\
&= O(\|T^{-1}\boldsymbol{\epsilon}^{\mathrm{T}}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})\|_{\max}\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1).
\end{aligned}
$$

We also have on $\mathcal{M}$,

$$
\|T^{-1}\boldsymbol{\epsilon}^{\mathrm{T}}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})\|_{\max} = \max_{1\le i,j\le N}\Big|T^{-1}\sum_{t=1}^T \epsilon_{tj}\boldsymbol{\gamma}^{\mathrm{T}}(\mathbf{b}_{ti} - \bar{\mathbf{b}}_{\cdot i})\Big| = O(c_T). \tag{S.32}
$$

Hence on $\mathcal{M}$, using S.32, we have

$$
|I_1| = O(c_T\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1).
$$

Similarly,

$$
|I_2| = O(c_T\|\mathbf{V}_0(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*)\|_1) = O(c_TN\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|_1).
$$

For I_3, on $\mathcal{M}$,

$$
\begin{aligned}
|I_3| &\le \|T^{-1}\boldsymbol{\epsilon}^{\mathrm{T}}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})\|_{\max} \cdot N^{-a}\|T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})^{\mathrm{T}}\mathbf{X}_{\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N)\|_1 \\
&= O(c_T) \cdot N^{-a}\sum_{i,j=1}^N \Big|\boldsymbol{\gamma}^{\mathrm{T}}\Big(T^{-1}\sum_{t=1}^T(\mathbf{b}_{ti} - \bar{\mathbf{b}}_{\cdot i})\mathbf{x}_{tj}^{\mathrm{T}}\Big)(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\Big| \\
&\le O(c_T) \cdot N^{-a} \cdot K\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1\big(c_TN^2(1 + \mu_{b,\max} + c_T) + C_{bx}N^{1+a}\big) \\
&= O(c_TN\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1).
\end{aligned}
$$

With similar techniques for handling I_1 to I_3, we can show that on $\mathcal{M}$,

$$
\begin{aligned}
|I_4| &= O(c_TN\|\boldsymbol{\beta}(\boldsymbol{\theta}^*) - \boldsymbol{\beta}^*\|_1), \\
|I_5| &= O\big(\|\boldsymbol{\beta}(\boldsymbol{\theta}^*) - \boldsymbol{\beta}^*\|_1(\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1 + N\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|_1)\big), \\
|I_6| &= O(N\|\boldsymbol{\beta}(\boldsymbol{\theta}^*) - \boldsymbol{\beta}^*\|_1\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}(\boldsymbol{\theta}^*)\|_1).
\end{aligned}
$$

With a similar technique as in the proof of Theorem 1, we can show that on $\mathcal{M}$,

$$
\|\boldsymbol{\beta}(\theta^*) - \boldsymbol{\beta}^*\|_1 = O(c_TN^{-\frac{1}{2}+\frac{1}{2w}}).
$$

Hence using the triangle inequality and the result for $\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1$ in Theorem 1, we also have

$$\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}(\theta^*)\|_1 = O(c_T N^{-\frac{1}{2}+\frac{1}{2w}} + N^{-1}\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1).$$

Substituting these rates back into those for $|I_4|$ to $|I_6|$, together with those for $|I_1|$ to $|I_3|$, we can conclude that on $\mathcal{M}$,

$$\begin{aligned}\frac{1}{2T}\|\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\boldsymbol{\xi}^* - \widetilde{\boldsymbol{\xi}}) + \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0(\boldsymbol{\delta}^* - \widetilde{\boldsymbol{\delta}}) + \mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\theta}^*)-\widetilde{\boldsymbol{\beta}}}\mathrm{vec}(\mathbf{I}_N)\|^2 \\ \le Cc_T\Big(c_T N^{\frac{1}{2}+\frac{1}{2w}} + \|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1\Big) + \lambda_T(\|\boldsymbol{\xi}^*\|_1 - \|\widetilde{\boldsymbol{\xi}}\|_1),\end{aligned} \tag{S.33}$$

where C is a positive constant.

Set $\lambda_T = 2Cc_T$, and add both sides of (S.33) by the term $Cc_T\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1$. Noting that

$$\begin{aligned}\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1 + \|\boldsymbol{\xi}^*\|_1 - \|\widetilde{\boldsymbol{\xi}}\|_1 &= \sum_{i=1}^{2}(\|\widetilde{\boldsymbol{\xi}}_{J_i} - \boldsymbol{\xi}^*_{J_i}\|_1 + \|\boldsymbol{\xi}^*_{J_i}\|_1 - \|\widetilde{\boldsymbol{\xi}}_{J_i}\|_1) \\ &\le 2\|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\|_1 + 2\|\boldsymbol{\xi}^*_{J_2}\|_1,\end{aligned}$$

we then have the following two inequalities:

$$\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1 \le h_{N,T} + 4\|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\|_1, \tag{S.34}$$

$$\begin{aligned}\frac{1}{2T}\|\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\boldsymbol{\xi}^* - \widetilde{\boldsymbol{\xi}}) + \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0(\boldsymbol{\delta}^* - \widetilde{\boldsymbol{\delta}}) + \mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\theta}^*)-\widetilde{\boldsymbol{\beta}}}\mathrm{vec}(\mathbf{I}_N)\|^2 \\ \le Cc_T(h_{N,T} + 4\|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\|_1).\end{aligned} \tag{S.35}$$

where we define $h_{N,T} = 4\|\boldsymbol{\xi}^*_{J_2}\|_1 + c_T N^{\frac{1}{2}+\frac{1}{2w}}$.

Since $c_T N^{\frac{1}{2}+\frac{1}{2w}}$ dominates $\|\boldsymbol{\xi}^*_{J_2}\|_1$ by Assumption R8, (S.34) becomes, on $\mathcal{M}$,

$$\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1 = O(c_T N^{\frac{1}{2}+\frac{1}{2w}} + n^{1/2}\|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\|), \tag{S.36}$$

which is the first inequality in the theorem.

To prove the second part, we need an intermediate result. Consider decomposing the left hand side of (S.33) into $\sum_{j=1}^6 D_j$, where

$$\begin{aligned}&D_1 = \frac{1}{2T}\|\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\boldsymbol{\xi}^* - \widetilde{\boldsymbol{\xi}})\|^2,\ D_2 = \frac{1}{2T}\|\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0(\boldsymbol{\delta}^* - \widetilde{\boldsymbol{\delta}})\|^2,\ D_3 = \frac{1}{2T}\|\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\theta}^*)-\widetilde{\boldsymbol{\beta}}}\mathrm{vec}(\mathbf{I}_N)\|^2, \\ &D_4 = \frac{1}{T}(\boldsymbol{\xi}^* - \widetilde{\boldsymbol{\xi}})^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0(\boldsymbol{\delta}^* - \widetilde{\boldsymbol{\delta}}),\ D_5 = \frac{1}{T}(\boldsymbol{\xi}^* - \widetilde{\boldsymbol{\xi}})^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\theta}^*)-\widetilde{\boldsymbol{\beta}}}\mathrm{vec}(\mathbf{I}_N), \\ &D_6 = \frac{1}{T}(\boldsymbol{\delta}^* - \widetilde{\boldsymbol{\delta}})^{\mathrm{T}}\mathbf{V}_0^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}(\boldsymbol{\theta}^*)-\widetilde{\boldsymbol{\beta}}}\mathrm{vec}(\mathbf{I}_N).\end{aligned}$$

Now on $\mathcal{M}$, using (S.28) and the result of Theorem 1,

$$\begin{aligned}\|D_4\|_1 &\le \|\boldsymbol{\xi}^* - \widetilde{\boldsymbol{\xi}}\|_1\|T^{-1}\mathbf{Z}^{\mathrm{T}}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})\|_\infty \cdot N^{-a}\|T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\gamma})^{\mathrm{T}}\mathbf{Z}\|_1\|\mathbf{V}_0(\boldsymbol{\delta}^* - \widetilde{\boldsymbol{\delta}})\|_{\max} \\ &= O(N^a\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|_1) \\ &= O(N^{a-1}\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1^2 + c_T N^{a-\frac{1}{2}+\frac{1}{2w}}\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1).\end{aligned}$$

Similarly, on $\mathcal{M}$,

$$\begin{aligned}\|D_5\|_1 &= O(\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}(\boldsymbol{\theta}^*)\|_1) = O(N^{-1}\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1^2 + c_T N^{-\frac{1}{2}+\frac{1}{2w}}\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1), \\ \|D_6\|_1 &= O(\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|_1\{c_T N^{\frac{1}{2}+\frac{1}{2w}} + \|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1\}) \\ &= O(N^{-1}\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1^2 + c_T N^{-\frac{1}{2}+\frac{1}{2w}}\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1 + c_T^2 N^{1/w}).\end{aligned}$$

Since D_2 and D_3 are positive, from (S.35) and the bounds above, on $\mathcal{M}$,

$$
\begin{aligned}
\frac{1}{2T}\big\|\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\widetilde{\boldsymbol{\xi}}-\boldsymbol{\xi}^*)\big\|^2 &\le \big\|D_4\big\|_1 + \big\|D_5\big\|_1 + \big\|D_6\big\|_1 \\
&\quad + Cc_T(c_T N^{\frac{1}{2}+\frac{1}{2w}} + 4\big\|\widetilde{\boldsymbol{\xi}}_{J_1}-\boldsymbol{\xi}^*_{J_1}\big\|_1) \\
&\le O(N^{a-1}\big\|\widetilde{\boldsymbol{\xi}}-\boldsymbol{\xi}^*\big\|_1^2 + c_T N^{a-\frac{1}{2}+\frac{1}{2w}}\big\|\widetilde{\boldsymbol{\xi}}-\boldsymbol{\xi}^*\big\|_1 + c_T^2 N^{1/w}) \\
&\quad + Cc_T(c_T N^{\frac{1}{2}+\frac{1}{2w}} + 4\big\|\widetilde{\boldsymbol{\xi}}_{J_1}-\boldsymbol{\xi}^*_{J_1}\big\|_1) \\
&= O(N^{a-1}\big\|\widetilde{\boldsymbol{\xi}}_{J_1}-\boldsymbol{\xi}^*_{J_1}\big\|_1^2 + c_T(1+N^{a-\frac{1}{2}+\frac{1}{2w}})\big\|\widetilde{\boldsymbol{\xi}}_{J_1}-\boldsymbol{\xi}^*_{J_1}\big\|_1 \\
&\quad + c_T^2(N^{a+\frac{1}{w}} + N^{\frac{1}{2}+\frac{1}{2w}}). \qquad \text{(S.37)}
\end{aligned}
$$

We are now ready to derive a bound for $\big\|\widetilde{\boldsymbol{\xi}}_{J_1}-\boldsymbol{\xi}^*_{J_1}\big\|_1$. By Assumption R4, the block diagonal matrix $\mathbf{G}$ has full rank, which means it is positive definite. Hence for any $\boldsymbol{\alpha}\in\mathbb{R}^{N^2}$, for some constant $u>0$,

$$
0 < u^{1/2} \le \frac{\big\|\mathbf{G}^{1/2}\boldsymbol{\alpha}\big\|}{\big\|\boldsymbol{\alpha}\big\|} \le \min\Big\{\frac{\big\|\mathbf{G}^{1/2}\boldsymbol{\alpha}\big\|}{\big\|\boldsymbol{\alpha}_{J_1}\big\|}, \frac{\big\|\mathbf{G}^{1/2}\boldsymbol{\alpha}\big\|}{\big\|\boldsymbol{\alpha}_{J_1^c}\big\|}\Big\},
$$

so that defining $\boldsymbol{\alpha}=\widetilde{\boldsymbol{\xi}}-\boldsymbol{\xi}^*$, on $\mathcal{M}$ and using (S.37),

$$
\begin{aligned}
u\big\|\boldsymbol{\alpha}_{J_1}\big\|^2,\ u\big\|\boldsymbol{\alpha}_{J_1^c}\big\|^2 &\le \big\|\mathbf{G}^{1/2}\boldsymbol{\alpha}\big\|^2 \le \big\|\mathbf{G}-T^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\big\|_{\max}\big\|\boldsymbol{\alpha}\big\|_1^2 + T^{-1}\big\|\mathbf{B}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\alpha}\big\|^2 \\
&= O(c_T\big\|\boldsymbol{\alpha}\big\|_1^2 + nN^{a-1}\big\|\boldsymbol{\alpha}_{J_1}\big\|^2 + c_T n^{1/2}(1+N^{a-\frac{1}{2}+\frac{1}{2w}})\big\|\boldsymbol{\alpha}_{J_1}\big\| \\
&\quad + c_T^2(N^{a+\frac{1}{w}} + N^{\frac{1}{2}+\frac{1}{2w}}) \\
&= O(n(c_T+N^{a-1})\big\|\boldsymbol{\alpha}_{J_1}\big\|^2 + c_T n^{1/2}(1+N^{a-\frac{1}{2}+\frac{1}{2w}})\big\|\boldsymbol{\alpha}_{J_1}\big\| \\
&\quad + c_T^2(N^{a+\frac{1}{w}} + N^{\frac{1}{2}+\frac{1}{2w}}), \qquad \text{(S.38)}
\end{aligned}
$$

where the last equality sign used (S.36) and Assumption R8 that $c_T N^{1-a}=o(1)$, and the second equality sign used

$$
\begin{aligned}
\big\|\mathbf{G}-T^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\big\|_{\max} &\le N^{1-a}\big\|T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}}-\bar{\mathbf{B}}_{\boldsymbol{\gamma}})^{\mathrm{T}}\mathbf{Z} - E(T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}}-\bar{\mathbf{B}}_{\boldsymbol{\gamma}})^{\mathrm{T}}\mathbf{Z})\big\|_{\max}^2 \\
&\quad + 2\big\|T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}}-\bar{\mathbf{B}}_{\boldsymbol{\gamma}})^{\mathrm{T}}\mathbf{Z} - E(T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}}-\bar{\mathbf{B}}_{\boldsymbol{\gamma}})^{\mathrm{T}}\mathbf{Z})\big\|_{\max} \\
&\quad \cdot N^{-a}\big\|E(T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}}-\bar{\mathbf{B}}_{\boldsymbol{\gamma}})^{\mathrm{T}}\mathbf{Z})\big\|_1 \\
&= O(c_T^2 N^{1-a} + c_T N^{-a}(N^a + NT^{-1})) = O(c_T).
\end{aligned}
$$

From (S.38), there exists a constant $C>0$ such that on $\mathcal{M}$,

$$
\begin{aligned}
&(u - Cn(c_T+N^{a-1}))\big\|\boldsymbol{\alpha}_{J_1}\big\|^2 - Cc_T n^{1/2}(1+N^{a-\frac{1}{2}+\frac{1}{2w}})\big\|\boldsymbol{\alpha}_{J_1}\big\| \\
&\quad - Cc_T^2(N^{a+\frac{1}{w}} + N^{\frac{1}{2}+\frac{1}{2w}}) \le 0.
\end{aligned}
$$

By Assumption R8, we have $n(c_T+N^{a-1})=o(1)$. Hence solving the above quadratic inequality for $\big\|\boldsymbol{\alpha}_{J_1}\big\|$, we have on $\mathcal{M}$ that

$$
\big\|\boldsymbol{\alpha}_{J_1}\big\| = O(c_T(n^{1/2}+N^{\frac{1}{4}+\frac{1}{4w}}) + c_T N^{\frac{a}{2}+\frac{1}{2w}}(1+n^{1/2}N^{\frac{a-1}{2}}),
$$

This completes the proof of the second inequality of the theorem.

The rates for $\big\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\big\|_1$ and $\big\|\widetilde{\boldsymbol{\delta}}-\boldsymbol{\delta}^*\big\|_1$ are obtained by substituting $\big\|\widetilde{\boldsymbol{\xi}}-\boldsymbol{\xi}^*\big\|_1$ by $c_T N^{\frac{1}{2}+\frac{1}{2w}} + n^{1/2}\big\|\widetilde{\boldsymbol{\xi}}_{J_1}-\boldsymbol{\xi}^*_{J_1}\big\|$ in the results of Theorem 1, and finally substituting $\big\|\widetilde{\boldsymbol{\xi}}_{J_1}-\boldsymbol{\xi}^*_{J_1}\big\|$ by the rate we proved above and simplifying using Assumption R8.

To prove the theorem for $M = 0$, note that since $\widetilde{\boldsymbol{\delta}} = \boldsymbol{\delta}^* = \mathbf{0}$, the decomposition of the left hand side of (S.33) in the proof has $D_2 = D_4 = D_6 = 0$, and so (S.37) becomes

$$\frac{1}{2T}\big\|\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*)\big\|^2 \leq \big\|D_5\big\|_1 + Cc_T(c_T N^{\frac{1}{2}+\frac{1}{2w}} + 4\big\|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\big\|_1)$$
$$= O(N^{-1}\big\|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\big\|_1^2 + c_T(1 + N^{-\frac{1}{2}+\frac{1}{2w}})\big\|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\big\|_1 + c_T^2 N^{\frac{1}{2}+\frac{1}{2w}}). \quad \text{(S.39)}$$

With this, (S.38) becomes

$$u\big\|\boldsymbol{\alpha}_{J_1}\big\|^2,\ u\big\|\boldsymbol{\alpha}_{J_1^c}\big\|^2 \leq \big\|\mathbf{G}^{1/2}\boldsymbol{\alpha}\big\|^2 \leq \big\|\mathbf{G} - T^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\big\|_{\max}\big\|\boldsymbol{\alpha}\big\|_1^2 + T^{-1}\big\|\mathbf{B}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\alpha}\big\|^2$$
$$= O(c_T\big\|\boldsymbol{\alpha}\big\|_1^2 + nN^{-1}\big\|\boldsymbol{\alpha}_{J_1}\big\|^2 + c_T n^{1/2}(1 + N^{-\frac{1}{2}+\frac{1}{2w}})\big\|\boldsymbol{\alpha}_{J_1}\big\|$$
$$+ c_T^2 N^{\frac{1}{2}+\frac{1}{2w}})$$
$$= O(n(c_T + N^{-1})\big\|\boldsymbol{\alpha}_{J_1}\big\|^2 + c_T n^{1/2}(1 + N^{-\frac{1}{2}+\frac{1}{2w}})\big\|\boldsymbol{\alpha}_{J_1}\big\|$$
$$+ c_T^2 N^{\frac{1}{2}+\frac{1}{2w}} + c_T^3 N^{1+\frac{1}{w}}). \quad \text{(S.40)}$$

Hence there exists a constant $C > 0$ such that on $\mathcal{M}$,

$$(u - Cn(c_T + N^{-1}))\big\|\boldsymbol{\alpha}_{J_1}\big\|^2 - Cc_T n^{1/2}(1 + N^{-\frac{1}{2}+\frac{1}{2w}})\big\|\boldsymbol{\alpha}_{J_1}\big\| - Cc_T^3 N^{1+\frac{1}{w}}) \leq 0.$$

By Assumption M2', we have $n = O(N)$, hence there is a constant C' such that $n \leq C'N$. We need to be sure that the constant u is large enough so that $u - Cn(c_T + N^{-1}) \geq u - CC' - Cc_T n > 0$ in the above inequality. And since in the definition of $\mathbf{G}$ in Assumption R4 the constant c^2 can be chosen to be large enough such that the constant u is enlarged accordingly with $\lambda_{\min}(\mathbf{G}) \geq u > CC' > 0$, we can without loss of generality assume $u - Cn(c_T + N^{-1}) > 0$ (recall that $c_T n = o(1)$ in Assumption R8').

Then solving this inequality leads to

$$\big\|\boldsymbol{\alpha}_{J_1}\big\| = O(c_T n^{1/2}(1 + N^{-\frac{1}{2}+\frac{1}{2w}}) + c_T^{3/2} N^{\frac{1}{2}+\frac{1}{2w}}).$$

The rate of $\big\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\big\|_1$ is obtained by substituting $\big\|\widetilde{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\big\|_1$ by $c_T N^{\frac{1}{2}+\frac{1}{2w}} + n^{1/2}\big\|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\big\|$ in the corresponding result of Theorem 1, and finally substituting $\big\|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\big\|$ by the rate we proved above and simplifying using Assumption R8'. This completes the proof of the theorem. □

Proof of Theorem 3. Consider $M > 0$. We first prove the sign consistency of $\widehat{\boldsymbol{\xi}}$ in (2.8). After that, we prove the asymptotic normality of $\widehat{\boldsymbol{\xi}}_{J_1}$.

Define the set $D = J_0 \cup J_1 \cup J_2$, so that D is the set of all indices excluding those corresponding to the diagonal of $\mathbf{A}^*$. The KKT condition implies that $\widehat{\boldsymbol{\xi}}$ is a solution to the adaptive LASSO problem in (2.8) if and only if there exists a subgradient

$$\mathbf{g} = \partial(\mathbf{v}^{\mathrm{T}}|\widehat{\boldsymbol{\xi}}|) = \left\{\mathbf{g} \in \mathbb{R}^{N^2} : \begin{cases} g_i = 0, & i \in D^c; \\ g_i = v_i \mathrm{sign}(\widehat{\xi}_i), & \widehat{\xi}_i \neq 0; \\ |g_i| \leq v_i, & \text{otherwise.} \end{cases}\right\}, \quad \text{(S.41)}$$

such that differentiating with respect to $\boldsymbol{\xi}_D$, we have

$$T^{-1}\mathbf{Z}_D^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_D\widehat{\boldsymbol{\xi}} - T^{-1}\mathbf{Z}_D^{\mathrm{T}}\mathbf{B}(\mathbf{B}^{\mathrm{T}}\mathbf{y} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0\widetilde{\boldsymbol{\delta}} - \mathbf{B}^{\mathrm{T}}\mathbf{X}_{\widetilde{\boldsymbol{\beta}}}\mathrm{vec}(\mathbf{I}_N)) + \lambda_T\mathbf{g}_D = \mathbf{0},$$

where for a matrix A, A_S is the sub-matrix of A with only those columns indexed by S. Using $\mathbf{B}^{\mathrm{T}}\mathbf{y} = \mathbf{B}^{\mathrm{T}}\mathbf{Z}_D\boldsymbol{\xi}^*_D + \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0\boldsymbol{\delta}^* + \mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N) + \mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}$, the above can be written as

$$T^{-1}\mathbf{Z}_D^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_D(\widehat{\boldsymbol{\xi}}_D - \boldsymbol{\xi}^*_D) + T^{-1}\mathbf{Z}_D^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*) - T^{-1}\mathbf{Z}_D^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}$$
$$+ T^{-1}\mathbf{Z}_D^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N) = -\lambda_T\mathbf{g}_D.$$

Hence, couple with the above, there is a partial sign consistent solution $\widehat{\boldsymbol{\xi}}$ such that $\widehat{\boldsymbol{\xi}}_{J_0\cup J_2} = \mathbf{0}$ and $\mathrm{sign}(\widehat{\boldsymbol{\xi}}_{J_1}) = \mathrm{sign}(\boldsymbol{\xi}^*_{J_1})$ if and only if

$$T^{-1}\mathbf{Z}_D^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1}(\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}) - T^{-1}\mathbf{Z}_D^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}^*_{J_2}\boldsymbol{\xi}^*_{J_2} + T^{-1}\mathbf{Z}_D^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*)$$
$$- T^{-1}\mathbf{Z}_D^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon} + T^{-1}\mathbf{Z}_D^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N) = -\lambda_T \mathbf{g}_D.$$

The above can be rewritten as two, one on J_1 and one on $J' = J_0 \cup J_2$:

$$\begin{cases} T^{-1}\mathbf{Z}_{J_1}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1}(\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}) - T^{-1}\mathbf{Z}_{J_1}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}^*_{J_2}\boldsymbol{\xi}^*_{J_2} + T^{-1}\mathbf{Z}_{J_1}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*) \\ -T^{-1}\mathbf{Z}_{J_1}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon} + T^{-1}\mathbf{Z}_{J_1}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N) = -\lambda_T \mathbf{g}_{J_1}, \\ \Big| T^{-1}\mathbf{Z}_{J'}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1}(\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}) - T^{-1}\mathbf{Z}_{J'}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}^*_{J_2}\boldsymbol{\xi}^*_{J_2} + T^{-1}\mathbf{Z}_{J'}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*) \\ -T^{-1}\mathbf{Z}_{J'}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon} + T^{-1}\mathbf{Z}_{J'}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N) \Big| \leq \lambda_T \mathbf{v}_{J'}. \end{cases} \tag{S.42}$$

Note that if $J_1 = \phi$, we only need to prove the second inequality above. So we assume $J_1 \neq \phi$ for now. Then Assumption R4 on $\mathbf{G}$ implies that $\mathbf{G}_{J_1J_1}$ is invertible, and so we can write $\widehat{\boldsymbol{\xi}}_{J_1} = \boldsymbol{\xi}^*_{J_1} + \sum_{j=1}^6 I_j$, where

$$\begin{aligned} &I_1 = -\mathbf{G}^{-1}_{J_1J_1}(T^{-1}\mathbf{Z}_{J_1}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{G}_{J_1J_1})(\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}_{J_1^*}),\ I_2 = \mathbf{G}^{-1}_{J_1J_1}T^{-1}\mathbf{Z}_{J_1}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_2}\boldsymbol{\xi}^*_{J_2}, \\ &I_3 = -\mathbf{G}^{-1}_{J_1J_1}T^{-1}\mathbf{Z}_{J_1}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*),\ I_4 = -\mathbf{G}^{-1}_{J_1J_1}T^{-1}\mathbf{Z}_{J_1}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N), \\ &I_5 = \mathbf{G}^{-1}_{J_1J_1}T^{-1}\mathbf{Z}_{J_1}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon},\ I_6 = -\mathbf{G}^{-1}_{J_1J_1}\lambda_T\mathbf{g}_{J_1}. \end{aligned} \tag{S.43}$$

By Assumption M2, since the number of elements in J_1 in each row of $\mathbf{A}^*$ is bounded by a constant uniformly (say by n_r), each block in $\mathbf{G}_{J_1J_1}$ has constant size $n_r \times n_r$. Then

$$\big\|\mathbf{G}^{-1}_{J_1J_1}\big\|_\infty \leq \frac{n_r^{1/2}}{\lambda_{\min}(\mathbf{G}_{J_1J_1})} \leq \frac{n_r^{1/2}}{u} = O(1). \tag{S.44}$$

Hence we have on $\mathcal{M}$, $\big\|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\big\|_{\max} \leq \sum_{j=1}^6 \big\|I_j\big\|_{\max}$

$$\begin{aligned} \big\|I_1\big\|_{\max} &\leq \big\|\mathbf{G}^{-1}_{J_1J_1}\big\|_\infty \big\|T^{-1}\mathbf{Z}_{J_1}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{G}_{J_1J_1}\big\|_\infty \big\|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\big\|_{\max} \\ &= O(c_T\big\|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\big\|_{\max}), \\ \big\|I_2\big\|_{\max} &\leq \big\|\mathbf{G}^{-1}_{J_1J_1}\big\|_\infty N^{-a}\big\|T^{-1}\mathbf{Z}_{J_1}^{\mathrm{T}}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})\big\|_\infty \big\|T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})^{\mathrm{T}}\mathbf{Z}_{J_2}\big\|_{\max} \max_{1\leq j\leq N}\big\|\mathbf{a}^*_{j,J_2}\big\|_1 \\ &= O(\max_{1\leq j\leq N}\big\|\mathbf{a}^*_{j,J_2}\big\|_1), \\ \big\|I_3\big\|_{\max} &= O(\big\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\big\|_1) = O(c_T N^{-\frac{1}{2}+\frac{1}{2w}}), \\ \big\|I_4\big\|_{\max} &= \big\|\mathbf{G}^{-1}_{J_1J_1}\big\|_\infty N^{-a}\big\|T^{-1}\mathbf{Z}_{J_1}^{\mathrm{T}}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})\big\|_\infty \\ &\quad \cdot \big\|\mathrm{vec}(T^{-1}\sum_{t=1}^T(\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^{\mathrm{T}}\mathbf{X}_t^{\mathrm{T}})\big\|_{\max} = O(\big\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\big\|_1) = O(c_T N^{-\frac{1}{2}+\frac{1}{2w}}), \\ \big\|I_5\big\|_{\max} &= O(\max_j \big\|T^{-1}\sum_{t=1}^T(\mathbf{B}_t - \bar{\mathbf{B}})\boldsymbol{\gamma}\epsilon_{tj}\big\|_{\max}) = O(c_T), \\ \big\|I_6\big\|_{\max} &\leq \frac{\lambda_T}{\min_{j\in J_1}|\xi^*_j| - \big\|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\big\|} = O(c_T), \end{aligned} \tag{S.45}$$

where in the above, $\mathbf{a}_j^{*\mathrm{T}}$ is the jth row of $\mathbf{A}^*$. For I_6, we also used Assumption R7 for the rate of λ_T, and the fact that $\big\|\boldsymbol{\alpha}_{J_1}\big\| = \big\|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\big\| = o(1)$ by the second inequality proved in this theorem, and the rates assumed in Assumption R8.

The above implies that on $\mathcal{M}$,

$$\left\|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\right\|_{\max} = O(c_T + \max_{1\leq j\leq N} \left\|\mathbf{a}^*_{j,J_2}\right\|_1) = o(1), \tag{S.46}$$

which implies that on $\mathcal{M}$, we have $\mathrm{sign}(\widehat{\boldsymbol{\xi}}_{J_1}) = \mathrm{sign}(\boldsymbol{\xi}^*_{J_1})$. To complete the sign consistency proof, we now prove the second part of (S.42) is satisfied indeed.

To this end we have on $\mathcal{M}$,

$$\begin{aligned}
\left\|T^{-1}\mathbf{Z}^{\mathrm{T}}_{J'}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1}(\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1})\right\|_{\max} &= O(\left\|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\right\|_{\max}) = O(c_T + \max_{1\leq j\leq N} \left\|\mathbf{a}^*_{j,J_2}\right\|_1),\\
\left\|T^{-1}\mathbf{Z}^{\mathrm{T}}_{J'}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_2}\boldsymbol{\xi}^*_{J_2}\right\|_{\max} &= O(\max_{1\leq j\leq N} \left\|\mathbf{a}^*_{j,J_2}\right\|_1),\\
\left\|T^{-1}\mathbf{Z}^{\mathrm{T}}_{J'}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*)\right\|_{\max} &= O(\left\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\right\|_1),\\
\left\|T^{-1}\mathbf{Z}^{\mathrm{T}}_{J'}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N)\right\|_{\max} &= O(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_1),\\
\left\|T^{-1}\mathbf{Z}^{\mathrm{T}}_{J'}\mathbf{B}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}\right\|_{\max} &= O(c_T).
\end{aligned}$$

On the other hand, the right hand side of the second part of (S.42) has minimum value of

$$\frac{\lambda_T}{\left\|\widetilde{\boldsymbol{\xi}}_{J_2}\right\|_{\max}} \geq \frac{\lambda_T}{\left\|\boldsymbol{\xi}^*_{J_2}\right\|_{\max} + \left\|\widetilde{\boldsymbol{\xi}}_{J_1^c} - \boldsymbol{\xi}^*_{J_1^c}\right\|},$$

so that it is sufficient to prove

$$(c_T + \max_{1\leq j\leq N} \left\|\mathbf{a}^*_{j,J_2}\right\|_1)(\left\|\boldsymbol{\xi}^*_{J_2}\right\|_{\max} + \left\|\widetilde{\boldsymbol{\xi}}_{J_1^c} - \boldsymbol{\xi}^*_{J_1^c}\right\|) = o(\lambda_T). \tag{S.47}$$

Since we have rate for $\left\|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\right\|$, from (S.38), we know that on $\mathcal{M}$,

$$\left\|\widetilde{\boldsymbol{\xi}}_{J_1^c} - \boldsymbol{\xi}^*_{J_1^c}\right\| = O(\left\|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\right\|),$$

so that it is sufficient to prove that on $\mathcal{M}$,

$$\begin{aligned}
(c_T + \max_{1\leq j\leq N} &\left\|\mathbf{a}^*_{j,J_2}\right\|_1)(\left\|\boldsymbol{\xi}^*_{J_2}\right\|_{\max} + c_T(n^{1/2} + N^{\frac{1}{4}+\frac{1}{4w}})\\
&+ c_T N^{\frac{a}{2}+\frac{1}{2w}}(1 + n^{1/2}N^{\frac{a-1}{2}})) = o(\lambda_T) = o(c_T),
\end{aligned}$$

since by Assumption R7, λ_T has the same rate as c_T. But the above is indeed satisfied by Assumption R8, and hence the proof of partial sign consistency for $\widehat{\boldsymbol{\xi}}_{J_1}$ completes.

For proving the asymptotic normality result, we go back to the decomposition in (S.43), and write for a constant vector $\boldsymbol{\alpha} \in \mathbb{R}^n$ such that $\left\|\boldsymbol{\alpha}\right\|_1 < \infty$,

$$\boldsymbol{\alpha}^{\mathrm{T}}(\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1} + \mathbf{G}^{-1}_{J_1J_1}\lambda_T\mathbf{g}_{J_1}) = \sum_{j=1}^{5}\boldsymbol{\alpha}^{\mathrm{T}}I_j,$$

where each I_j is defined exactly the same as in (S.43). If we can show that $\boldsymbol{\alpha}^{\mathrm{T}}I_5$ is $T^{1/2}N^{(a-b)/2}$-convergent, then by Theorem S.1, we can actually conclude from (S.45) that

$$\begin{aligned}
|\boldsymbol{\alpha}^{\mathrm{T}}I_1| &= O_p(c_T\left\|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\right\|_{\max}) = O_p(c_T^2 + c_T\max_{1\leq j\leq N} \left\|\mathbf{a}^*_{j,J_2}\right\|_1) = o_p(T^{-1/2}N^{-(a-b)/2}),\\
|\boldsymbol{\alpha}^{\mathrm{T}}I_2| &= O_p(\max_{1\leq j\leq N} \left\|\mathbf{a}^*_{j,J_2}\right\|_1) = o_p(T^{-1/2}N^{-(a-b)/2}),\\
|\boldsymbol{\alpha}^{\mathrm{T}}I_3| &= O_p(\left\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\right\|_1) = o_p(T^{-1/2}N^{-(a-b)/2}),\\
|\boldsymbol{\alpha}^{\mathrm{T}}I_4| &= O_p(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right\|_1) = o_p(T^{-1/2}N^{-(a-b)/2}),
\end{aligned}$$

where we used Assumption R8 and the results of Theorem 2 to determine if the above rates are dominated by $T^{-1/2}N^{-(a-b)/2}$. Hence if we can show that $\boldsymbol{\alpha}^{\mathrm{T}}I_5$ is also asymptotically normal with rate $T^{-1/2}N^{-(a-b)/2}$, then asymptotic normality of $\boldsymbol{\alpha}^{\mathrm{T}}(\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1} + \mathbf{G}^{-1}_{J_1J_1}\lambda_T\mathbf{g}_{J_1})$ follows.

To this end, we decompose further

$$\begin{aligned}\boldsymbol{\alpha}^{\mathrm{T}}I_5 &= \boldsymbol{\alpha}^{\mathrm{T}}(N^{-a}\mathbf{G}^{-1}_{J_1J_1})(T^{-1}\mathbf{Z}^{\mathrm{T}}_{J_1}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}}) - E(T^{-1}\mathbf{Z}^{\mathrm{T}}_{J_1}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})))T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})^{\mathrm{T}}\boldsymbol{\epsilon} \\ &+ \boldsymbol{\alpha}^{\mathrm{T}}(N^{-a}\mathbf{G}^{-1}_{J_1J_1})E(T^{-1}\mathbf{Z}^{\mathrm{T}}_{J_1}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}}))T^{-1}(E(\mathbf{B}_{\boldsymbol{\gamma}}) - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})^{\mathrm{T}}\boldsymbol{\epsilon}, \\ &+ \boldsymbol{\alpha}^{\mathrm{T}}(N^{-a}\mathbf{G}^{-1}_{J_1J_1})E(T^{-1}\mathbf{Z}^{\mathrm{T}}_{J_1}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}}))T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}} - E(\mathbf{B}_{\boldsymbol{\gamma}}))^{\mathrm{T}}\boldsymbol{\epsilon},\end{aligned}$$

where clearly the third term on the right hand side above dominated the rest. To show asymptotic normality of this particular dominating term, we use Theorem 3(ii) of Wu (2011). Denote $\|Y\| = E^{1/2}(Y)^2$ for a random variable Y. Rewriting the term as

$$T^{-1}\sum_{t=1}^{T}\boldsymbol{\alpha}^{\mathrm{T}}(N^{-a}\mathbf{G}^{-1}_{J_1J_1})E(T^{-1}\mathbf{Z}^{\mathrm{T}}_{J_1}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}}))(\boldsymbol{\epsilon}_t \otimes (\mathbf{B}_t - \boldsymbol{\mu}_b)\boldsymbol{\gamma}),$$

it is clear that we need to show that

$$\sum_{t\geq 0}\big\|P_0(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}\boldsymbol{\epsilon}_t \otimes (\mathbf{B}_t - \boldsymbol{\mu}_b)\boldsymbol{\gamma})\big\| < \infty, \tag{S.48}$$

where $\mathbf{R} = \Big(E(T^{-1}\mathbf{Z}^{\mathrm{T}}_{J_1}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}}))E(T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}})^{\mathrm{T}}\mathbf{Z}_{J_1})\Big)^{-1}E(T^{-1}\mathbf{Z}^{\mathrm{T}}_{J_1}(\mathbf{B}_{\boldsymbol{\gamma}} - \bar{\mathbf{B}}_{\boldsymbol{\gamma}}))$, and $P_0(\cdot) = E(\cdot|\mathcal{G}_0 \cup \mathcal{H}_0) - E(\cdot|\mathcal{G}_{-1} \cup \mathcal{H}_{-1})$. Then Theorem 3(ii) of Wu (2011) implies that

$$\begin{aligned}T^{1/2}s_0^{-1/2}\boldsymbol{\alpha}^{\mathrm{T}}I_5 &= T^{-1/2}s_0^{-1/2}\sum_{t=1}^{T}\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}(\boldsymbol{\epsilon}_t \otimes (\mathbf{B}_t - \boldsymbol{\mu}_b)\boldsymbol{\gamma})(1 + o_p(1)) \\ &\xrightarrow{\mathcal{D}} N(0,1),\end{aligned}$$

where

$$\begin{aligned}s_0 &= \sum_{\tau}\mathrm{cov}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}\boldsymbol{\epsilon}_t \otimes (\mathbf{B}_t - \boldsymbol{\mu}_b)\boldsymbol{\gamma}, \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}\boldsymbol{\epsilon}_{t+\tau} \otimes (\mathbf{B}_{t+\tau} - \boldsymbol{\mu}_b)\boldsymbol{\gamma}) \\ &= \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}\Big(\sum_{\tau}E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}^{\mathrm{T}}_{t+\tau}) \otimes E((\mathbf{B}_t - \boldsymbol{\mu}_b)\boldsymbol{\gamma}\boldsymbol{\gamma}^{\mathrm{T}}(\mathbf{B}_t - \boldsymbol{\mu}_b)^{\mathrm{T}})\Big)\mathbf{R}^{\mathrm{T}}\boldsymbol{\alpha}.\end{aligned}$$

Note that

$$\|\boldsymbol{\alpha}\|^2\lambda_{\min}(\mathbf{R}\mathbf{R}^{\mathrm{T}})\lambda_{\min}(\boldsymbol{\Sigma}) \leq s_0 \leq \|\boldsymbol{\alpha}\|^2\lambda_{\max}(\mathbf{R}\mathbf{R}^{\mathrm{T}})\lambda_{\max}(\boldsymbol{\Sigma}),$$

so that

$$\|\boldsymbol{\alpha}\|^2N^{-a}\lambda_{\min}(\mathbf{G}^{-1}_{J_1J_1})\lambda_{\min}(\boldsymbol{\Sigma}) \leq s_0 \leq \|\boldsymbol{\alpha}\|^2N^{-a}\lambda_{\max}(\mathbf{G}^{-1}_{J_1J_1})\lambda_{\max}(\boldsymbol{\Sigma}).$$

Since the eigenvalues of $N^{-b}\boldsymbol{\Sigma}$ are easily seen to be uniformly bounded away from 0 and infinity by Assumption R6, we can see from the above that both sides are of order $N^{-(a-b)}$, and hence the term $\boldsymbol{\alpha}^{\mathrm{T}}I_5$ is indeed $T^{1/2}N^{(a-b)/2}$-convergent. It remains to show (S.48).

Define $E_i(\cdot) = E(\cdot|\mathcal{G}_i \cup \mathcal{H}_i)$. Then observe that

$$\begin{aligned}P_0(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}\boldsymbol{\epsilon}_t \otimes (\mathbf{B}_t - \boldsymbol{\mu}_b)\boldsymbol{\gamma}) &= \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}\Big(E_0(\boldsymbol{\epsilon}_t) \otimes E_0((\mathbf{B}_t - \boldsymbol{\mu}_b)\boldsymbol{\gamma}) - E_{-1}(\boldsymbol{\epsilon}_t) \otimes E_{-1}((\mathbf{B}_t - \boldsymbol{\mu}_b)\boldsymbol{\gamma})\Big) \\ &= \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}P_0(\epsilon_t) \otimes E_0((\mathbf{B}_t - \boldsymbol{\mu}_b)\gamma) + \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}E_{-1}(\boldsymbol{\epsilon}_t) \otimes P_0((\mathbf{B}_t - \boldsymbol{\mu}_b)\gamma).\end{aligned}$$

Hence

$$\begin{aligned}
&\big\|P_0(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}\boldsymbol{\epsilon}_t\otimes(\mathbf{B}_t-\boldsymbol{\mu}_b)\boldsymbol{\gamma})\big\| \\
&\leq \Big\{2\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}E(P_0(\boldsymbol{\epsilon}_t)P_0(\boldsymbol{\epsilon}_t)^{\mathrm{T}})\otimes E\big(E_0((\mathbf{B}_t-\boldsymbol{\mu}_b)\boldsymbol{\gamma})E_0(\boldsymbol{\gamma}^{\mathrm{T}}(\mathbf{B}_t-\boldsymbol{\mu}_b)^{\mathrm{T}})\big)\mathbf{R}^{\mathrm{T}}\boldsymbol{\alpha}\Big\}^{1/2} \\
&\quad+\Big\{2\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}E(E_{-1}(\boldsymbol{\epsilon}_t)E_{-1}(\boldsymbol{\epsilon}_t)^{\mathrm{T}})\otimes E\big(P_0((\mathbf{B}_t-\boldsymbol{\mu}_b)\boldsymbol{\gamma})P_0(\boldsymbol{\gamma}^{\mathrm{T}}(\mathbf{B}_t-\boldsymbol{\mu}_b)^{\mathrm{T}})\big)\mathbf{R}^{\mathrm{T}}\boldsymbol{\alpha}\Big\}^{1/2} \\
&\leq 2^{1/2}\big\|\boldsymbol{\alpha}\big\|_1\big\|\mathbf{R}\big\|_\infty\max_{1\leq j\leq N}\big\|P_0(\epsilon_{tj})\big\|\cdot\max_{1\leq j\leq N}\mathrm{var}^{1/2}(\mathbf{b}_{t,j}^{\mathrm{T}}\boldsymbol{\gamma}) \\
&\quad+2^{1/2}\big\|\boldsymbol{\alpha}\big\|_1\big\|\mathbf{R}\big\|_\infty\cdot\sigma_{\max}\cdot\max_{1\leq j\leq N}\big\|P_0(\mathbf{b}_{t,j}^{\mathrm{T}}\boldsymbol{\gamma})\big\| \\
&\leq 2^{1/2}\big\|\boldsymbol{\alpha}\big\|_1\big\|\mathbf{R}\big\|_\infty\max_{1\leq j\leq N}\big\|P_0(\epsilon_{tj})\big\|\cdot\sigma_{\max}\big\|\boldsymbol{\gamma}\big\|_1 \\
&\quad+2^{1/2}\big\|\boldsymbol{\alpha}\big\|_1\big\|\mathbf{R}\big\|_\infty\cdot\sigma_{\max}\cdot\max_{\substack{1\leq j\leq N\\1\leq k\leq K}}\big\|P_0(B_{t,jk})\big\|\big\|\boldsymbol{\gamma}\big\|_1 \\
&=O\big(\max_{1\leq j\leq N}\big\|P_0(\epsilon_{tj})\big\|+\max_{\substack{1\leq j\leq N\\1\leq k\leq K}}\big\|P_0(B_{t,jk})\big\|\big),
\end{aligned}$$

where the second inequality used the decomposition

$$\mathrm{var}(\cdot)=\mathrm{var}(E_i(\cdot))+E(\mathrm{var}_i(\cdot))\geq\mathrm{var}(E_i(\cdot)),$$

and the third inequality used Assumption R2, while the last equality used $\big\|\boldsymbol{\gamma}\big\|_1=1$ and $\big\|\mathbf{R}\big\|_\infty=O(1)$. Since $P_0(B_{t,sk})=P_0^b(B_{t,sk})$ and $P_0(\epsilon_{tj})=P_0^\epsilon(\epsilon_{tj})$, our assumptions (3.16) immediately implies from the above that (S.48) is satisfied. The proof of the case $M>0$ is completed by noting that a fixed dimensional multivariate version is true also by theorem 3(ii) of Wu (2011), replacing $\boldsymbol{\alpha}$ by $\mathbf{M}=(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m)^{\mathrm{T}}$.

We turns to the proof of the case $M=0$. Consider $\widehat{\boldsymbol{\xi}}$ in (3.10). The KKT condition implies that $\widehat{\boldsymbol{\xi}}$ is a solution to the adaptive LASSO problem (3.10) if and only if there exists a subgradient $\mathbf{g}$ in (S.41) such that differentiating with respect to $\boldsymbol{\xi}_D$, we have

$$T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_D-\mathbf{K}'_D)^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_D-\mathbf{K}'_D)\widehat{\boldsymbol{\xi}}_D-T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_D-\mathbf{K}'_D)^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{y}-\mathbf{K}\mathbf{y}^v)=-\lambda_T\mathbf{g}_D.$$

Since

$$\begin{aligned}
\mathbf{B}^{\mathrm{T}}\mathbf{y}&=\mathbf{B}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\xi}^*+\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N)+\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon} \\
&=\mathbf{B}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\xi}^*+N^{-a/2}T^{-1/2}\sum_{t=1}^T(\mathbf{X}_t\otimes(\mathbf{B}_t-\bar{\mathbf{B}})^{\mathrm{T}}\boldsymbol{\gamma})\boldsymbol{\beta}^*+\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}, \\
\mathbf{K}\mathbf{y}^v&=N^{-a/2}T^{-1/2}\sum_{t=1}^T(\mathbf{X}_t\otimes(\mathbf{B}_t-\bar{\mathbf{B}})^{\mathrm{T}}\boldsymbol{\gamma})(\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\mathbf{A}^{*\otimes}\mathbf{y}^v \\
&\quad+\mathbf{K}\mathbf{X}\boldsymbol{\beta}^*+\mathbf{K}\boldsymbol{\epsilon}^v \\
&=\mathbf{K}'\boldsymbol{\xi}^*+N^{-a/2}T^{-1/2}\sum_{t=1}^T(\mathbf{X}_t\otimes(\mathbf{B}_t-\bar{\mathbf{B}})^{\mathrm{T}}\boldsymbol{\gamma})\boldsymbol{\beta}^*+\mathbf{K}\boldsymbol{\epsilon}^v,
\end{aligned}$$

we have

$$\mathbf{B}^{\mathrm{T}}\mathbf{y}-\mathbf{K}\mathbf{y}^v=(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_D-\mathbf{K}'_D)\boldsymbol{\xi}^*_D+\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}-\mathbf{K}\boldsymbol{\epsilon}^v.$$

Hence $\widehat{\boldsymbol{\xi}}$ is a partial sign consistent solution to the adaptive LASSO problem (3.10) if and only if

$$\begin{aligned}
&T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_D-\mathbf{K}'_D)^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1}-\mathbf{K}'_{J_1})(\widehat{\boldsymbol{\xi}}_{J_1}-\boldsymbol{\xi}^*_{J_1})-T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_D-\mathbf{K}'_D)^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}-\mathbf{K}\boldsymbol{\epsilon}^v) \\
&-T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_D-\mathbf{K}'_D)^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_2}-\mathbf{K}'_{J_2})\boldsymbol{\xi}^*_{J_2}=-\lambda_T\mathbf{g}_D.
\end{aligned}$$

The above can be rewritten as two, one on J_1 and one on $J' = J_0 \cup J_2$:

$$\begin{cases} T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{K}'_{J_1})^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{K}'_{J_1})(\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}) - T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{K}'_{J_1})^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon} - \mathbf{K}\boldsymbol{\epsilon}^v) \\ -T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{K}'_{J_1})^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_2} - \mathbf{K}'_{J_2})\boldsymbol{\xi}^*_{J_2} = -\lambda_T \mathbf{g}_{J_1}. \\ |T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J'} - \mathbf{K}'_{J'})^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{K}'_{J_1})(\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}) - T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J'} - \mathbf{K}'_{J'})^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon} - \mathbf{K}\boldsymbol{\epsilon}^v) \\ -T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J'} - \mathbf{K}'_{J'})^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_2} - \mathbf{K}'_{J_2})\boldsymbol{\xi}^*_{J_2}| \le \lambda_T \mathbf{g}_{J'}. \end{cases} \tag{S.49}$$

Similar to the proof before, we assume $J_1 \neq \phi$, and since $\mathbf{G}_{J_1 J_1}$ is invertible by Assumption R4, we can write $\widehat{\boldsymbol{\xi}}_{J_1} = \boldsymbol{\xi}^*_{J_1} + \sum_{j=1}^4 I_j$ using (S.49), where

$$\begin{aligned} I_1 &= -\mathbf{G}^{-1}_{J_1J_1}(T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{K}'_{J_1})^{\mathrm{T}}((\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{K}'_{J_1})) - \mathbf{G}_{J_1J_1})(\widehat{\boldsymbol{\xi}}_{J_1} - \widehat{\boldsymbol{\xi}}^*_{J_1}), \\ I_2 &= \mathbf{G}^{-1}_{J_1J_1} T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{K}'_{J_1})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}, \\ I_3 &= -\mathbf{G}^{-1}_{J_1J_1} T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{K}'_{J_1})^{\mathrm{T}}\mathbf{K}\boldsymbol{\epsilon}^v, \\ I_4 &= \mathbf{G}^{-1}_{J_1J_1} T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{K}'_{J_1})^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_2} - \mathbf{K}'_{J_2})\boldsymbol{\xi}^*_{J_2}, \\ I_5 &= -\mathbf{G}^{-1}_{J_1J_1}\lambda_T \mathbf{g}_{J_1}. \end{aligned} \tag{S.50}$$

Recall from (S.28) that $\|T^{-1/2}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\|_1 = O(N^{a/2})$ on $\mathcal{M}$. Also on $\mathcal{M}$, using Assumption R5' and M2 that $n = O(N)$,

$$\begin{aligned} \|T^{-1/2}\mathbf{K}'_{J_1}\|_\infty &= O(N^{-a/2} \cdot (c_T + 1) \cdot (N^{-2}(N^2 c_T)N^{-2} + N^{-2}) \cdot (Nc_T + N) \\ &\qquad \cdot (c_T N + o(N) + c_T N)) = o(N^{-a/2}). \end{aligned} \tag{S.51}$$

$$\begin{aligned} \|T^{-1/2}\mathbf{K}'\|_1 &= O(N^{-a/2} \cdot (c_T N^2 + N) \cdot (c_T N^{-2} + N^{-2}) \cdot N) \\ &= O(N^{-a/2}). \end{aligned} \tag{S.52}$$

We have $\|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\|_{\max} \le \sum_{j=1}^5 \|I_j\|_{\max}$, where on the set $\mathcal{M}$, using (S.44), (S.51) and (S.52),

$$\begin{aligned} \|I_1\|_{\max} &\le O(\|\mathbf{G}^{-1}_{J_1J_1}\|_\infty \cdot \Big[\|T^{-1}\mathbf{Z}^{\mathrm{T}}_{J_1}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{G}_{J_1J_1}\|_\infty + \|T^{-1}\mathbf{Z}^{\mathrm{T}}_{J_1}\mathbf{B}\mathbf{K}'_{J_1}\|_\infty \\ &\qquad \|T^{-1}\mathbf{K}'^{\mathrm{T}}_{J_1}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1}\|_\infty + \|T^{-1}\mathbf{K}'^{\mathrm{T}}_{J_1}\mathbf{K}'_{J_1}\|_\infty\Big] \cdot \|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\|_{\max}) \\ &= O((c_T + N^{a/2}o(N^{-a/2}) + N^{-a/2} \cdot N^{-a/2} + N^{-a/2}o(N^{-a/2})) \cdot \|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\|_{\max}) \\ &= o(\|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\|_{\max}), \\ \|I_2\|_{\max} & O((\|T^{-1/2}\mathbf{Z}^{\mathrm{T}}_{J_1}\mathbf{B}\|_\infty + \|T^{-1/2}\mathbf{K}'^{\mathrm{T}}_{J_1}\|_\infty) \cdot \|T^{-1/2}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}\|_{\max}) \\ &= O((N^{a/2} + N^{-a/2})N^{-a/2} \cdot c_T) = O(c_T), \\ \|I_3\|_{\max} &= O((\|T^{-1/2}\mathbf{Z}^{\mathrm{T}}_{J_1}\mathbf{B}\|_\infty + \|T^{-1/2}\mathbf{K}'^{\mathrm{T}}_{J_1}\|_\infty) \cdot \|T^{-1/2}\mathbf{K}\boldsymbol{\epsilon}^v\|_{\max}) \\ &= O((N^{a/2} + N^{-a/2})c_T N^{-1/2-a/2+1/(2w)}) = O(c_T N^{-1/2+1/(2w)}), \\ \|I_4\|_{\max} &= O(\|T^{-1/2}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{K}'_{J_1})^{\mathrm{T}}\|_\infty(\|T^{-1/2}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_2}\boldsymbol{\xi}_{J_2}\|_{\max} + \|T^{-1/2}\mathbf{K}'_{J_2}\|_\infty\|\boldsymbol{\xi}^*_{J_2}\|_{\max})) \\ &= O((N^{a/2} + N^{-a/2})(N^{-a/2}\max_{1\le j\le N}\|\mathbf{a}^*_{j,J_2}\|_1 + N^{-a/2}\|\boldsymbol{\xi}^*_{J_2}\|_{\max})) \\ &= O(\max_{1\le j\le N}\|\mathbf{a}^*_{j,J_2}\|_1), \\ \|I_5\|_{\max} &\le \frac{\lambda_T}{\min_{j\in J_1}|\xi^*_j| - \|\widetilde{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\|} = O(\lambda_T) = O(c_T), \end{aligned}$$

where the rate of I_3 uses the rate on $\mathcal{A}_3$. Combining the above, we have on $\mathcal{M}$,

$$\|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\|_{\max} = O(c_T + \max_{1\le j\le N}\|\mathbf{a}^*_{j,J_2}\|_1) = o(1). \tag{S.53}$$

This means that we have $\text{sign}(\widehat{\boldsymbol{\xi}}_{J_1}) = \text{sign}(\boldsymbol{\xi}^*_{J_1})$ on $\mathcal{M}$. To complete the sign consistency proof, we now prove the second part of (S.49). On $\mathcal{M}$, using (S.51) and (S.52),

$$\begin{aligned}
&\big\|T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J'} - \mathbf{K}_{J'})^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1} - \mathbf{K}'_{J_1})(\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1})\big\|_{\max} \\
&= O(\big\|T^{-1/2}\mathbf{Z}^{\mathrm{T}}_{J'}\mathbf{B}\big\|_{\infty}\big\|T^{-1/2}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1}\big\|_{\infty} + \big\|T^{-1}\mathbf{Z}^{\mathrm{T}}_{J'}\mathbf{B}\mathbf{K}'_{J_1}\big\|_{\infty} + \big\|T^{-1}\mathbf{K}'^{\mathrm{T}}_{J'}\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_1}\big\|_{\infty} \\
&\quad + \big\|T^{-1}\mathbf{K}'^{\mathrm{T}}_{J'}\mathbf{K}'_{J_1}\big\|_{\infty}) \cdot O(\big\|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\big\|_{\max}) \\
&= O(N^{a/2}N^{-a/2} + N^{a/2}o(N^{-a/2}) + N^{-a/2}N^{-a/2} + N^{-a/2}o(N^{-a/2})) \\
&\quad \cdot O(c_T + \max_{1\le j\le N}\big\|\mathbf{a}^*_{j,J_2}\big\|_1) = O(c_T + \max_{1\le j\le N}\big\|\mathbf{a}^*_{j,J_2}\big\|_1), \\
&\big\|T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J'} - \mathbf{K}_{J'})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}\big\|_{\max} = O(c_T), \\
&\big\|T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J'} - \mathbf{K}_{J'})^{\mathrm{T}}\mathbf{K}\boldsymbol{\epsilon}^v\big\|_{\max} = O(c_T N^{-1/2+1/(2w)}), \\
&\big\|T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J'} - \mathbf{K}'_{J'})^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}_{J_2} - \mathbf{K}'_{J_2})\boldsymbol{\xi}^*_{J_2}\big\|_{\max} = O(\max_{1\le j\le N}\big\|\mathbf{a}^*_{j,J_2}\big\|_1).
\end{aligned}$$

On the other hand, the right hand side of the second part of (S.49) has a minimum value of

$$\frac{\lambda_T}{\big\|\widetilde{\boldsymbol{\xi}}_{J'}\big\|_{\max}} \ge \frac{\lambda_T}{\big\|\boldsymbol{\xi}^*_{J_2}\big\|_{\max} + \big\|\widetilde{\boldsymbol{\xi}}_{J_1^c} - \boldsymbol{\xi}^*_{J_1^c}\big\|_{\max}},$$

so that it is sufficient to prove

$$(c_T + \max_{1\le j\le N}\big\|\mathbf{a}^*_{j,J_2}\big\|_1)(\big\|\boldsymbol{\xi}^*_{J_2}\big\|_{\max} + \big\|\widetilde{\boldsymbol{\xi}}_{J_1^c} - \boldsymbol{\xi}^*_{J_1^c}\big\|) = o(\lambda_T).$$

But this is exactly the same condition as (S.47) and is proved in the first part of the proof for $M > 0$. Hence we have established the partial sign consistency of the solution $\widehat{\boldsymbol{\xi}}$ for (3.10).

Finally, for asymptotic normality of $\widehat{\boldsymbol{\xi}}$ in (3.10), consider $\boldsymbol{\alpha} \in \mathbb{R}^n$ such that $\big\|\boldsymbol{\alpha}\big\|_1 < \infty$,

$$\boldsymbol{\alpha}^{\mathrm{T}}(\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1} + \mathbf{G}^{-1}_{J_1J_1}\lambda_T\mathbf{g}_{J_1}) = \sum_{j=1}^4 \boldsymbol{\alpha}^{\mathrm{T}}I_j,$$

where each I_j is defined exactly the same as in (S.50). Then since $|\boldsymbol{\alpha}^{\mathrm{T}}I_j| \le \big\|\boldsymbol{\alpha}\big\|_1\big\|I_j\big\|_{\max} = O(\big\|I_j\big\|_{\max})$, we have on $\mathcal{M}$,

$$\begin{aligned}
&|\boldsymbol{\alpha}^{\mathrm{T}}I_1| = o(\big\|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\big\|_{\max}) = O(c_T + \max_{1\le j\le N}\big\|\mathbf{a}^*_{j,J_2}\big\|_1), \\
&|\boldsymbol{\alpha}^{\mathrm{T}}I_2| = O(c_T),\ \ |\boldsymbol{\alpha}^{\mathrm{T}}I_3| = O(c_T N^{-1/2+1/(2w)}),\ \ |\boldsymbol{\alpha}^{\mathrm{T}}I_4| = O(\max_{1\le j\le N}\big\|\mathbf{a}^*_{j,J_2}\big\|_1).
\end{aligned}$$

Hence $\boldsymbol{\alpha}^{\mathrm{T}}I_2$ is the dominating term, which can be further decomposed as $I_2 = I_{2,1} + I_{2,2}$, where

$$I_{2,1} = \mathbf{G}^{-1}_{J_1J_1}T^{-1}\mathbf{Z}^{\mathrm{T}}_{J_1}\mathbf{B}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}, \quad I_{2,2} = -\mathbf{G}^{-1}_{J_1J_1}T^{-1}\mathbf{K}'^{\mathrm{T}}_{J_1}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}.$$

On $\mathcal{M}$, we have

$$|\boldsymbol{\alpha}^{\mathrm{T}}I_{2,2}| = O(\big\|T^{-1/2}\mathbf{K}'_{J_1}\big\|_1 \cdot \big\|T^{-1/2}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}\big\|_{\max}) = O(N^{-a/2}N^{-a/2}c_T) = O(N^{-a}c_T) = o(c_T).$$

Hence $I_{2,1}$ is dominating, and the asymptotic normality of $\boldsymbol{\alpha}^{\mathrm{T}}I_{2,1}$ is exactly the same as proved before for the case $M > 0$ (the treatment of $\boldsymbol{\alpha}^{\mathrm{T}}I_5$ where I_5 is defined in (S.43)). This completes the proof of the theorem. □

Proof of Theorem 4. Consider $M > 0$. For $\boldsymbol{\alpha} \in \mathbb{R}^K$ such that $\|\boldsymbol{\alpha}\| = 1$, we can decompose $\boldsymbol{\alpha}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = \sum_{j=1}^4 \boldsymbol{\alpha}^{\mathrm{T}} I_j$, where

$$
\begin{aligned}
I_1 &= \big(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t)\big)^{-1}\Big(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t) - T^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\mathbf{X}\Big)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*),\\
I_2 &= \big(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t)\big)^{-1} \cdot T^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\boldsymbol{\epsilon}^v,\\
I_3 &= \big(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t)\big)^{-1} \cdot T^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\Big((\mathbf{A}^{*\otimes} - \widehat{\mathbf{A}}^{\otimes}) + \sum_{i=1}^M (\delta_i^* - \widetilde{\delta}_i)\mathbf{W}_{0i}^{\otimes}\Big)\boldsymbol{\Pi}^{*\otimes}\mathbf{X}\boldsymbol{\beta}^*,\\
I_4 &= \big(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t)\big)^{-1} \cdot T^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v\mathbf{B}^{v\mathrm{T}}\Big((\mathbf{A}^{*\otimes} - \widehat{\mathbf{A}}^{\otimes}) + \sum_{i=1}^M (\delta_i^* - \widetilde{\delta}_i)\mathbf{W}_{0i}^{\otimes}\Big)\boldsymbol{\Pi}^{*\otimes}\boldsymbol{\epsilon}^v,
\end{aligned}
$$

Similar to the treatment of I_1 in the proof of Theorem 1, for the I_1 defined above, through Theorem S.1, we have

$$|\boldsymbol{\alpha}^{\mathrm{T}} I_1| \le \|\boldsymbol{\alpha}\|_{\max}\|I_1\|_1 = O_p(c_T\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1).$$

For I_3 and I_4, similar to the treatment of the respective terms in the proof of Theorem 1, through Theorem S.1, we have

$$
\begin{aligned}
|\boldsymbol{\alpha}^{\mathrm{T}} I_3| &= O_p(\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|_1 + N^{-1}\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1),\\
|\boldsymbol{\alpha}^{\mathrm{T}} I_4| &= O_p(c_T(\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|_1 + N^{-1}\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1)).
\end{aligned}
$$

Using the sign consistency of $\widehat{\boldsymbol{\xi}}$ and (S.46), and the rate for $\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|_1$ from Theorem 2, we have through Theorem S.1 that

$$
\begin{aligned}
|\boldsymbol{\alpha}^{\mathrm{T}} I_3| &= O_p(c_T N^{-\frac{1}{2}+\frac{1}{2w}} + N^{-1}n(c_T + \max_{1\le j\le N}\|\mathbf{a}_{j,J_2}^*\|_1)) = O_p(c_T N^{-\frac{1}{2}+\frac{1}{2w}})\\
&= o_p(T^{-1/2}N^{-(1-b)/2}),
\end{aligned}
$$

where we used Assumption R8 to conclude with the last line above. Hence if we can show that $\boldsymbol{\alpha}^{\mathrm{T}} I_2$ is $T^{1/2}N^{(1-b)/2}$-convergent, then the remaining task is to show the asymptotic normality of $\boldsymbol{\alpha}^{\mathrm{T}} I_2$ as presented in the theorem.

We are going to show the asymptotic normality of $\boldsymbol{\alpha}^{\mathrm{T}} I_2$ now and find its rate of convergence. To this end, we further decompose

$$
\begin{aligned}
\boldsymbol{\alpha}^{\mathrm{T}} I_2 = {} & \boldsymbol{\alpha}^{\mathrm{T}}\big(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t)\big)^{-1}\big(T^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v - E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)\big)T^{-1}\mathbf{B}^{v\mathrm{T}}\boldsymbol{\epsilon}^v\\
& + \boldsymbol{\alpha}^{\mathrm{T}}\big(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t)\big)^{-1}E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)T^{-1}\mathbf{B}^{v\mathrm{T}}\boldsymbol{\epsilon}^v,
\end{aligned}
$$

with the second term clearly dominating. Rewriting this dominating term as

$$T^{-1}\sum_{t=1}^T \boldsymbol{\alpha}^{\mathrm{T}}\big(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t)\big)^{-1}E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)\mathbf{B}_t^{\mathrm{T}}\boldsymbol{\epsilon}_t,$$

if we can show that, defining $\mathbf{R}_0 = \big(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t)\big)^{-1}E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)$,

$$\sum_{t\ge 0}\|P_0(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_0\mathbf{B}_t^{\mathrm{T}}\boldsymbol{\epsilon}_t)\| \le \infty, \tag{S.54}$$

then by Theorem 3(ii) of Wu (2011),

$$T^{1/2}s_0^{-1/2}\boldsymbol{\alpha}I_2 = T^{-1/2}s_0^{-1/2}\sum_{t=1}^T \boldsymbol{\alpha}^T\mathbf{R}_0\mathbf{B}_t^{\mathrm{T}}\boldsymbol{\epsilon}_t(1 + o_p(1)) \xrightarrow{\mathcal{D}} N(0,1),$$

where

$$s_0 = \sum_\tau \mathrm{cov}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_0\mathbf{B}_t^{\mathrm{T}}\boldsymbol{\epsilon}_t, \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_0\mathbf{B}_{t+\tau}^{\mathrm{T}}\boldsymbol{\epsilon}_{t+\tau}) = \sum_\tau \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_0 E(\mathbf{B}_t^{\mathrm{T}}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t+\tau}^{\mathrm{T}}\mathbf{B}_{t+\tau})\mathbf{R}_0^{\mathrm{T}}\boldsymbol{\alpha}.$$

To determine the rate of s_0, consider the (k,k) element of $\sum_\tau E(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t+\tau}^{\mathrm{T}}\mathbf{B}_{t+\tau,k})$,

$$\begin{aligned}\sum_\tau E(\mathbf{B}_{t,k}^{\mathrm{T}}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t+\tau}^{\mathrm{T}}\mathbf{B}_{t+\tau,k}) &= \sum_\tau \mathrm{tr}\big(E(\mathbf{B}_{t+\tau,k}\mathbf{B}_{t,k}^{\mathrm{T}})E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t+\tau}^{\mathrm{T}})\big)\\ &= \sum_\tau \mathrm{tr}\big(\mathrm{cov}(\mathbf{B}_{t+\tau,k},\mathbf{B}_{t,k})\mathrm{cov}(\boldsymbol{\epsilon}_t,\boldsymbol{\epsilon}_{t+\tau})\big) + \sum_\tau \mathrm{tr}\big(\boldsymbol{\mu}_{b,k}\boldsymbol{\mu}_{b,k}^{\mathrm{T}}\mathrm{cov}(\boldsymbol{\epsilon}_t,\boldsymbol{\epsilon}_{t+\tau})\big).\end{aligned}$$

By Assumption R6, the left term in the last line above has order N^{1+b} exactly, while the right term is bounded by

$$\sum_\tau \boldsymbol{\mu}_{b,k}^{\mathrm{T}}\mathrm{cov}(\boldsymbol{\epsilon}_t,\boldsymbol{\epsilon}_{t+\tau})\boldsymbol{\mu}_{b,k} \le \lambda_{\max}\Big(\sum_\tau \mathrm{cov}(\epsilon_t,\epsilon_{t+\tau})\Big)\big\|\boldsymbol{\mu}_{b,k}\big\|^2 = O(\big\|\boldsymbol{\mu}_{b,k}\big\|^2) = O(N).$$

Hence since $\sum_\tau E(\mathbf{B}_t^{\mathrm{T}}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t+\tau}^{\mathrm{T}}\mathbf{B}_{t+\tau})$ is of size $K\times K$ which is finite, the order of the eigenvalues of this matrix is exactly N^{1+b}. It means that, since

$$\begin{aligned}\big\|\boldsymbol{\alpha}\big\|^2\lambda_{\min}(\mathbf{R}_0\mathbf{R}_0^{\mathrm{T}})\lambda_{\min}\big(\sum_\tau E(\mathbf{B}_t^{\mathrm{T}}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t+\tau}^{\mathrm{T}}\mathbf{B}_{t+\tau})\big) \le s_0 \le \big\|\boldsymbol{\alpha}\big\|^2\lambda_{\max}(\mathbf{R}_0\mathbf{R}_0^{\mathrm{T}})\\ \cdot\lambda_{\max}\big(\sum_\tau E(\mathbf{B}_t^{\mathrm{T}}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t+\tau}^{\mathrm{T}}\mathbf{B}_{t+\tau})\big),\end{aligned}$$

and $\mathbf{R}_0\mathbf{R}_0^{\mathrm{T}} = (E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t))^{-1}$ has all K eigenvalues of order N^{-2}, the order for s_0 is exactly $N^{-(1-b)}$. It means also that $\boldsymbol{\alpha}^{\mathrm{T}}I_2$ is indeed $T^{1/2}N^{(1-b)/2}$-convergent. It remains to show (S.54).

Since we can decompose

$$P_0(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_0\mathbf{B}_t^{\mathrm{T}}\boldsymbol{\epsilon}_t) = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_0P_0(\mathbf{B}_t^{\mathrm{T}})E_0(\boldsymbol{\epsilon}_t) + \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_0E_{-1}(\mathbf{B}_t^{\mathrm{T}})P_0(\boldsymbol{\epsilon}_t),$$

we have $\big\|P_0(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_0\mathbf{B}_t^{\mathrm{T}}\boldsymbol{\epsilon}_t)\big\| \le K_{1,t}+K_{2,t}$, where

$$\begin{aligned}K_{1,t}^2 &= E(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_0P_0(\mathbf{B}_t^{\mathrm{T}})E_0(\boldsymbol{\epsilon}_t)E_0(\boldsymbol{\epsilon}_t^{\mathrm{T}})P_0(B_t)\mathbf{R}_0^{\mathrm{T}}\boldsymbol{\alpha})\\ &\le \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_0E(P_0(\mathbf{B}_t^{\mathrm{T}})P_0(\mathbf{B}_t))\mathbf{R}_0^{\mathrm{T}}\boldsymbol{\alpha}\cdot E\big(\lambda_{\max}(E_0(\boldsymbol{\epsilon}_t)E_0(\boldsymbol{\epsilon}_t^{\mathrm{T}}))\big)\\ &\le \big\|\boldsymbol{\alpha}\big\|_1^2\big\|\mathbf{R}_0\big\|_\infty^2\max_{1\le k\le K}E(P_0(\mathbf{B}_{t,k}^{\mathrm{T}})P_0(\mathbf{B}_{t,k}))\cdot E(E_0(\boldsymbol{\epsilon}_t^{\mathrm{T}})E_0(\boldsymbol{\epsilon}_t))\\ &= O(N^{-1}\max_{1\le k\le K}E(P_0(\mathbf{B}_{t,k}^{\mathrm{T}})P_0(\mathbf{B}_{t,k}))\cdot E(N^{-1}E_0(\boldsymbol{\epsilon}_t^{\mathrm{T}})E_0(\boldsymbol{\epsilon}_t)))\\ &= O(\max_{1\le k\le K}\max_{1\le s\le N}\big\|P_0(B_{t,sk})\big\|^2\cdot\max_{1\le j\le N}E(E_0^2(\epsilon_{t,j})))\\ &= O(\max_{1\le k\le K}\max_{1\le s\le N}\big\|P_0^b(B_{t,sk})\big\|^2\cdot\sigma_{\max}^2),\end{aligned}$$

so that $\sum_{t\ge0}K_{1,t}<\infty$ by our assumption $\sum_{t\ge0}\max_{1\le k\le K}\max_{1\le s\le N}\big\|P_0^b(B_{t,sk})\big\|<\infty$.

Similarly, we have

$$\begin{aligned}K_{2,t}^2 &= E(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_0E_{-1}(\mathbf{B}_t^{\mathrm{T}})P_0(\boldsymbol{\epsilon}_t)P_0(\boldsymbol{\epsilon}_t^{\mathrm{T}})E_{-1}(B_t)\mathbf{R}_0^{\mathrm{T}}\boldsymbol{\alpha})\\ &\le \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_0E(E_{-1}(\mathbf{B}_t^{\mathrm{T}})E_{-1}(\mathbf{B}_t))\mathbf{R}_0^{\mathrm{T}}\boldsymbol{\alpha}\cdot E(\lambda_{\max}(P_0(\boldsymbol{\epsilon}_t)P_0(\boldsymbol{\epsilon}_t^{\mathrm{T}})))\\ &\le \big\|\boldsymbol{\alpha}\big\|_1^2\big\|\mathbf{R}_0\big\|_\infty^2\max_{1\le k\le K}E(E_{-1}(\mathbf{B}_{t,k}^{\mathrm{T}})E_{-1}(\mathbf{B}_{t,k}))E(P_0(\boldsymbol{\epsilon}_t^{\mathrm{T}})P_0(\boldsymbol{\epsilon}_t))\\ &= O(\max_{1\le k\le K}\max_{1\le s\le N}E(E_{-1}^2(B_{t,sk}))\cdot\max_{1\le j\le N}\big\|P_0(\epsilon_{t,j})\big\|^2)\\ &= O((\sigma_{\max}^2+\max_{t,s}\mu_{b,ts}^2)\cdot\max_{1\le j\le N}\big\|P_0(\epsilon_{t,j})\big\|^2)\\ &= O(\max_{1\le j\le N}\big\|P_0^\epsilon(\epsilon_{t,j})\big\|^2),\end{aligned}$$

so that $\sum_{t\geq 0} K_{2,t} < \infty$ by our assumption of $\sum_{t\geq 0}\max_{1\leq j\leq N}\|P_0^\epsilon(\epsilon_{t,j})\| < \infty$. This completes the proof of the theorem, by noting that extension to multivariate case is straight forward and use the same Theorem 3(ii) of Wu (2011).

Now consider the proof when $M = 0$. By Assumption M2', we have $n = O(N)$, so that for $\boldsymbol{\alpha} \in \mathbb{R}^K$ with $\|\boldsymbol{\alpha}\| = 1$, using the same definitions of I_1 to I_4 at the beginning of the proof of this theorem (obviously setting $\widetilde{\boldsymbol{\delta}} = \boldsymbol{\delta}^* = \mathbf{0}$),

$$|\boldsymbol{\alpha}^{\mathrm{T}} I_3| = O_p(N^{-1}\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1) = O_p(N^{-1}n(c_T + \max_{1\leq j\leq N}\|\mathbf{a}^*_{j,J_2}\|_1)) = O_p(c_T + \max_{1\leq j\leq N}\|\mathbf{a}^*_{j,J_2}\|_1).$$

This makes $\boldsymbol{\alpha}^{\mathrm{T}} I_3$ the dominating term. Write, on $\mathcal{M}$,

$$\begin{aligned}
\boldsymbol{\alpha}^{\mathrm{T}} I_3 &= \boldsymbol{\alpha}^{\mathrm{T}}(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t))^{-1}\cdot T^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{B}^v \cdot T^{-1}\mathbf{B}^{v\mathrm{T}}(\mathbf{A}^{*\otimes} - \widehat{\mathbf{A}}^{\otimes})\boldsymbol{\Pi}^{*\otimes}\mathbf{X}\boldsymbol{\beta}^* \\
&= \boldsymbol{\alpha}^{\mathrm{T}}(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t))^{-1}\cdot E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t) \\
&\quad \cdot T^{-1}\sum_{t=1}^{T}(\mathbf{B}_t - \bar{\mathbf{B}})^{\mathrm{T}}\otimes \boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_t^{\mathrm{T}}\boldsymbol{\Pi}^{*\mathrm{T}}(\boldsymbol{\xi}^* - \widehat{\boldsymbol{\xi}})(1 + o(1)) \\
&= \boldsymbol{\alpha}^{\mathrm{T}}(E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)E(\mathbf{B}_t^{\mathrm{T}}\mathbf{X}_t))^{-1}\cdot E(\mathbf{X}_t^{\mathrm{T}}\mathbf{B}_t)\cdot E(\mathbf{B}_t^{\mathrm{T}}\otimes\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_t^{\mathrm{T}}\boldsymbol{\Pi}^{*\mathrm{T}})(\boldsymbol{\xi}^* - \widehat{\boldsymbol{\xi}})(1 + o(1)) \\
&= \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0(\boldsymbol{\xi}^* - \widehat{\boldsymbol{\xi}})(1 + o(1)) \\
&= (\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0(\boldsymbol{\xi}^*_{J_1} - \widehat{\boldsymbol{\xi}}_{J_1}) + \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0\boldsymbol{\xi}^*_{J_2})(1 + o(1)) \\
&= (\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0(\boldsymbol{\xi}^*_{J_1} - \widehat{\boldsymbol{\xi}}_{J_1} - \mathbf{G}^{-1}_{J_1J_1}\lambda_T\mathbf{g}_{J_1}) + \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0\mathbf{G}^{-1}_{J_1J_1}\lambda_T\mathbf{g}_{J_1} + \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0\boldsymbol{\xi}^*_{J_2})(1 + o(1)).
\end{aligned}$$

Finally, since $\|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0\|_\infty = O(1)$, we can apply the asymptotic normality in Theorem 3 for $M = 0$ to arrive at

$$\begin{aligned}
&T^{1/2}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0\mathbf{R}\boldsymbol{\Sigma}\mathbf{R}^{\mathrm{T}}\mathbf{K}_0^{\mathrm{T}}\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \\
&= T^{1/2}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0\mathbf{R}\boldsymbol{\Sigma}\mathbf{R}^{\mathrm{T}}\mathbf{K}_0^{\mathrm{T}}\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0(\boldsymbol{\xi}^*_{J_1} - \widehat{\boldsymbol{\xi}}_{J_1} - \mathbf{G}^{-1}_{J_1J_1}\lambda_T\mathbf{g}_{J_1})(1 + o_p(1)) \\
&\quad + T^{1/2}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0\mathbf{R}\boldsymbol{\Sigma}\mathbf{R}^{\mathrm{T}}\mathbf{K}_0^{\mathrm{T}}\boldsymbol{\alpha})^{-1/2}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0\mathbf{G}^{-1}_{J_1J_1}\lambda_T\mathbf{g}_{J_1} + \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0\boldsymbol{\xi}^*_{J_2})(1 + o_p(1)).
\end{aligned}$$

Since the last term in the above is of order $o_p(1)$ by using Assumption R8', and the first term is asymptotically standard normal by using Theorem 3 when $M = 0$ and the property that $\|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{K}_0\|_\infty = O(1)$, the proof completes by noting that $\boldsymbol{\alpha}$ can be replaced by a matrix $\mathbf{M} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m)^{\mathrm{T}}$ where m is finite and $\|\boldsymbol{\alpha}_i\|_1 < \infty$. □

Proof of Theorem 5. By the KKT condition, there exists a solution $\widehat{\boldsymbol{\delta}}$ to (3.8) if and only if there exists a subgradient

$$\mathbf{h} = \partial(\mathbf{u}^{\mathrm{T}}|\widehat{\boldsymbol{\delta}}|) = \left\{h \in \mathbb{R}^M : \begin{cases} h_i = u_i\mathrm{sign}(\widehat{\delta}_i), & \widehat{\delta}_i \neq 0; \\ |h_i| \leq u_i, & \text{otherwise.}\end{cases}\right\},$$

such that differentiating the expression in (3.8) with respect to $\boldsymbol{\delta}$, we have

$$T^{-1}(\mathbf{H} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0)^{\mathrm{T}}(\mathbf{H} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0)\boldsymbol{\delta} - T^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0 - \mathbf{H})^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{y} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\widehat{\boldsymbol{\xi}} - \widehat{\mathbf{h}}) = -\lambda_T'\mathbf{h}.$$

By $\mathbf{B}^{\mathrm{T}}\mathbf{y} = \mathbf{B}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\xi}^* + \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0\boldsymbol{\delta}^* + \mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N) + \mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}$, the above can be written as

$$\begin{aligned}
&T^{-1}(\mathbf{H} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0)^{\mathrm{T}}(\mathbf{H} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0)(\boldsymbol{\delta} - \boldsymbol{\delta}^*) + T^{-1}(\mathbf{H} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0)^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\boldsymbol{\xi}^* - \widehat{\boldsymbol{\xi}}) \\
&+ T^{-1}(\mathbf{H} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0)^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N) + \mathbf{H}\boldsymbol{\delta}^* - \widehat{\mathbf{h}}) + T^{-1}(\mathbf{H} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_0)^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon} = -\lambda_T'\mathbf{h}.
\end{aligned}$$

Noting that $\mathbf{H}\boldsymbol{\delta}^* - \widehat{\mathbf{h}} = -\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}(\widehat{\boldsymbol{\xi}},\boldsymbol{\delta}^*)}\mathrm{vec}(\mathbf{I}_N)$, we can conclude that there exists a sign-consistent solution $\widehat{\boldsymbol{\delta}}$ if and only if

$$\begin{cases} T^{-1}(\mathbf{H}_H - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H})^{\mathrm{T}}(\mathbf{H}_H - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H})(\widehat{\boldsymbol{\delta}}_H - \boldsymbol{\delta}_H^*) + T^{-1}(\mathbf{H}_H - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\boldsymbol{\xi}^* - \widehat{\boldsymbol{\xi}}) \\ +T^{-1}(\mathbf{H}_H - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}^* - \boldsymbol{\beta}(\widehat{\boldsymbol{\xi}},\boldsymbol{\delta}^*)}\mathrm{vec}(\mathbf{I}_N) + T^{-1}(\mathbf{H}_H - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon} = -\lambda_T'\mathbf{h}_H, \\ \Big|T^{-1}(\mathbf{H}_{H^c} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H^c})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\boldsymbol{\xi}^* - \widehat{\boldsymbol{\xi}}) + T^{-1}(\mathbf{H}_{H^c} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H^c})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}^* - \boldsymbol{\beta}(\widehat{\boldsymbol{\xi}},\boldsymbol{\delta}^*)}\mathrm{vec}(\mathbf{I}_N) \\ +T^{-1}(\mathbf{H}_{H^c} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H^c})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}\Big| \le \lambda_T'\mathbf{h}_{H^c}, \end{cases} \tag{S.55}$$

where $H = \{j : \delta_j^* \neq 0\}$. From the first equation above, we can decompose $\widehat{\boldsymbol{\delta}}_H = \boldsymbol{\delta}_H^* + \sum_{j=1}^5 I_j$, where

$$\begin{aligned} I_1 &= -(N^{-a}(\mathbf{H}_{20} - \mathbf{H}_{10})_H^{\mathrm{T}}(\mathbf{H}_{20} - \mathbf{H}_{10})_H)^{-1}\Big(T^{-1}(\mathbf{H}_H - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H})^{\mathrm{T}}(\mathbf{H}_H - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H}) \\ &\quad - N^{-a}(\mathbf{H}_{20} - \mathbf{H}_{10})_H^{\mathrm{T}}(\mathbf{H}_{20} - \mathbf{H}_{10})_H\Big)(\widehat{\boldsymbol{\delta}}_H - \boldsymbol{\delta}_H^*), \\ I_2 &= (N^{-a}(\mathbf{H}_{20} - \mathbf{H}_{10})_H^{\mathrm{T}}(\mathbf{H}_{20} - \mathbf{H}_{10})_H)^{-1}T^{-1}(\mathbf{H}_H - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}^*), \\ I_3 &= (N^{-a}(\mathbf{H}_{20} - \mathbf{H}_{10})_H^{\mathrm{T}}(\mathbf{H}_{20} - \mathbf{H}_{10})_H)^{-1}T^{-1}(\mathbf{H}_H - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}(\widehat{\boldsymbol{\xi}},\boldsymbol{\delta}^*) - \boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N), \\ I_4 &= -(N^{-a}(\mathbf{H}_{20} - \mathbf{H}_{10})_H^{\mathrm{T}}(\mathbf{H}_{20} - \mathbf{H}_{10})_H)^{-1}\lambda_T'\mathbf{h}_H, \\ I_5 &= -(N^{-a}(\mathbf{H}_{20} - \mathbf{H}_{10})_H^{\mathrm{T}}(\mathbf{H}_{20} - \mathbf{H}_{10})_H)^{-1}T^{-1}(\mathbf{H}_H - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}. \end{aligned}$$

By the same treatment of F_1 (see (S.23)) in the proof of Theorem 1, using Theorem S.1,

$$\|I_1\|_{\max} = o_p(c_T N^{1-a}\|\widehat{\boldsymbol{\delta}}_H - \boldsymbol{\delta}_H^*\|_{\max}).$$

Similarly, using Theorem S.1, the same treatments of F_2 and F_3 (see (S.24) and (S.25)) in the proof of Theorem 1 lead to

$$\|I_2\|_{\max} = O_p(N^{-1}\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\|_1) = O_p(N^{-1}(\|\boldsymbol{\xi}_{J_2}^*\|_1 + c_T n)),$$

where we used the partial sign consistency property of $\widehat{\boldsymbol{\xi}}$ and (S.46).

Similarly, we also have

$$\|I_4\|_{\max} = O(c_T N^{-1}).$$

We now show that for $\boldsymbol{\alpha} \in \mathbb{R}^{|H|}$ with $\|\boldsymbol{\alpha}\|_1 \le c < \infty$, both $\boldsymbol{\alpha}^{\mathrm{T}}I_3$ and $\boldsymbol{\alpha}^{\mathrm{T}}I_5$ are asymptotically normal, with the rate of $\boldsymbol{\alpha}^{\mathrm{T}}I_3$ dominating the rates of all other terms.

We first show the asymptotical normality of $\boldsymbol{\alpha}^{\mathrm{T}}I_5$. To this end, decompose $\boldsymbol{\alpha}^{\mathrm{T}}I_5$ further into

$$\boldsymbol{\alpha}^{\mathrm{T}}I_5 = \boldsymbol{\alpha}^{\mathrm{T}}\big((\mathbf{H}_{10} - \mathbf{H}_{20})_H^{\mathrm{T}}(\mathbf{H}_{10} - \mathbf{H}_{20})_H\big)^{-1}(\mathbf{H}_{10} - \mathbf{H}_{20})_H^{\mathrm{T}}T^{-1}\sum_{t=1}^T \boldsymbol{\epsilon}_t \otimes (\mathbf{B}_t - \boldsymbol{\mu}_b)\boldsymbol{\gamma}(1 + o_p(1)),$$

so that to utilize Theorem 3(ii) of Wu (2011), we need to show

$$\sum_{t\ge 0}\|P_0(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1\boldsymbol{\epsilon}_t \otimes (\mathbf{B}_t - \boldsymbol{\mu}_b)\boldsymbol{\gamma})\| < \infty, \tag{S.56}$$

where $\mathbf{R}_1 = \big((\mathbf{H}_{10} - \mathbf{H}_{20})_H^{\mathrm{T}}(\mathbf{H}_{10} - \mathbf{H}_{20})_H\big)^{-1}(\mathbf{H}_{10} - \mathbf{H}_{20})_H^{\mathrm{T}}$. With $\|\mathbf{R}_1\|_\infty = O(1)$, we can use the same lines used for proving the asymptotic normality of $\widehat{\boldsymbol{\xi}}_{J_1}$ in the proof of Theorem 3 to conclude that

$$\|P_0(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1\boldsymbol{\epsilon}_t \otimes (\mathbf{B}_t - \boldsymbol{\mu}_b)\boldsymbol{\gamma})\| = O(\max_{1\le j\le N}\|P_0(\epsilon_{t,j})\| + \max_{1\le j\le N}\max_{1\le k\le K}\|B_{t,jk}\|),$$

so that again our assumptions on the predictive dependence measures complete the proof of (S.56).

And with $\lambda_{\min}(\mathbf{R}_1\mathbf{R}_1^{\mathrm{T}})$ and $\lambda_{\max}(\mathbf{R}_1\mathbf{R}_1^{\mathrm{T}})$ both of order N^{-1-a} similar to the derivation of (S.21) in the proof of Theorem 1, we have

$$s_0 = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1\boldsymbol{\Sigma}\mathbf{R}_1^{\mathrm{T}}\boldsymbol{\alpha}$$

having rate N^{-1-a+b} exactly since $N^{-b}\boldsymbol{\Sigma}$ has all eigenvalues uniformly bounded from 0 and infinity by Assumption R6. Also,

$$T^{1/2}s_0^{-1/2}\boldsymbol{\alpha}^{\mathrm{T}}I_5 \xrightarrow{\mathcal{D}} N(0,1)$$

by Theorem 3(ii) of Wu (2011), and hence $\boldsymbol{\alpha}^{\mathrm{T}}I_5$ is $T^{1/2}N^{(1+a-b)/2}$-convergent, so that $\boldsymbol{\alpha}^{\mathrm{T}}I_5$ has rate $T^{-1/2}N^{-(1+a-b)/2}$.

To show the asymptotic normality of $\boldsymbol{\alpha}I_3$, note that we can decompose it further into

$$\begin{aligned}
\boldsymbol{\alpha}^{\mathrm{T}}I_3 &= \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1T^{-1}(\mathbf{B}_{\boldsymbol{\gamma}} - E(\mathbf{B}_{\boldsymbol{\gamma}}))^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}(\widehat{\boldsymbol{\xi}},\boldsymbol{\delta}^*)-\boldsymbol{\beta}^*}\mathrm{vec}(\mathbf{I}_N)(1+o_p(1)) \\
&= \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1\mathrm{vec}\Big(T^{-1}\sum_{t=1}^{T}(\mathbf{B}_t - E(\mathbf{B}_t))\boldsymbol{\gamma}(\boldsymbol{\beta}(\widehat{\boldsymbol{\xi}},\boldsymbol{\delta}^*) - \boldsymbol{\beta}^*)^{\mathrm{T}}\mathbf{X}_t^{\mathrm{T}}\Big)(1+o_p(1)) \\
&= \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1\mathrm{vec}\Big(\big[\boldsymbol{\gamma}^{\mathrm{T}}(\mathbf{b}_{t,i} - E(\mathbf{b}_{t,i}))\mathbf{x}_{t,j}^{\mathrm{T}}(\boldsymbol{\beta}(\widehat{\boldsymbol{\xi}},\boldsymbol{\delta}^*) - \boldsymbol{\beta}^*)\big]_{1\leq i,j\leq N}\Big)(1+o_p(1)) \\
&= \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1\mathrm{vec}\Big(\big[\boldsymbol{\gamma}^{\mathrm{T}}\mathrm{cov}(\mathbf{b}_{t,i},\mathbf{x}_{t,j})(\boldsymbol{\beta}(\widehat{\boldsymbol{\xi}},\boldsymbol{\delta}^*) - \boldsymbol{\beta}^*)\big]_{1\leq i,j\leq N}\Big)(1+o_p(1)) \\
&= \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1\mathbf{S}_{\boldsymbol{\gamma}}(\boldsymbol{\beta}(\widehat{\boldsymbol{\xi}},\boldsymbol{\delta}^*) - \boldsymbol{\beta}^*)(1+o_p(1)).
\end{aligned}$$

But we can easily see from the proof of Theorem 4 that we have

$$T^{1/2}\mathbf{S}_0^{-1/2}(\boldsymbol{\beta}(\widehat{\boldsymbol{\xi}},\boldsymbol{\delta}^*) - \boldsymbol{\beta}^*) \xrightarrow{\mathcal{D}} N(\mathbf{0},\mathbf{I}_K),$$

where $\mathbf{S}_0$ is as defined in Theorem 4 with $\mathbf{M} = \mathbf{I}_K$. We then immediately have

$$T^{1/2}\boldsymbol{\alpha}^{\mathrm{T}}I_3 = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{S}_0^{1/2}\big(T^{1/2}\mathbf{S}_0^{-1/2}(\boldsymbol{\beta}(\widehat{\boldsymbol{\xi}},\boldsymbol{\delta}^*) - \boldsymbol{\beta}^*)\big)$$

being asymptotically normal with asymptotic variance $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{S}_0\mathbf{S}_{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{R}_1^{\mathrm{T}}\boldsymbol{\alpha}$, i.e.,

$$T^{1/2}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{S}_0\mathbf{S}_{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{R}_1^{\mathrm{T}}\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}^{\mathrm{T}}I_3 \xrightarrow{\mathcal{D}} N(0,1).$$

But we have

$$\big\|\boldsymbol{\alpha}\big\|^2\lambda_{\min}(\mathbf{R}_1\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{S}_{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{R}_1^{\mathrm{T}})\lambda_{\min}(\mathbf{S}_0) \leq \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{S}_0\mathbf{S}_{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{R}_1^{\mathrm{T}}\boldsymbol{\alpha} \leq \big\|\boldsymbol{\alpha}\big\|^2\lambda_{\max}(\mathbf{R}_1\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{S}_{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{R}_1^{\mathrm{T}})\lambda_{\max}(\mathbf{S}_0).$$

Observe that

$$\lambda_{\max}(\mathbf{R}_1\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{S}_{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{R}_1^{\mathrm{T}}) \leq \lambda_{\max}(\mathbf{R}_1\mathbf{R}_1^{\mathrm{T}})\lambda_{\max}(\mathbf{S}_{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{S}_{\boldsymbol{\gamma}}) = O(N^{-1-a}\cdot N^{1+a}) = O(1)$$

by Assumption R5 and what we derived for the rate of $\lambda_{\min}(\mathbf{R}_1\mathbf{R}_1^{\mathrm{T}})$ and $\lambda_{\max}(\mathbf{R}_1\mathbf{R}_1^{\mathrm{T}})$ in the proof of the asymptotic normality of $\boldsymbol{\alpha}^{\mathrm{T}}I_5$. At the same time, we also assume in the theorem that $\mathbf{R}_1\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{S}_{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{R}_1^{\mathrm{T}}$ has its smallest eigenvalue of constant order. It means that then $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{S}_0\mathbf{S}_{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{R}_1^{\mathrm{T}}\boldsymbol{\alpha}$ is of order exactly $N^{-(1-b)}$. Hence $\boldsymbol{\alpha}^{\mathrm{T}}I_3$ is of order exactly $T^{-1/2}N^{-(1-b)/2}$.

By Assumption R8, it is not difficult to see that $\big\|I_2\big\|_{\max}$, $\big\|I_4\big\|_{\max}$ and $\boldsymbol{\alpha}^{\mathrm{T}}I_5$ are all of order smaller than $T^{-1/2}N^{-(1-b)/2}$. Hence we have proved that

$$T^{1/2}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{R}_1\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{S}_0\mathbf{S}_{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{R}_1^{\mathrm{T}}\boldsymbol{\alpha})^{-1/2}\boldsymbol{\alpha}^{\mathrm{T}}(\widehat{\boldsymbol{\delta}}_H - \boldsymbol{\delta}_H^*) \xrightarrow{\mathcal{D}} N(0,1).$$

Note that the above can be extended in a straight forward fashion to proving multivariate asymptotic normality presented in Theorem 5, since Theorem 3(ii) of Wu (2011) is in fact applicable to proving multivariate asymptotic normality. This completes the proof of asymptotic normality for $\widehat{\boldsymbol{\delta}}_H$.

Incidentally, since $\|I_1\|_{\max}$ to $\|I_5\|_{\max}$ are all $o_p(1)$, we have $\mathrm{sign}(\widehat{\boldsymbol{\delta}}_H) = \mathrm{sign}(\boldsymbol{\delta}^*_H)$. It remains to show the second part of (S.55) for the sign consistency of $\widehat{\boldsymbol{\delta}}$.

To this end, observe that

$$\begin{aligned}\big\|T^{-1}(\mathbf{H}_{H^c} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H^c})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{Z}(\boldsymbol{\xi}^* - \widehat{\boldsymbol{\xi}})\big\|_{\max} &= O_p(\|\boldsymbol{\xi}^*_{J_2}\|_1 + c_T n),\\ \big\|T^{-1}(\mathbf{H}_{H^c} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H^c})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{X}_{\boldsymbol{\beta}^*-\boldsymbol{\beta}(\widehat{\boldsymbol{\xi}},\boldsymbol{\delta}^*)}\mathrm{vec}(\mathbf{I}_N)\big\|_{\max} &= O_p(T^{-1/2}N^{(1+b)/2}),\\ \big\|T^{-1}(\mathbf{H}_{H^c} - \mathbf{B}^{\mathrm{T}}\mathbf{Z}\mathbf{V}_{0H^c})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\boldsymbol{\epsilon}\big\|_{\max} &= O_p(T^{-1/2}N^{(1+b-a)/2}),\end{aligned}$$

while the right hand side of the second inequality in (S.55) has minimum value of

$$\frac{\lambda'_T}{\|\widetilde{\boldsymbol{\delta}}_{H^c}\|_{\max}} \geq \frac{\lambda'_T}{\|\widetilde{\boldsymbol{\delta}}_{H^c} - \boldsymbol{\delta}^*_{H^c}\|},$$

so that it is sufficient to prove

$$(\|\boldsymbol{\xi}^*_{J_2}\|_1 + c_T n + T^{-1/2}(N^{(1+b)/2} + N^{(1+b-a)/2}))(\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|) = o_p(c_T).$$

With the rate for $\|\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|_1$ in Theorem 2, it is straight forward to verify that the above is indeed satisfied under Assumption R8. This completes the proof of the theorem. □

Proof of Theorem 6. We have

$$\begin{aligned}\|\widehat{\mathbf{W}} - \mathbf{W}^*\| \leq \|\widehat{\mathbf{W}} - \mathbf{W}^*\|_\infty &= O_p(\|\widehat{\mathbf{A}} - \mathbf{A}^*\|_\infty + \sum_{i=1}^M |\widehat{\delta}_i - \delta^*_i|\|\mathbf{W}_{0i}\|_\infty)\\ &= O_p(\|\widehat{\boldsymbol{\xi}}_{J_1} - \boldsymbol{\xi}^*_{J_1}\|_{\max} + \|\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^*\|_1 + \max_{1\leq j\leq N}\|\mathbf{a}^*_{j,j_2}\|_1)\\ &= O_p(c_T + \max_{1\leq j\leq N}\|\mathbf{a}^*_{j,j_2}\|_1) = O_p(c_T).\end{aligned}$$

Also, we can decompose

$$\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}^* = (\mathbf{W}^* - \widehat{\mathbf{W}})(\boldsymbol{\Pi}^*\boldsymbol{\mu}^* + \boldsymbol{\Pi}^*\bar{\mathbf{X}}\boldsymbol{\beta}^* + \boldsymbol{\Pi}^*\bar{\boldsymbol{\epsilon}}) + \bar{\mathbf{X}}(\boldsymbol{\beta}^* - \widetilde{\boldsymbol{\beta}}) + \bar{\boldsymbol{\epsilon}},$$

so that

$$\begin{aligned}\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\|_{\max} &= O_p(\|\mathbf{W}^* - \widehat{\mathbf{W}}\|_\infty(1-\eta)^{-1}(\|\boldsymbol{\mu}^*\|_{\max} + \|\bar{\mathbf{X}}\|_{\max}\|\boldsymbol{\beta}^*\|_1 + \|\bar{\boldsymbol{\epsilon}}\|_{\max})\\ &\quad + \|\bar{\mathbf{X}}\|_{\max}\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 + \|\bar{\boldsymbol{\epsilon}}\|_{\max})\\ &= O_p(c_T + \max_{1\leq j\leq N}\|\mathbf{a}^*_{j,j_2}\|_1) = O_p(c_T).\end{aligned}$$

This completes the proof of the theorem. □

References

Chen, X., M. Xu, and W. B. Wu (2013, 12). Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics 41*(6), 2994–3021.

Li, Y. (2003). A martingale inequality and large deviations. *Statistics & Probability Letters 62*, 317–321.

Liu, W., H. Xiao, and W. Wu (2013). Probability and moment inequalities under dependence. *Statistica Sinica 23*, 1257–1272.

Wu, W. B. (2011). Asymptotic theory for stationary processes. *Statistics and Its Interface 4*, 207–226.