Online Theory Supplement to "Variable Selection and Forecasting in High Dimensional Linear Regressions with Structural Breaks"

Alexander Chudik, M. Hashem Pesaran and Mahrad Sharifvaghefi

Globalization Institute Working Paper 394 Supplement
August 2020 (Revised May 2021)

Research Department

https://doi.org/10.24149/gwp394suppr1

Online Theory Supplement to "Variable Selection and Forecasting in High Dimensional Linear Regressions with Structural Breaks"

Alexander Chudik
Federal Reserve Bank of Dallas

M. Hashem Pesaran
University of Southern California, USA and Trinity College, Cambridge, UK

Mahrad Sharifvaghefi
University of Pittsburgh

April 17, 2021

This online theory supplement has three sections. First section provides the main lemmas needed for the proofs of Theorems 1-3 in Appendix A of the paper. Second section contains the complementary lemmas needed for the proofs of the main lemmas in the previous section. Third section explains the algorithms used for implementing Lasso, Adaptive Lasso and Cross-validation.

Notations: Generic finite positive constants are denoted by C_i for $i = 1, 2, \cdots$ and c. They can take different values in different instances. $\|\mathbf{A}\|_2$, $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|_\infty$ and $\|\mathbf{A}\|_1$ denote the spectral, Frobenius, row, and column norms of matrix $\mathbf{A}$, respectively. $\lambda_i(\mathbf{A})$ denote the i^{th} eigenvalue of a square matrix A. $\|\mathbf{x}\|$ denote the ℓ_2 norm of vector $\mathbf{x}$. If $\{f_n\}_{n=1}^{\infty}$ is any real sequence and $\{g_n\}_{n=1}^{\infty}$ is a sequence of positive real numbers, then $f_n = O(g_n)$, if there exists a positive constant C_0 and n_0 such that $|f_n|/g_n \leq C_0$ for all $n > n_0$. $f_n = o(g_n)$ if $f_n/g_n \to 0$ as $n \to \infty$. If $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ are both positive sequences of real numbers, then $f_n = \ominus(g_n)$ if there exist $n_0 \geq 1$ and positive constants C_0 and C_1, such that $\inf_{n \geq n_0} (f_n/g_n) \geq C_0$ and $\sup_{n \geq n_0} (f_n/g_n) \leq C_1$. respectively. If $\{f_n\}_{n=1}^{\infty}$ is a sequence of random variables and $\{g_n\}_{n=1}^{\infty}$ is a sequence of positive real numbers, then $f_n = O_p(g_n)$, if for any $\varepsilon > 0$, there exists a positive constant B_ε and n_ε such that $\Pr(|f_n| > g_n B_\varepsilon) < \varepsilon$ for all $n > n_\varepsilon$.

Main Lemmas

Lemma S.1 *Let y_t be a target variable generated by equation (1), $\mathbf{z}_t = (z_{1t}, z_{2t}, \cdots, z_{mt})'$ be the $m \times 1$ vector of conditioning covariates in DGP(1) and x_{it} be a covariate in the active set $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$. Under Assumptions 1, 3, and 4 we have*

$$\mathbb{E}\left[y_t x_{it} - \mathbb{E}(y_t x_{it}) | \mathcal{F}_{t-1}\right] = 0,$$

for $i = 1, 2, \cdots, N$,

$$\mathbb{E}\left[y_t z_{\ell t} - \mathbb{E}(y_t z_{\ell t}) | \mathcal{F}_{t-1}\right] = 0,$$

for $\ell = 1, 2, \cdots, m$, *and*

$$\mathbb{E}\left[y_t^2 - \mathbb{E}(y_t^2) | \mathcal{F}_{t-1}\right] = 0.$$

Proof. Note that y_t can be written as

$$y_t = \mathbf{z}_t' \mathbf{a}_t + \mathbf{x}_{k,t}' \boldsymbol{\beta}_t + u_t = \sum_{\ell=1}^m \mathrm{a}_{\ell t} z_{\ell t} + \sum_{j=1}^k \beta_{jt} x_{jt} + u_t,$$

where $\mathbf{x}_{k,t} = (x_{1t}, x_{2t}, \cdots, x_{kt})'$, and $\boldsymbol{\beta}_t = (\beta_{1t}, \beta_{2t}, \cdots, \beta_{kt})'$. Moreover, By Assumption 4, $\mathrm{a}_{\ell t}$ is independent of $x_{it'}$ and $z_{\ell' t'}$ for all i, ℓ', and t'. Hence, for $i = 1, 2, \cdots, N$, we have

$$\mathbb{E}(y_t x_{it} | \mathcal{F}_{t-1}) = \sum_{\ell=1}^m \mathbb{E}(\mathrm{a}_{\ell t} | \mathcal{F}_{t-1}) \mathbb{E}(z_{\ell t} x_{it} | \mathcal{F}_{t-1}) + \sum_{j=1}^k \mathbb{E}(\beta_{jt} | \mathcal{F}_{t-1}) \mathbb{E}(x_{jt} x_{it} | \mathcal{F}_{t-1}) + \mathbb{E}(u_t x_{it} | \mathcal{F}_{t-1}).$$

By Assumption 1, we have $\mathbb{E}(\mathrm{a}_{\ell t} | \mathcal{F}_{t-1}) = \mathbb{E}(\mathrm{a}_{\ell t})$, $\mathbb{E}(z_{\ell t} x_{it} | \mathcal{F}_{t-1}) = \mathbb{E}(z_{\ell t} x_{it})$, $\mathbb{E}(\beta_{jt} | \mathcal{F}_{t-1}) = \mathbb{E}(\beta_{jt})$, $\mathbb{E}(x_{jt} x_{it} | \mathcal{F}_{t-1}) = \mathbb{E}(x_{jt} x_{it})$, and $\mathbb{E}(u_t x_{it} | \mathcal{F}_{t-1}) = \mathbb{E}(u_t x_{it})$. Therefore,

$$\mathbb{E}(y_t x_{it} | \mathcal{F}_{t-1}) = \sum_{\ell=1}^m \mathbb{E}(\mathrm{a}_{\ell t}) \mathbb{E}(z_{\ell t} x_{it}) + \sum_{j=1}^k \mathbb{E}(\beta_{jt}) \mathbb{E}(x_{jt} x_{it}) + \mathbb{E}(u_t x_{it}) = \mathbb{E}(y_t x_{it}).$$

Similarly, we can show that for $\ell = 1, 2, \cdots, m$,

$$\begin{aligned}\mathbb{E}(y_t z_{\ell t} | \mathcal{F}_{t-1}) &= \sum_{\ell'=1}^m \mathbb{E}(\mathrm{a}_{\ell' t} | \mathcal{F}_{t-1}) \mathbb{E}(z_{\ell' t} z_{\ell t} | \mathcal{F}_{t-1}) + \sum_{j=1}^k \mathbb{E}(\beta_{jt} | \mathcal{F}_{t-1}) \mathbb{E}(x_{jt} z_{\ell t} | \mathcal{F}_{t-1}) + \mathbb{E}(u_t z_{\ell t} | \mathcal{F}_{t-1}) \\ &= \sum_{\ell'=1}^m \mathbb{E}(\mathrm{a}_{\ell' t}) \mathbb{E}(z_{\ell' t} z_{\ell t}) + \sum_{j=1}^k \mathbb{E}(\beta_{jt}) \mathbb{E}(x_{jt} z_{\ell t}) + \mathbb{E}(u_t z_{\ell t}) = \mathbb{E}(y_t z_{\ell t}).\end{aligned}$$

Also to establish the last result, we can write y_t as $y_t = \mathbf{q}_t' \boldsymbol{\delta}_t + u_t$, where $\mathbf{q}_t = (\mathbf{z}_t', \mathbf{x}_{k,t}')'$ and $\boldsymbol{\delta}_t = (\mathbf{a}_t', \boldsymbol{\beta}_t')'$. We have,

$$\begin{aligned}\mathbb{E}(y_t^2 | \mathcal{F}_{t-1}) &= \mathbb{E}(\ \boldsymbol{\delta}_t' | \mathcal{F}_{t-1}) \mathbb{E}(\mathbf{q}_t \mathbf{q}_t' | \mathcal{F}_{t-1}) \mathbb{E}(\boldsymbol{\delta}_t | \mathcal{F}_{t-1}) + \mathbb{E}(u_t^2 | \mathcal{F}_{t-1}) + 2\mathbb{E}(\boldsymbol{\delta}_t' | \mathcal{F}_{t-1}) \mathbb{E}(\mathbf{q}_t u_t | \mathcal{F}_{t-1}) \\ &= \mathbb{E}(\boldsymbol{\delta}_t') \mathbb{E}(\mathbf{q}_t \mathbf{q}_t') \mathbb{E}(\boldsymbol{\delta}_t) + \mathbb{E}(u_t^2) + 2\mathbb{E}(\boldsymbol{\delta}_t')\ \mathbb{E}(\mathbf{q}_t u_t) = \mathbb{E}(y_t^2).\end{aligned}$$

■

Lemma S.2 *Let* y_t *be a target variable generated by equation (1). Under Assumptions 2-4, for any value of* $\alpha > 0$, *there exist some positive constants* C_0 *and* C_1 *such that*

$$\sup_t \Pr(|y_t| > \alpha) \leq C_0 \exp(C_1 \alpha^{s/2})$$

Proof. Note that

$$|y_t| \leq \sum_{\ell=1}^m |\mathrm{a}_{\ell t} z_{\ell t}| + \sum_{j=1}^k |\beta_{jt} x_{jt}| + |u_t|.$$

Therefore,

$$\Pr(|y_t| > \alpha) \le \Pr(\textstyle\sum_{\ell=1}^{m}|\mathrm{a}_{\ell t} z_{\ell t}| + \sum_{j=1}^{k}|\beta_{jt} x_{jt}| + |u_t| > \alpha),$$

and by Lemma S.10 for any $0 < \pi_i < 1$, $i = 1, 2, \cdots, k+m+1$, with $\sum_{i=1}^{k+m+1} \pi_j = 1$, we can further write

$$\Pr(|y_t| > \alpha) \le \textstyle\sum_{\ell=1}^{m} \Pr(|\mathrm{a}_{\ell t} z_{\ell t}| > \pi_\ell \alpha) + \sum_{j=1}^{k} \Pr(|\beta_{jt} x_{jt}| > \pi_j \alpha) + \Pr(|u_t| > \pi_{k+m+1}\alpha).$$

Moreover, by Lemma S.11, we have

$$\begin{aligned} \Pr(|\mathrm{a}_{\ell t} z_{\ell t}| > \pi_\ell \alpha) &\le \Pr[|z_{\ell t}| > (\pi_\ell \alpha)^{1/2}] + \Pr[|\mathrm{a}_{\ell t}| > (\pi_\ell \alpha)^{1/2}], \\ \Pr(|\beta_{jt} x_{jt}| > \pi_j \alpha) &\le \Pr[|x_{jt}| > (\pi_j \alpha)^{1/2}] + \Pr[|\beta_{jt}| > (\pi_i \alpha)^{1/2}], \end{aligned}$$

and hence

$$\begin{aligned} \Pr(|y_t| > \alpha) \le{} & \textstyle\sum_{\ell=1}^{m} \Pr[|z_{\ell t}| > (\pi_\ell \alpha)^{1/2}] + \sum_{\ell=1}^{m} \Pr[|\mathrm{a}_{\ell t}| > (\pi_\ell \alpha)^{1/2}] + \\ & \textstyle\sum_{j=1}^{k} \Pr[|x_{jt}| > (\pi_j \alpha)^{1/2}] + \sum_{j=1}^{k} \Pr[|\beta_{jt}| > (\pi_j \alpha)^{1/2}] + \Pr(|u_t| > \pi_{k+1}\alpha), \end{aligned}$$

Therefore, under Assumptions 2-4, we can conclude that for any value of $\alpha > 0$, there exist some positive constants C_0 and C_1 such that

$$\sup_t \Pr(|y_t| > \alpha) \le C_0 \exp(C_1 \alpha^{s/2}).$$

■

Lemma S.3 *Let x_{it} be a covariate in the active set, $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$ and $\mathbf{z}_t = (z_{1t}, z_{2t}, \cdots, z_{mt})'$ be the $m \times 1$ vector of conditioning covariates in the DGP, given by (1). Define the projection regression of x_{it} on $\mathbf{z}_t$ as*

$$x_{it} = \bar{\boldsymbol{\psi}}_i' \mathbf{z}_t + \tilde{x}_{it},$$

where $\bar{\boldsymbol{\psi}}_i = (\psi_1, \cdots, \psi_m)'$ is the $m \times 1$ vector of projection coefficients which is equal to $[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t\mathbf{z}_t')^{-1}][T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t x_{it})]$. Under Assumptions 1, 2, and 4, there exist some finite positive constants C_0, C_1 and C_2 such that if $0 < \lambda \le (s+2)/(s+4)$, then

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j - \mathbb{E}(\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_j)| > \zeta_T) \le \exp(-C_0 T^{-1}\zeta_T^2) + \exp(-C_1 T^{C_2})$$

and if $\lambda > (s+2)/(s+4)$, then

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j - \mathbb{E}(\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_j)| > \zeta_T) \le \exp(-C_0 \zeta_T^{s/(s+1)}) + \exp(-C_1 T^{C_2})$$

for all i and j, where $\tilde{\mathbf{x}}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{iT})'$, $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iT})'$, and $\mathbf{M}_z = \mathbf{I} - T^{-1}\mathbf{Z}\hat{\boldsymbol{\Sigma}}_{zz}^{-1}\mathbf{Z}'$ with $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_T)'$ and $\hat{\boldsymbol{\Sigma}}_{zz} = T^{-1}\sum_{t=1}^{T}(\mathbf{z}_t\mathbf{z}_t')$.

Proof. By Assumption 1 we have

$$\mathbb{E}\left[z_{\ell t}z_{\ell' t}-\mathbb{E}(z_{\ell t}z_{\ell' t})|\mathcal{F}_{t-1}\right]=0.$$

for $\ell,\ell'=1,2,\cdots,m$,

$$\mathbb{E}\left[x_{it}x_{jt}-\mathbb{E}(x_{it}x_{jt})|\mathcal{F}_{t-1}\right]=0,$$

for $i,j=1,2,\cdots,N$, and

$$\mathbb{E}\left[z_{\ell t}x_{it}-\mathbb{E}(z_{\ell t}x_{it})|\mathcal{F}_{t-1}\right]=0,$$

for $\ell=1,2,\cdots,m$, $i=1,2,\cdots,N$. Moreover, by Assumption 2, for all i, ℓ, and t, x_{it}, and $z_{\ell t}$ have exponential decaying probability tails. Additionally, by Assumption 4 the number of pre-selected covariates m is finite. Therefore by Lemma S.27, we can conclude that there exist sufficiently large positive constants C_0, C_1, and C_2 such that if $0<\lambda\leq(s+2)/(s+4)$,

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j-\mathbb{E}(\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_j)|>\zeta_T)\leq\exp(-C_0T^{-1}\zeta_T^2)+\exp(-C_1T^{C_2})$$

and if $\lambda>(s+2)/(s+4)$

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j-\mathbb{E}(\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_j)|>\zeta_T)\leq\exp(-C_0\zeta_T^{s/(s+1)})+\exp(-C_1T^{C_2})$$

for all i and j. ■

Lemma S.4 *Let y_t be a target variable generated by the DGP given by (1), $\mathbf{z}_t=(z_{1t},z_{2t},\cdots,z_{mt})'$ be the $m\times1$ vector of conditioning covariates in DGP(1) and x_{it} be a covariate in the active set, $\mathcal{S}_{Nt}=\{x_{1t},x_{2t},\cdots,x_{Nt}\}$. Define the projection regression of x_{it} on $\mathbf{z}_t$ as*

$$x_{it}=\mathbf{z}_t'\bar{\boldsymbol{\psi}}_{i,T}+\tilde{x}_{it},$$

where $\bar{\boldsymbol{\psi}}_{i,T}=(\psi_{1i,T},\cdots,\psi_{mi,T})'$ is the $m\times1$ vector of projection coefficients which is equal to $\left[T^{-1}\sum_{t=1}^T\mathbb{E}(\mathbf{z}_t\mathbf{z}_t')\right]^{-1}[T^{-1}\sum_{t=1}^T\mathbb{E}(\mathbf{z}_tx_{it})]$. Additionally define the projection regression of y_t on $\mathbf{z}_t$ as

$$y_t=\mathbf{z}_t'\bar{\boldsymbol{\psi}}_{y,T}+\tilde{y}_t,$$

where $\bar{\boldsymbol{\psi}}_{y,T}=(\psi_{1y,T},\cdots,\psi_{my,T})'$ is equal to $\left[T^{-1}\sum_{t=1}^T\mathbb{E}(\mathbf{z}_t\mathbf{z}_t')\right]^{-1}[T^{-1}\sum_{t=1}^T\mathbb{E}(\mathbf{z}_ty_t)]$. Under Assumptions 1- 4, if $0<\lambda\leq(s+2)/(s+4)$,

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{y}-\theta_{i,T}|>\zeta_T)\leq\exp(-C_0T^{-1}\zeta_T^2)+\exp(-C_1T^{C_2}),$$

and if $\lambda > (s+2)/(s+4)$

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > \zeta_T) \leq \exp(-C_0\zeta_T^{s/(s+1)}) + \exp(-C_1T^{C_2}),$$

for all $i = 1, 2, \cdots, N$; *where* $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iT})'$, $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$, $\theta_{i,T} = T\bar{\theta}_{i,T} = \mathbb{E}(\ \tilde{\mathbf{x}}_i'\tilde{\mathbf{y}})$, $\tilde{\mathbf{x}}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{iT})'$, $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_T)'$, $\mathbf{M}_z = \mathbf{I} - T^{-1}\mathbf{Z}\hat{\mathbf{\Sigma}}_{zz}^{-1}\mathbf{Z}'$, $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_T)'$ *and* $\hat{\mathbf{\Sigma}}_{zz} = T^{-1}\sum_{t=1}^{T}\mathbf{z}_t\mathbf{z}_t'$.

Proof. Note that by Assumption 1 and Lemma S.1, for all i and ℓ, cross products of x_{it}, $z_{\ell t}$ and y_t minus their expected values are martingale difference processes with respect to filtration $\mathcal{F}_{t-1}$. Moreover, by Assumption 2 and Lemma S.2 , for all i, ℓ, and t, x_{it}, $z_{\ell t}$ and y_t have exponential decaying probability tails. Additionally, by Assumption 4 the number of pre-selected covariates m is finite. Therefore by Lemma S.27, we can conclude that there exist sufficiently large positive constants C_0, C_1, and C_2 such that if $0 < \lambda \leq (s+2)/(s+4)$, then

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > \zeta_T) \leq \exp(-C_0T^{-1}\zeta_T^2) + \exp(-C_1T^{C_2}),$$

and if $\lambda > (s+2)/(s+4)$, then

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > \zeta_T) \leq \exp(-C_0\zeta_T^{s/(s+1)}) + \exp(-C_1T^{C_2}),$$

for all $i = 1, 2, \cdots, N$. ■

Lemma S.5 *Let* y_t *be a target variable generated by equation (1),* $\mathbf{z}_t$ *be the* $m \times 1$ *vector of conditioning covariates in DGP(1) and* x_{it} *be a covariate in the active set,* $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$. *Define the projection regression of* y_t *on* $\mathbf{q}_t \equiv (\mathbf{z}_t', x_{it})'$ *as*

$$y_t = \bar{\boldsymbol{\phi}}_{i,T}'\mathbf{q}_t + \eta_{it},$$

where $\bar{\boldsymbol{\phi}}_{i,T} \equiv \left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{q}_t\mathbf{q}_t')\right]^{-1}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{q}_t y_t)\right]$ *is the projection coefficients. Under Assumptions 1-4, there exist sufficiently large positive constants* C_0, C_1 *and* C_2 *such that if* $0 < \lambda \leq (s+2)/(s+4)$, *then*

$$\Pr\left[|\boldsymbol{\eta}_i'\mathbf{M}_q\ \boldsymbol{\eta}_i - \mathbb{E}(\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)| > \zeta_T\right] \leq \exp(-C_0T^{-1}\zeta_T^2) + \exp(-C_1T^{C_2}),$$

and if $\lambda > (s+2)/(s+4)$, *then*

$$\Pr\left[|\boldsymbol{\eta}_i'\mathbf{M}_q\ \boldsymbol{\eta}_i - \mathbb{E}(\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)| > \zeta_T\right] \leq \exp(-C_0\zeta_T^{s/(s+1)}) + \exp(-C_1T^{C_2}),$$

for all $i = 1, 2, \cdots, N$; *where* $\boldsymbol{\eta}_i = (\eta_{i1}, \eta_{i2}, \cdots, \eta_{iT})'$, $\mathbf{M}_q = \mathbf{I}_T - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$, *and* $\mathbf{Q} = (\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_T)'$.

Proof. Note that $\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i = \mathbf{y}'\mathbf{M}_q\mathbf{y}$ where $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$. By Lemma S.1 we have

$$\mathbb{E}\left[y_t x_{it} - \mathbb{E}(y_t x_{it}) | \mathcal{F}_{t-1}\right] = 0,$$

for $i = 1, 2, \cdots, N$,

$$\mathbb{E}\left[y_t z_{\ell t} - \mathbb{E}(y_t z_{\ell t}) | \mathcal{F}_{t-1}\right] = 0,$$

for $\ell = 1, 2, \cdots, m$, and

$$\mathbb{E}\left[y_t^2 - \mathbb{E}(y_t^2) | \mathcal{F}_{t-1}\right] = 0.$$

Moreover, by Assumption 2 and Lemma S.2, for all i, ℓ, and t, x_{it}, $z_{\ell t}$ and y_t have exponential decaying probability tails. Additionally, by Assumption 4 the number of pre-selected covariates m is finite. Therefore by Lemma S.27, we can conclude that there exist sufficiently large positive constants C_0, C_1, and C_2 such that if $0 < \lambda \leq (s+2)/(s+4)$, then

$$\Pr\left[|\boldsymbol{\eta}_i'\mathbf{M}_q\ \boldsymbol{\eta}_i - \mathbb{E}(\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)| > \zeta_T\right] \leq \exp(-C_0 T^{-1}\zeta_T^2) + \exp(-C_1 T^{C_2}),$$

and if $\lambda > (s+2)/(s+4)$, then

$$\Pr\left[|\boldsymbol{\eta}_i'\mathbf{M}_q\ \boldsymbol{\eta}_i - \mathbb{E}(\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)| > \zeta_T\right] \leq \exp(-C_0 \zeta_T^{s/(s+1)}) + \exp(-C_1 T^{C_2}),$$

for all $i = 1, 2, \cdots, N$. ■

Lemma S.6 *Let y_t be a target variable generated by equation (1), $\mathbf{z}_t$ be the $m \times 1$ vector of conditioning covariates in DGP(1) and x_{it} be a covariate in the active set $\mathcal{S}_{Nt} = \{x_{1t}, x_{2t}, \cdots, x_{Nt}\}$. Define the projection regression of x_{it} on $\mathbf{z}_t$ as*

$$x_{it} = \mathbf{z}_t'\bar{\boldsymbol{\psi}}_{i,T} + \tilde{x}_{it},$$

where $\bar{\boldsymbol{\psi}}_{i,T} = (\psi_{1i,T}, \cdots, \psi_{mi,T})'$ is the $m \times 1$ vector of projection coefficients which is equal to $[T^{-1}\sum_{t=1}^T \mathbb{E}(\mathbf{z}_t\mathbf{z}_t')^{-1}][T^{-1}\sum_{t=1}^T \mathbb{E}(\mathbf{z}_t x_{it})]$. Additionally define the projection regression of y_t on $\mathbf{z}_t$ as

$$y_t = \mathbf{z}_t'\bar{\boldsymbol{\psi}}_{y,T} + \tilde{y}_t,$$

where $\bar{\boldsymbol{\psi}}_{y,T} = (\psi_{1y,T}, \cdots, \psi_{my,T})'$ is equal to $\left[T^{-1}\sum_{t=1}^T \mathbb{E}(\mathbf{z}_t\mathbf{z}_t')\right]^{-1}[T^{-1}\sum_{t=1}^T \mathbb{E}(\mathbf{z}_t y_t)]$. Lastly, define the projection regression of y_t on $\mathbf{q}_t \equiv (\mathbf{z}_t', x_{it})'$ as

$$y_t = \bar{\boldsymbol{\phi}}_{i,T}'\mathbf{q}_t + \eta_{it},$$

where $\bar{\boldsymbol{\phi}}_{i,T} \equiv \left[T^{-1}\sum_{t=1}^T \mathbb{E}(\mathbf{q}_t\mathbf{q}_t')\right]^{-1}[T^{-1}\sum_{t=1}^T \mathbb{E}(\mathbf{q}_t y_t)]$ is the vector of projection coefficients.

Consider

$$t_{i,T}=\frac{T^{-1/2}\mathbf{x}_i'\mathbf{M}_z\mathbf{y}}{\sqrt{T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i}\sqrt{T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i}},$$

for all $i=1,2,\cdots,N$; *where* $\mathbf{x}_i=(x_{i1},x_{i2},\cdots,x_{iT})'$, $\mathbf{y}=(y_1,y_2,\cdots,y_T)'$, $\boldsymbol{\eta}_i=(\eta_{i1},\eta_{i2},\cdots,\eta_{iT})'$, $\mathbf{M}_z=\mathbf{I}-\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, $\mathbf{Z}=(\mathbf{z}_1,\mathbf{z}_2,\cdots,\mathbf{z}_T)'$, $\mathbf{M}_q=\mathbf{I}-\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$, $\mathbf{Q}=(\mathbf{q}_1,\mathbf{q}_2,\cdots,\mathbf{q}_T)'$. *Under Assumptions 1-4, there exist sufficiently large positive constants* C_0, C_1 *and* C_2 *such that*

$$\Pr\left[|t_{i,T}|>c_p(N,\delta)|\theta_{i,T}=\ominus(T^{1-\epsilon_i})\right]\le\exp\left[-C_0c_p^2(N,\delta)\right]+\exp(-C_1T^{C_2}),\ \textit{for }\epsilon_i>\frac{1}{2}$$

where $c_p(N,\delta)$ *is defined by (8),* $\theta_{i,T}=T\bar{\theta}_{i,T}=\mathbb{E}(\tilde{\mathbf{x}}_i'\tilde{\mathbf{y}})$, $\tilde{\mathbf{x}}_i=(\tilde{x}_{i1},\tilde{x}_{i2},\cdots,\tilde{x}_{iT})'$, *and* $\tilde{\mathbf{y}}=(\tilde{y}_1,\tilde{y}_2,\cdots,\tilde{y}_T)'$. *Moreover, if* $c_p(N,\delta)=o(T^{1/2-\vartheta-c})$ *for any* $0\le\vartheta<1/2$ *and a finite positive constant c, then there exist some finite positive constants* C_0 *and* C_1 *such that,*

$$\Pr\left[|t_{i,T}|>c_p(N,\delta)|\theta_{i,T}=\ominus(T^{1-\vartheta_i})\right]\ge1-\exp(-C_0T^{C_1}),\ \textit{for }0\le\vartheta_i<\tfrac{1}{2}.$$

Proof. Let $\sigma_{\eta_i}^2=\mathbb{E}(T^{-1}\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)$, and $\sigma_{\tilde{x}_i}^2=\mathbb{E}(T^{-1}\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i)$. We have $|t_{i,T}|=\mathcal{A}_{iT}\mathcal{B}_{iT}$, where,

$$\mathcal{A}_{iT}=\frac{|T^{-1/2}\mathbf{x}_i\mathbf{M}_z\mathbf{y}|}{\sigma_{\eta_i}\sigma_{\tilde{x}_i}},$$

and

$$\mathcal{B}_{iT}=\frac{\sigma_{\eta_i}\sigma_{\tilde{x}_i}}{\sqrt{T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i}\sqrt{T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i}}.$$

In the first case where $\theta_{i,T}=\ominus(T^{1-\epsilon_i})$ for some $\epsilon_i>1/2$, by using Lemma S.11 we have

$$\begin{aligned}\Pr\left[|t_{i,T}|>c_p(n,\delta)|\theta_{i,T}=\ominus(T^{1-\epsilon_i})\right]\le{}&\Pr\left[\mathcal{A}_{iT}>c_p(N,\delta)/(1+d_T)|\theta_{i,T}=\ominus(T^{1-\epsilon_i})\right]+\\&\Pr\left[\mathcal{B}_{iT}>1+d_T|\theta_{i,T}=\ominus(T^{1-\epsilon_i})\right],\end{aligned}$$

where $d_T\to0$ as $T\to\infty$. By using Lemma S.13,

$$\begin{aligned}&\Pr\left[\mathcal{B}_{iT}>1+d_T|\theta_{i,T}=\ominus(T^{1-\epsilon_i})\right]\\&=\Pr\left(|\frac{\sigma_{\eta_i}\sigma_{\tilde{x}_i}}{\sqrt{T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i}\sqrt{T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i}}-1|>d_T|\theta_{i,T}=\ominus(T^{1-\epsilon_i})\right)\\&\le\Pr\left(|\frac{(T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i)(T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i)}{\sigma_{\eta_i}^2\sigma_{\tilde{x}_i}^2}-1|>d_T|\theta_{i,T}=\ominus(T^{1-\epsilon_i})\right)\\&=\Pr\left[\mathcal{M}_{iT}+\mathcal{R}_{iT}+\mathcal{M}_{iT}\mathcal{R}_{iT}>d_T|\theta_{i,T}=\ominus(T^{1-\epsilon_i})\right]\end{aligned}$$

where $\mathcal{R}_{iT}=|(T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i)/\sigma_{\eta_i}^2-1|$ and $\mathcal{M}_{iT}=|(T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i)/\sigma_{\tilde{x}_i}^2-1|$. By using Lemmas S.10 and S.11 , for any values of $0<\pi_i<1$ with $\sum_{i=1}^3\pi_i=1$ and a strictly positive constant,

c, we have

$$\begin{aligned}&\Pr\left[\mathcal{B}_{iT} > 1 + d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \\ &\quad \leq \Pr\left[\mathcal{M}_{iT} > \pi_1 d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] + \Pr\left[\mathcal{R}_{iT} > \pi_2 d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] + \\ &\qquad \Pr\left[\mathcal{M}_{iT} > \tfrac{\pi_3}{c} d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] + \Pr\left[\mathcal{R}_{iT} > c|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right].\end{aligned}$$

First, consider $\Pr\left[\mathcal{M}_{iT} > \pi_1 d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right]$, and note that

$$\Pr\left[\mathcal{M}_{iT} > \pi_1 d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] = \Pr\left[|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i - \mathbb{E}(\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i)| > \pi_1\sigma_{\tilde{x}_i}^2 T d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right].$$

Therefore, by Lemma S.3, there exist some constants C_0 and C_1 such that,

$$\Pr\left[\mathcal{M}_{iT} > \pi_1 d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Similarly,

$$\Pr\left[\mathcal{M}_{iT} > \tfrac{\pi_3}{c} d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Also note that

$$\Pr\left[\mathcal{R}_{iT} > \pi_2 d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] = \Pr\left[|\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i - \mathbb{E}(\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)| > \pi_2\sigma_{\eta_i}^2 T d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right].$$

Therefore, by Lemma S.5, there exist some constants C_0 and C_1 such that,

$$\Pr\left[\mathcal{R}_{iT} > \pi_2 d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Similarly,

$$\Pr\left[\mathcal{R}_{iT} > c|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Therefore, we can conclude that there exist some constants C_0 and C_1 such that,

$$\Pr\left[\mathcal{B}_{iT} > 1 + d_T|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp(-C_0 T^{C_1})$$

Now consider $\Pr\left[\mathcal{A}_{iT} > c_p(N,\delta)/(1+d_T)|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right]$, which is equal to

$$\begin{aligned}&\Pr\left(\frac{|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T} + \theta_{i,T}|}{\sigma_{\eta_i}\sigma_{\tilde{x}_i}} > T^{1/2}\frac{c_p(N,\delta)}{1+d_T}|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right) \\ &\quad \leq \Pr\left(|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > \frac{\sigma_{\eta_i}\sigma_{\tilde{x}_i}}{1+d_T}T^{1/2}c_p(N,\delta) - |\theta_{i,T}||\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right).\end{aligned}$$

Note that since $\epsilon_i > 1/2$ the first term on the right hand side of the inequality dominate the second one. Moreover, Since $c_p(N,\delta) = o(T^\lambda)$ for all values of $\lambda > 0$, by Lemma S.4, there exists a finite positive constant C_0 such that

$$\Pr\left[|\mathbf{x}_i'\mathbf{M}_z\mathbf{y}| > k_1 T^{1/2} c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp\left[-C_0 c_p^2(N,\delta)\right],$$

where $k_1 = \frac{\sigma_{\eta_i}\sigma_{\tilde{x}_i}}{1+d_T}$.

Given the probability upper bound for $\mathcal{A}_{iT}$ and $\mathcal{B}_{iT}$, we can conclude that there exist some finite positive constants C_0, C_1 and C_2 such that

$$\Pr\left[|t_{i,T}| > c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\epsilon_i})\right] \leq \exp\left[-C_0 c_p^2(N,\delta)\right] + \exp(-C_1 T^{C_2}).$$

Let's consider the next case where $\theta_{i,T} = \ominus(T^{1-\vartheta_i})$ for some $0 \leq \vartheta_i < 1/2$. We know that

$$\Pr\left[|t_{i,T}| > c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] = 1 - \Pr\left[t_{i,T} < c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right].$$

By Lemma S.15,

$$\Pr\left[|t_{i,T}| < c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \Pr\left[\mathcal{A}_{iT} < \sqrt{1+d_T}c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] + \Pr\left[\mathcal{B}_{iT} < 1/\sqrt{1+d_T}|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right].$$

Since $\theta_{i,T} = \ominus(T^{1-\vartheta_i})$, for some $0 \leq \vartheta_i < 1/2$ and $c_p(N,\delta) = o(T^{1/2-\vartheta-c})$, for any $0 \leq \vartheta < 1/2$, $|\theta_{i,T}| - \sigma_{\eta_i}\sigma_{\tilde{x}_i}[(1+d_T)T]^{1/2}c_p(N,\delta) = \ominus(T^{1-\vartheta_i}) > 0$ and by Lemma S.12, we have

$$\begin{aligned}
&\Pr\left[\mathcal{A}_{iT} < \sqrt{1+d_T}c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \\
&\quad = \Pr\left[\frac{|T^{-1/2}\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - T^{-1/2}\theta_{i,T} + T^{-1/2}\theta_{i,T}|}{\sigma_{\eta_i}\sigma_{\tilde{x}_i}} < \sqrt{1+d_T}c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \\
&\quad \leq \Pr\left[|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > |\theta_{i,T}| - \sigma_{\eta_i}\sigma_{\tilde{x}_i}[(1+d_T)T]^{1/2}c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right].
\end{aligned}$$

Therefore, by Lemma S.4, there exist some finite positive constants C_0 and C_1 such that,

$$\Pr\left[|\mathbf{x}_i'\mathbf{M}_z\mathbf{y} - \theta_{i,T}| > |\theta_{i,T}| - \sigma_{\eta_i}\sigma_{\tilde{x}_i}[(1+d_T)T]^{1/2}c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0 T^{C_1}),$$

and therefore

$$\Pr\left[\mathcal{A}_{iT} < \sqrt{1+d_T}c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Now let consider the probability of $\mathcal{B}_{iT}$,

$$\begin{aligned}
&\Pr\left(\mathcal{B}_{iT} < 1/\sqrt{1+d_T}|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right) \\
&= \Pr\left(\frac{\sigma_{\eta_i}\sigma_{\tilde{x}_i}}{\sqrt{T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i}\sqrt{T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i}} < \frac{1}{\sqrt{1+d_T}}|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right) \\
&\quad = \Pr\left(\frac{(T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i)(T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i)}{\sigma_{\eta_i}^2\sigma_{\tilde{x}_i}^2} > 1+d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right) \\
&\quad \leq \Pr(\mathcal{M}_{iT} + \mathcal{R}_{iT} + \mathcal{M}_{iT}\mathcal{R}_{iT} > d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})),
\end{aligned}$$

where $\mathcal{R}_{iT} = |(T^{-1}\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i)/\sigma_{\eta_i}^2 - 1|$ and $\mathcal{M}_{iT} = |(T^{-1}\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i)/\sigma_{\tilde{x}_i}^2 - 1|$. By using Lemmas S.10 and S.11 , for any values of $0 < \pi_i < 1$ with $\sum_{i=1}^{3}\pi_i = 1$ and a positive constant, c, we have

$$\begin{aligned}
&\Pr\left[\mathcal{B}_{iT} < 1/\sqrt{1+d_T}|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \\
&\quad \leq \Pr\left[\mathcal{M}_{iT} > \pi_1 d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] + \Pr\left[\mathcal{R}_{iT} > \pi_2 d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] + \\
&\qquad \Pr\left[\mathcal{M}_{iT} > \tfrac{\pi_3}{c} d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] + \Pr\left[\mathcal{R}_{iT} > c|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right].
\end{aligned}$$

Let's first consider the $\Pr\left[\mathcal{M}_{iT} > \pi_1 d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right]$. Note that

$$\Pr\left[\mathcal{M}_{iT} > \pi_1 d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] = \Pr\left[|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_i - \mathbb{E}(\boldsymbol{\nu}_i'\boldsymbol{\nu}_i)| > \pi_1\sigma_{\tilde{x}_i}^2 T d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right].$$

So, by Lemma S.3, we know that there exist some constants C_0 and C_1 such that,

$$\Pr\left[\mathcal{M}_{iT} > \pi_1 d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Similarly,

$$\Pr\left[\mathcal{M}_{iT} > \tfrac{\pi_3}{c} d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0 T^{C_1}).$$

Also note that

$$\Pr\left[\mathcal{R}_{iT} > \pi_2 d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] = \Pr\left[|\boldsymbol{\eta}_i'\mathbf{M}_q\boldsymbol{\eta}_i - \mathbb{E}(\boldsymbol{\eta}_i'\boldsymbol{\eta}_i)| > \pi_2\sigma_{\eta_i}^2 T d_T|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right].$$

Therefore, by Lemma S.5, there exist some constants C_0 and C_1 such that,

$$\Pr(\mathcal{R}_{iT} > \pi_2 d_T|\theta_{i,T} \neq 0) \leq \exp(-C_0 T^{C_1}).$$

Similarly,

$$\Pr(\mathcal{R}_{iT} > c|\theta_{i,T} \neq 0) \leq \exp(-C_0 T^{C_1}).$$

Therefore, we can conclude that there exist some constants C_0 and C_1 such that,

$$\Pr\left[\mathcal{B}_{iT} < 1/\sqrt{1+d_T}|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \leq \exp(-C_0T^{C_1}).$$

So, overall we conclude that

$$\Pr\left[|t_{i,T}| > c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right]$$
$$= 1 - \Pr\left[t_{i,T} < c_p(N,\delta)|\theta_{i,T} = \ominus(T^{1-\vartheta_i})\right] \geq 1 - \exp(-C_0T^{C_1}).$$

■

Lemma S.7 *Consider the following data generating process (DGP) for y_t:*

$$y_t = \sum_{i=1}^{k} x_{it}\beta_{it} + u_t \text{ for } t = 1, 2, \cdots, T. \tag{S.1}$$

Estimate the following regression

$$y_t = \sum_{i=1}^{k} x_{it}\phi_i + \sum_{j=1}^{l_T} x_{k+j,t}\delta_j + \eta_t = \mathbf{q}_t'\boldsymbol{\phi} + \mathbf{s}_t'\boldsymbol{\delta}_T + \eta_t, \tag{S.2}$$

by least squares (LS), where $\mathbf{q}_t = (x_{1t}, x_{2t}, \cdots, x_{kt})'$, $\boldsymbol{\phi} = (\phi_1, \phi_2, \cdots, \phi_k)'$, $\mathbf{s}_t = (x_{k+1,t}, x_{k+2,t}, \cdots, x_{k+l_T,t})'$, and $\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots, \delta_{l_T})'$. The LS estimator of $\boldsymbol{\gamma}_T = (\boldsymbol{\phi}', \boldsymbol{\delta}_T')'$ is

$$\hat{\boldsymbol{\gamma}}_T = \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{y}\right), \tag{S.3}$$

where $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_T)'$, $\mathbf{w}_t = (\mathbf{q}_t', \mathbf{s}_t')'$ and $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$. The model error is

$$\hat{\boldsymbol{\eta}} = \mathbf{y} - \mathbf{W}\hat{\boldsymbol{\gamma}}_T. \tag{S.4}$$

Suppose that $\lambda_{\min}\left[T^{-1}\mathbb{E}(\mathbf{W}'\mathbf{W})\right] > c > 0$, and $l_T = \ominus(T^d)$, where $0 \leq d < \frac{1}{2}$. Moreover suppose that Assumptions 1-4 holds. Now,

(i) If $\mathbb{E}(\beta_{it}) = \beta_i$ for all t, then

$$\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\| = O_p\left(T^{\frac{d-1}{2}}\right), \tag{S.5}$$

where $\boldsymbol{\gamma}_T^* = (\boldsymbol{\beta}', \mathbf{0}_{l_T}')'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_k)'$. If Assumption 6 also holds, then

$$T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\sigma_{ijt,x}\sigma_{ijt,\beta}\right) + \bar{\sigma}_{u,T}^2 + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{l_T}{T}\right), \tag{S.6}$$

where $\sigma_{ijt,x} = \mathbb{E}\left(x_{it}x_{jt}\right)$, $\sigma_{ijt,\beta} = \mathbb{E}\left[(\beta_{it} - \beta_i)(\beta_{jt} - \beta_j)\right]$, and $\bar{\sigma}_{u,T}^2 = T^{-1}\mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)$.

(ii) If $\mathbb{E}(\mathbf{w}_t\mathbf{w}_t')$ *is time invariant, then*

$$\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^{\diamond}\| = O_p\left(T^{\frac{d-1}{2}}\right), \tag{S.7}$$

where $\boldsymbol{\gamma}_T^{\diamond} = (\bar{\boldsymbol{\beta}}_T', \mathbf{0}_{l_T}')'$, $\bar{\boldsymbol{\beta}}_T = (\bar{\beta}_{1T}, \bar{\beta}_{2T}, \cdots, \bar{\beta}_{kT})'$, *and* $\bar{\beta}_{iT} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(\beta_{it})$. *If Assumption 6 also holds, then*

$$T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\sigma_{ijt,x}\sigma_{ijt,\beta}^{*}\right) + \bar{\sigma}_{u,T}^2 + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{l_T}{T}\right), \tag{S.8}$$

where $\sigma_{ijt,\beta}^{*} = \mathbb{E}\left[(\beta_{it} - \bar{\beta}_{i,T})(\beta_{jt} - \bar{\beta}_{j,T})\right]$.

Proof. In the first scenario, where $\mathbb{E}(\beta_{it}) = \beta_i$ for all t , we can write (S.1) as

$$y_t = \sum_{i=1}^{k} x_{it}\beta_i + \sum_{i=1}^{k} x_{it}\left(\beta_{it} - \beta_i\right) + u_t = \sum_{i=1}^{k} x_{it}\beta_i + \sum_{i=1}^{k} r_{it} + u_t = \mathbf{q}_t'\boldsymbol{\beta} + \mathbf{r}_t'\boldsymbol{\tau} + u_t,$$

where $r_{it} = x_{it}\left(\beta_{it} - \beta_i\right)$, $\mathbf{r}_t = (r_{1t}, r_{2t}, \cdots, r_{kt})'$, and $\boldsymbol{\tau}$ is a $k \times 1$ vector of ones. We can further write the DGP in a following matrix format,

$$\mathbf{y} = \mathbf{Q}\boldsymbol{\beta} + \mathbf{R}\boldsymbol{\tau} + \mathbf{u}, \tag{S.9}$$

where $\mathbf{Q} = (\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_T)'$, $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_T)'$ and $\mathbf{u} = (u_1, u_2, \cdots, u_T)'$. By substituting (S.9) into (S.3), we obtain

$$\hat{\boldsymbol{\gamma}}_T = \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{Q}\boldsymbol{\beta}\right) + \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) + \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{u}\right), \tag{S.10}$$

where $\mathbf{W} = (\mathbf{Q}, \mathbf{S})$, and $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_T)'$. Since $\boldsymbol{\gamma}_T^{*} = (\boldsymbol{\beta}', \mathbf{0}_{l_T}')'$, $\mathbf{Q}\boldsymbol{\beta} = \mathbf{Q}\boldsymbol{\beta} + \mathbf{S}\mathbf{0}_{l_T} = \mathbf{W}\boldsymbol{\gamma}_T^{*}$, which in turn allows us to write the above result as:

$$\hat{\boldsymbol{\gamma}}_T = \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\ \boldsymbol{\gamma}_T^{*} + \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) + \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{u}\right),$$

and hence

$$\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^{*} = \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) + \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{u}\right). \tag{S.11}$$

We can further write

$$
\begin{aligned}
\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^* = & \left\{\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\}\left\{T^{-1}\left[\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right)\right]\right\} + \\
& \left\{\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\}\left[T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right)\right] + \\
& \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\left\{T^{-1}\left[\left(\mathbf{W}'\mathbf{R}\ \boldsymbol{\tau}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{R}\ \boldsymbol{\tau}\right)\right]\right\} + \\
& \left\{\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\}\left\{T^{-1}\left[\left(\mathbf{W}'\mathbf{u}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\} + \\
& \left\{\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\}\left[T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right] + \\
& \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\left\{T^{-1}\left[\left(\mathbf{W}'\mathbf{u}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\}.
\end{aligned}
$$

Hence, by the sub-additive property of norms and Lemma S.16, we have

$$
\begin{aligned}
\left\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right\| \leq & \left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\left[\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right)\right]\right\| + \\
& \left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right)\right\| + \\
& \left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 \left\|T^{-1}\left[\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right)\right]\right\| + \\
& \left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\left[\left(\mathbf{W}'\mathbf{u}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\|_F + \\
& \left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right\| + \\
& \left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 \left\|T^{-1}\left[\left(\mathbf{W}'\mathbf{u}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{u}\right)\right]\right\|.
\end{aligned}
$$

Since, by Assumption 3, β_{it} for $i = 1, 2, \cdots, k$ are distributed independently of $\mathbf{w}_t$ for $t = 1, 2, \cdots, T$,

$$
\begin{aligned}
T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right) = & \sum_{i=1}^{k}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{w}_t r_{it})\right] = \sum_{i=1}^{k}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{w}_t x_{it}\left(\beta_{it} - \beta_i\right))\right] \\
= & \sum_{i=1}^{k}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{w}_t x_{it})\mathbb{E}\left(\beta_{it} - \beta_i\right)\right] = \mathbf{0}.
\end{aligned}
$$

Also,

$$
T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{u}\right) = T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{w}_t u_t) = \mathbf{0}.
$$

Hence,

$$\begin{aligned}\left\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right\| \leq & \left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right\| + \\ & \left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 \left\|T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right\| + \\ & \left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\mathbf{W}'\mathbf{u}\right\| + \\ & \left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 \left\|T^{-1}\mathbf{W}'\mathbf{u}\right\|.\end{aligned}$$

Since Assumptions 1 and 2 imply that $\mathbf{W}$, and $\mathbf{u}$ satisfy condition (i) and (ii) of Lemma S.19, by Lemmas S.19 and S.20, we have

$$\left\|T^{-1}\mathbf{W}'\mathbf{u}\right\| = O_p\left(\sqrt{\frac{l_T}{T}}\right).$$

Similarly,

$$\left\|T^{-1}\left[\left(\mathbf{W}'\mathbf{W}\right) - \mathbb{E}\left(\mathbf{W}'\mathbf{W}\right)\right]\right\|_F = O_p\left(\frac{l_T}{\sqrt{T}}\right),$$

and since $l_T = \ominus(T^d)$ with $0 \leq d < 1/2$, by Lemma S.21,

$$\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F = O_p\left(\frac{l_T}{\sqrt{T}}\right).$$

Now consider $\|T^{-1}\mathbf{W}'\mathbf{R}\ \boldsymbol{\tau}\|$. Note that the row j and column i of $l_T \times p$ matrix $T^{-1}\mathbf{W}'\mathbf{R}$ is equal to $T^{-1}\sum_{t=1}^T w_{jt}r_{it}$. Hence the j^{th} element of $l_T \times 1$ vector $T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}$ is equal $T^{-1}\sum_{i=1}^k\sum_{t=1}^T w_{jt}r_{it}$. In other words, $T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau} = T^{-1}\sum_{i=1}^k\sum_{t=1}^T \mathbf{w}_t r_{it}$. Therefore, (recalling that $r_{it} = x_{it}\left(\beta_{it} - \beta_i\right)$)

$$\begin{aligned}\left\|T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right\|^2 = & \left\|T^{-1}\sum_{i=1}^k\sum_{t=1}^T(\mathbf{w}_t r_{it})\right\|^2 \leq \sum_{i=1}^k\left\|T^{-1}\sum_{t=1}^T \mathbf{w}_t x_{it}\left(\beta_{it} - \beta_i\right)\right\|^2 \\ = & T^{-2}\sum_{i=1}^k\sum_{t=1}^T\sum_{t'=1}^T \mathbf{w}_t'\mathbf{w}_{t'}x_{it}x_{it'}\left(\beta_{it} - \beta_i\right)\left(\beta_{it'} - \beta_i\right) \\ = & T^{-2}\sum_{i=1}^k\sum_{t=1}^T\sum_{t'=1}^T\sum_{\ell=1}^{k+l_T} w_{\ell t}w_{\ell t'}x_{it}x_{it'}\left(\beta_{it} - \beta_i\right)\left(\beta_{it'} - \beta_i\right).\end{aligned}$$

Since, by Assumption 1, β_{it} for $i = 1, 2, \cdots, k$ are distributed independently of $\mathbf{w}_t$ for $t = 1, 2, \cdots, T$, we can further write,

$$\begin{aligned}
\mathbb{E}\left\|T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right\|^2 &\leq T^{-2}\sum_{i=1}^{k}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{\ell=1}^{k+\ell_T}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\beta_i\right)\left(\beta_{it'}-\beta_i\right)\right] \\
&\leq T^{-2}\sum_{i=1}^{k}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{\ell=1}^{k+\ell_T}\left|\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\right| \times \left|\mathbb{E}\left[\left(\beta_{it}-\beta_i\right)\left(\beta_{it'}-\beta_i\right)\right]\right| \\
&\leq T^{-2}\left(k+\ell_T\right) sup_{i,\ell,t,t'}\left|\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\right|\sum_{i=1}^{k}\sum_{t=1}^{T}\sum_{t'=1}^{T}\left|\mathbb{E}\left[\left(\beta_{it}-\beta_i\right)\left(\beta_{it'}-\beta_i\right)\right]\right|
\end{aligned}$$

Since $\mathbf{W}$ satisfy condition (i) of Lemma S.19, we have $\sup_{i,\ell,t,t'}|\mathbb{E}(w_{\ell t}w_{\ell t'}x_{it}x_{it'})| < C < \infty$. Also, note that for any $t' < t$,

$$\mathbb{E}\left[\left(\beta_{it}-\beta_i\right)\left(\beta_{it'}-\beta_i\right)\right] = \mathbb{E}\left[\left(\beta_{it'}-\beta_i\right)\mathbb{E}\left(\beta_{it}-\beta_i|\mathcal{F}_{t-1}\right)\right],$$

and by Assumption 1, $\mathbb{E}\left(\beta_{it}-\beta_i|\mathcal{F}_{t-1}\right) = 0$. Therefore,

$$\begin{aligned}
\sum_{t=1}^{T}\sum_{t'=1}^{T}\left|\mathbb{E}\left[\left(\beta_{it}-\beta_i\right)\left(\beta_{it'}-\beta_i\right)\right]\right| &= \sum_{t=1}^{T}\left|\mathbb{E}\left[\left(\beta_{it}-\beta_i\right)^2\right]\right| + 2\sum_{t=2}^{T}\sum_{t'=1}^{t}\left|\mathbb{E}\left[\left(\beta_{it}-\beta_i\right)\left(\beta_{it'}-\beta_i\right)\right]\right| \\
&= \sum_{t=1}^{T}\left|\mathbb{E}\left[\left(\beta_{it}-\beta_i\right)^2\right]\right| = O(T).
\end{aligned}$$

Since, by Assumption 3, k is also a finite fixed integer, we conclude that

$$\mathbb{E}\left\|T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right\|^2 = O\left(\frac{l_T}{T}\right),$$

and hence, by Lemma S.20,

$$\left\|T^{-1}\mathbf{W}'\mathbf{R}\boldsymbol{\tau}\right\| = O_p\left(\sqrt{\frac{l_T}{T}}\right).$$

So, we can conclude that

$$\left\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right\| = O_p\left(\sqrt{\frac{l_T}{T}}\right),$$

as required.

In the next step, consider the mean square error of the model, $T^{-1}\hat{\boldsymbol{\eta}}_T'\hat{\boldsymbol{\eta}}_T$. By substituting y from (S.9) into equation (S.4) for the model error, we have

$$\hat{\boldsymbol{\eta}} = \mathbf{y} - \mathbf{W}\hat{\boldsymbol{\gamma}}_T = \mathbf{Q}\boldsymbol{\beta} + \mathbf{R}\boldsymbol{\tau} + \mathbf{u} - \mathbf{W}\hat{\boldsymbol{\gamma}}_T.$$

Since $\mathbf{Q}\boldsymbol{\beta} = \mathbf{W}\boldsymbol{\gamma}_T^*$, where $\boldsymbol{\gamma}_T^* = (\boldsymbol{\beta}', \mathbf{0}_{l_T}')'$, we can further write,

$$\hat{\boldsymbol{\eta}} = \mathbf{R}\boldsymbol{\tau} + \mathbf{u} - \mathbf{W}\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right).$$

Therefore,

$$\begin{aligned}
T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} &= T^{-1}\left[\mathbf{R}\boldsymbol{\tau} + \mathbf{u} - \mathbf{W}\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)\right]'\left[\mathbf{R}\boldsymbol{\tau} + \mathbf{u} - \mathbf{W}\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)\right] \\
&= T^{-1}\left(\mathbf{R}\boldsymbol{\tau} + \mathbf{u}\right)'\left(\mathbf{R}\boldsymbol{\tau} + \mathbf{u}\right) + T^{-1}\left[\mathbf{W}\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)\right]'\left[\mathbf{W}\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)\right] - \\
&\qquad 2T^{-1}\left[\mathbf{W}\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)\right]'\left(\mathbf{R}\boldsymbol{\tau} + \mathbf{u}\right) \\
&= T^{-1}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} + \mathbf{u}'\mathbf{u}\right) + 2T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} + \left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)'\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right) - \\
&\qquad 2\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*\right)'\left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right].
\end{aligned}$$

By substituting for $\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^*$ from (S.11), we get

$$\begin{aligned}
T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} &= T^{-1}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\,\boldsymbol{\tau} + \mathbf{u}'\mathbf{u}\right) + 2T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} + \\
&\qquad \left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right]'\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right] - \\
&\qquad 2\left[T^{-1}\left(\mathbf{W}'\mathbf{R}\,\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right]'\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right] \\
&= T^{-1}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} + \mathbf{u}'\mathbf{u}\right) + 2T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} - \\
&\qquad \left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right]'\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right].
\end{aligned}$$

we can further write

$$\begin{aligned}
T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} &= T^{-1}\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} + \mathbf{u}'\mathbf{u}\right) + T^{-1}\left\{\left[\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\,\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau}\right)\right] + \left[\mathbf{u}'\mathbf{u} - \mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)\right]\right\} + \\
&\quad 2T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} - \left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right]'\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right] - \\
&\quad \left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right]'\left\{\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\}\left[T^{-1}\left(\mathbf{W}'\mathbf{R}\boldsymbol{\tau} + \mathbf{W}'\mathbf{u}\right)\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
&T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} - T^{-1}\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} + \mathbf{u}'\mathbf{u}\right) \leq \\
&\quad T^{-1}\left[\boldsymbol{\tau}'\mathbf{R}'\,\mathbf{R}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau}\right)\right] + T^{-1}\left[\mathbf{u}'\mathbf{u} - \mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)\right] + \\
&\quad 2T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} + \left\|T^{-1}\mathbf{W}'\left(\mathbf{R}\boldsymbol{\tau} + \mathbf{u}\right)\right\|^2\left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 + \\
&\quad \left\|T^{-1}\mathbf{W}'\left(\mathbf{R}\boldsymbol{\tau} + \mathbf{u}\right)\right\|^2\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F.
\end{aligned} \tag{S.12}$$

First, consider $T^{-1}\left[\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau}\right)\right]$. Note that

$$\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} = \boldsymbol{\tau}'\left(\sum_{t=1}^{T}\mathbf{r}_t\mathbf{r}_t'\right)\boldsymbol{\tau} = \sum_{t=1}^{T}\left(\boldsymbol{\tau}'\mathbf{r}_t\right)\left(\mathbf{r}_t'\boldsymbol{\tau}\right) = \sum_{t=1}^{T}\left(\sum_{i=1}^{k}r_{it}\right)\left(\sum_{j=1}^{k}r_{jt}\right) = \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{t=1}^{T}r_{it}r_{jt}.$$

Recalling that $r_{it} = x_{it}(\beta_{it} - \beta_i)$, and hence,

$$T^{-1}\left[\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau}\right)\right] = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T} a_{ij,t}\right),$$

where

$$a_{ij,t} = x_{it}x_{jt}(\beta_{it} - \beta_i)(\beta_{jt} - \beta_j) - \mathbb{E}\left(x_{it}x_{jt}\right)\mathbb{E}\left[(\beta_{it} - \beta_i)(\beta_{jt} - \beta_j)\right].$$

Now consider $\mathbb{E}\left(T^{-1}\sum_{t=1}^{T} a_{ij,t}\right)^2$ and note that

$$\begin{aligned}
\mathbb{E}\left(T^{-1}\sum_{t=1}^{T} a_{ij,t}\right)^2 &= T^{-2}\sum_{t=1}^{T}\mathbb{E}\left(a_{ij,t}^2\right) + 2T^{-2}\sum_{t=2}^{T}\sum_{t'=1}^{t}\mathbb{E}\left(a_{ij,t}a_{ij,t'}\right) \\
&= T^{-2}\sum_{t=1}^{T}\mathbb{E}\left(a_{ij,t}^2\right) + 2T^{-2}\sum_{t=2}^{T}\sum_{t'=1}^{t}\mathbb{E}\left[a_{ij,t'}\mathbb{E}\left(a_{ij,t}|\mathcal{F}_{t-1}\right)\right].
\end{aligned}$$

But, by Assumptions 1, 3, and 6,

$$\begin{aligned}
\mathbb{E}\left(a_{ij,t}|\mathcal{F}_{t-1}\right) &= \mathbb{E}\left(x_{it}x_{jt}|\mathcal{F}_{t-1}\right)\mathbb{E}\left[(\beta_{it} - \beta_i)(\beta_{jt} - \beta_j)|\mathcal{F}_{t-1}\right] - \mathbb{E}\left(x_{it}x_{jt}\right)\mathbb{E}\left[(\beta_{it} - \beta_i)(\beta_{jt} - \beta_j)\right] \\
&= \mathbb{E}\left(x_{it}x_{jt}\right)\mathbb{E}\left[(\beta_{it} - \beta_i)(\beta_{jt} - \beta_j)\right] - \mathbb{E}\left(x_{it}x_{jt}\right)\mathbb{E}\left[(\beta_{it} - \beta_i)(\beta_{jt} - \beta_j)\right] = 0.
\end{aligned}$$

Therefore,

$$\mathbb{E}\left(T^{-1}\sum_{t=1}^{T} a_{ij,t}\right)^2 = T^{-2}\sum_{t=1}^{T}\mathbb{E}\left(a_{ij,t}^2\right) = O\left(\frac{1}{T}\right),$$

and by Lemma S.20 we conclude that

$$\left|T^{-1}\sum_{t=1}^{T} a_{ij,t}\right| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Since by Assumption 3, k is a finite fixed integer, we can further conclude that

$$T^{-1}\left[\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{R}'\mathbf{R}\boldsymbol{\tau}\right)\right] = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T} a_{ij,t}\right) = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{S.13}$$

Now, consider, $T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u}$. Note that

$$T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} = T^{-1}\boldsymbol{\tau}'\left(\sum_{t=1}^{T}\mathbf{r}_t u_t\right) = T^{-1}\sum_{t=1}^{T}\boldsymbol{\tau}'\mathbf{r}_t u_t = T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{k} r_{it}u_t = \sum_{i=1}^{k}\left(T^{-1}\sum_{t=1}^{T} r_{it}u_t\right).$$

We have

$$\mathbb{E}\left(T^{-1}\sum_{t=1}^{T} r_{it}u_t\right)^2 = T^{-2}\sum_{t=1}^{T}\mathbb{E}\left(r_{it}^2 u_t^2\right) + 2T^{-2}\sum_{t=2}^{T}\sum_{t'=1}^{t}\mathbb{E}\left(r_{it}r_{it'}u_t u_{t'}\right).$$

Since $r_{it} = x_{it}(\beta_{it} - \beta_i)$, and β_{it} for $i = 1, 2, \cdots, k$ are distributed independently of x_{js}, $j = 1, 2, \cdots, N$, and u_s for all t and s, we can further write for any $t' < t$

$$\begin{aligned}\mathbb{E}\left(r_{it}r_{it'}u_t u_{t'}\right) &= \mathbb{E}\left(x_{it}u_t x_{it'}u_{t'}\right)\mathbb{E}\left[(\beta_{it} - \beta_i)(\beta_{it'} - \beta_i)\right] \\ &= \mathbb{E}\left(x_{it}u_t x_{it'}u_{t'}\right)\mathbb{E}\left\{(\beta_{it'} - \beta_i)\mathbb{E}\left[(\beta_{it} - \beta_i)|\mathcal{F}_{t-1}\right]\right\}.\end{aligned}$$

But, by Assumption 1, $\mathbb{E}\left[(\beta_{it} - \beta_i)|\mathcal{F}_{t-1}\right] = 0$ and thus $\mathbb{E}\left(r_{it}r_{it'}u_t u_{t'}\right) = 0$ for any $t' < t$. Therefore,

$$\mathbb{E}\left(T^{-1}\sum_{t=1}^{T} r_{it}u_t\right)^2 = T^{-2}\sum_{t=1}^{T}\mathbb{E}\left(r_{it}^2 u_t^2\right) = O\left(\frac{1}{T}\right).$$

Hence, by Lemma S.20, $\left|T^{-1}\sum_{t=1}^{T} r_{it}u_t\right| = O_p\left(\frac{1}{\sqrt{T}}\right)$. Since, by Assumption 3, k is a finite fixed integer, we conclude that

$$T^{-1}\boldsymbol{\tau}'\mathbf{R}'\mathbf{u} = \sum_{i=1}^{k}\left(T^{-1}\sum_{t=1}^{T} r_{it}u_t\right) = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{S.14}$$

By substituting (S.13) and (S.14) into (S.12), and noting that $\left\|T^{-1}\mathbf{W}'\left(\mathbf{R}\boldsymbol{\tau} + \mathbf{u}\right)\right\|^2 = O_p(l_T/T)$, $\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F = O_p(l_T/\sqrt{T})$, and $T^{-1}\left[\mathbf{u}'\mathbf{u} - \mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)\right] = O_p(1/\sqrt{T})$, we conclude that

$$T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\sigma_{ijt,x}\sigma_{ijt,\beta}\right) + \bar{\sigma}_{u,T}^2 + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{l_T}{T}\right),$$

where $\sigma_{ijt,x} = \mathbb{E}\left(x_{it}x_{jt}\right)$, $\sigma_{ijt,\beta} = \mathbb{E}\left[(\beta_{it} - \beta_i)(\beta_{jt} - \beta_j)\right]$, and $\bar{\sigma}_{u,T}^2 = T^{-1}\mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)$.

In the second scenario, where $\mathbb{E}\left(\mathbf{w}_t\mathbf{w}_t'\right)$ is time invariant, we can write (S.1) as

$$y_t = \sum_{i=1}^{k} x_{it}\bar{\beta}_{iT} + \sum_{i=1}^{k} x_{it}\left(\beta_{it} - \bar{\beta}_{iT}\right) + u_t = \sum_{i=1}^{k} x_{it}\bar{\beta}_{iT} + \sum_{i=1}^{k} h_{it} + u_t = \mathbf{q}_t'\bar{\boldsymbol{\beta}} + \mathbf{h}_t'\boldsymbol{\tau} + u_t,$$

where $h_{it} = x_{it}\left(\beta_{it} - \bar{\beta}_{iT}\right)$, and $\mathbf{h}_t = (h_{1t}, h_{2t}, \cdots, h_{kt})'$. We can further write the DGP in a following matrix format,

$$\mathbf{y} = \mathbf{Q}\bar{\boldsymbol{\beta}} + \mathbf{H}\ \boldsymbol{\tau} + \mathbf{u},$$

where $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_T)'$. Now, by using the similar lines of arguments as in the first scenario, we get

$$\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^\diamond = \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right) + \left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(T^{-1}\mathbf{W}'\ \mathbf{u}\right).$$

Notice that

$$\begin{aligned}
T^{-1}\mathbb{E}\left(\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right) &= \sum_{i=1}^{k}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{w}_t h_{it})\right] = \sum_{i=1}^{k}\left\{T^{-1}\sum_{t=1}^{T}\mathbb{E}[\mathbf{w}_t x_{it}(\beta_{it} - \bar{\beta}_{iT})]\right\} \\
&= \sum_{i=1}^{k}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{w}_t x_{it})\mathbb{E}(\beta_{it} - \bar{\beta}_{iT})\right] \\
&= \sum_{i=1}^{k}\left[\mathbb{E}(\mathbf{w}_t x_{it})T^{-1}\sum_{t=1}^{T}\mathbb{E}(\beta_{it} - \bar{\beta}_{iT})\right] = \mathbf{0}.
\end{aligned}$$

Hence, we can further use the similar lines of arguments as in the first scenario and conclude that

$$\begin{aligned}
\|\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_T^\diamond\| \le{}& \left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\| + \\
& \left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 \left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\| + \\
& \left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F \left\|T^{-1}\mathbf{W}'\mathbf{u}\right\| + \\
& \left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2 \left\|T^{-1}\mathbf{W}'\mathbf{u}\right\|.
\end{aligned}$$

We know that

$$\left\|T^{-1}\mathbf{W}'\mathbf{u}\right\| = O_p\left(\sqrt{\frac{l_T}{T}}\right),$$

and

$$\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F = O_p\left(\frac{l_T}{\sqrt{T}}\right).$$

Now consider $\|T^{-1}\mathbf{W}'\mathbf{H}\ \boldsymbol{\tau}\|$. By using the similar lines of arguments as in the first scenario, we have

$$\left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\|^2 \le T^{-2}\sum_{i=1}^{k}\sum_{\ell=1}^{k+l_T}\sum_{t=1}^{T}\sum_{t'=1}^{T} w_{\ell t}w_{\ell t'}x_{it}x_{it'}\left(\beta_{it} - \bar{\beta}_i\right)\left(\beta_{it'} - \bar{\beta}_i\right).$$

Since, by Assumption 3, β_{it} for $i = 1, 2, \cdots, k$ are distributed independently of $\mathbf{w}_t$ for $t = 1, 2, \cdots, T$, we can further write,

$$\mathbb{E}\left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\|^2 \leq T^{-2}\sum_{i=1}^{k}\sum_{\ell=1}^{k+l_T}\sum_{t=1}^{T}\sum_{t'=1}^{T}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right]$$
$$= T^{-2}\sum_{i=1}^{k}\sum_{\ell=1}^{k+l_T}\sum_{t=1}^{T}\mathbb{E}\left(w_{\ell t}^2 x_{it}^2\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)^2\right]+$$
$$T^{-2}\sum_{i=1}^{k}\sum_{\ell=1}^{k+l_T}\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right].$$

Since, by Assumption 1, $\mathbb{E}\left[w_{\ell t}w_{\ell' t} - \mathbb{E}(w_{\ell t}w_{\ell' t})|\mathcal{F}_{t-1}\right] = 0$ for all ℓ, ℓ' and $t = 1, 2, \cdots, T$, we have for any $t' \neq t$

$$\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right) = \mathbb{E}\left(w_{\ell t}x_{it}\right)\mathbb{E}\left(w_{\ell t'}x_{it'}\right).$$

Therefore,

$$\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right]$$
$$= \sum_{t=}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}x_{it}\right)\mathbb{E}\left(w_{\ell t'}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right].$$

Since $\mathbb{E}\left(\mathbf{w}_t\mathbf{w}_t'\right)$ is time invariant, we can further write

$$\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right]$$
$$= \mathbb{E}\left(w_{\ell t}x_{it}\right)^2\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right].$$

Note that, by Assumption 1, for any $t' \neq t$, $\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right] = \left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]\left[\mathbb{E}\left(\beta_{it'}\right)-\bar{\beta}_i\right]$. Therefore

$$\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right]$$
$$= \left[\mathbb{E}\left(w_{\ell t}x_{it}\right)\right]^2\sum_{t=1}^{T}\sum_{t'\neq t}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]\left[\mathbb{E}\left(\beta_{it'}\right)-\bar{\beta}_i\right].$$

We can further write,

$$\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right]$$
$$=\left[\mathbb{E}\left(w_{\ell t}x_{it}\right)\right]^2\left\{\sum_{t=1}^{T}\sum_{t'=1}^{T}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]\left[\mathbb{E}\left(\beta_{it'}\right)-\bar{\beta}_i\right]-\sum_{t=1}^{T}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]^2\right\}$$
$$=\left[\mathbb{E}\left(w_{\ell t}x_{it}\right)\right]^2\left\{\sum_{t=1}^{T}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]\right\}\left\{\sum_{t'=1}^{T}\left[\mathbb{E}\left(\beta_{it'}\right)-\bar{\beta}_i\right]\right\}-$$
$$\left[\mathbb{E}\left(w_{\ell t}x_{it}\right)\right]^2\sum_{t=1}^{T}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]^2.$$

But, $\sum_{t=1}^{T}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]=0$, and therefore,

$$\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(w_{\ell t}w_{\ell t'}x_{it}x_{it'}\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)\left(\beta_{it'}-\bar{\beta}_i\right)\right]=-\left[\mathbb{E}\left(w_{\ell t}x_{it}\right)\right]^2\sum_{t=1}^{T}\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]^2.$$

So,

$$\mathbb{E}\left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\|^2$$
$$\leq T^{-2}\sum_{i=1}^{p}\sum_{\ell=1}^{p+l_T}\sum_{t=1}^{T}\left\{\mathbb{E}\left(w_{\ell t}^2x_{it}^2\right)\mathbb{E}\left[\left(\beta_{it}-\bar{\beta}_i\right)^2\right]-\left[\mathbb{E}\left(w_{\ell t}x_{it}\right)\right]^2\left[\mathbb{E}\left(\beta_{it}\right)-\bar{\beta}_i\right]^2\right\}$$
$$=O\left(\frac{l_T}{T}\right),$$

and hence, by Lemma S.20,

$$\left\|T^{-1}\mathbf{W}'\mathbf{H}\boldsymbol{\tau}\right\|=O_p\left(\sqrt{\frac{l_T}{T}}\right).$$

So, we conclude that

$$\left\|\hat{\boldsymbol{\gamma}}_T-\boldsymbol{\gamma}_T^{\diamond}\right\|=O_p\left(\sqrt{\frac{l_T}{T}}\right).$$

Lastly, consider the model mean square error for the second scenario. Following the same lines of argument as in the first scenario, we can write,

$$\begin{aligned}&T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}}-T^{-1}\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}+\mathbf{u}'\mathbf{u}\right)\leq\\&\quad T^{-1}\left[\boldsymbol{\tau}'\mathbf{H}'\ \mathbf{H}\boldsymbol{\tau}-\mathbb{E}\left(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}\right)\right]+T^{-1}\left[\mathbf{u}'\mathbf{u}-\mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)\right]+\\&\quad 2T^{-1}\boldsymbol{\tau}'\mathbf{H}'\mathbf{u}+\left\|T^{-1}\mathbf{W}'\left(\mathbf{H}\boldsymbol{\tau}+\mathbf{u}\right)\right\|^2\left\|\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_2+\\&\quad\left\|T^{-1}\mathbf{W}'\left(\mathbf{H}\boldsymbol{\tau}+\mathbf{u}\right)\right\|^2\left\|\left(T^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}-\left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F.\end{aligned}\tag{S.15}$$

First, consider $T^{-1}\left[\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}\right)\right]$. Note that

$$\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau} = \boldsymbol{\tau}'\left(\sum_{t=1}^{T}\mathbf{h}_t\mathbf{h}_t'\right)\boldsymbol{\tau} = \sum_{t=1}^{T}\left(\ \boldsymbol{\tau}'\mathbf{r}_t\right)\left(\mathbf{r}_t'\boldsymbol{\tau}\right) = \sum_{t=1}^{T}\left(\sum_{i=1}^{k}h_{it}\right)\left(\sum_{j=1}^{k}h_{jt}\right) = \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{t=1}^{T}h_{it}h_{jt}.$$

Recalling that $h_{it} = x_{it}(\beta_{it} - \bar{\beta}_{iT})$, and hence,

$$T^{-1}\left[\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}\right)\right] = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}b_{ij,t}\right),$$

where

$$b_{ij,t} = x_{it}x_{jt}(\beta_{it} - \bar{\beta}_{iT})(\beta_{jt} - \bar{\beta}_{jT}) - \mathbb{E}\left(x_{it}x_{jt}\right)\mathbb{E}\left[(\beta_{it} - \bar{\beta}_{iT})(\beta_{jt} - \bar{\beta}_{jT})\right].$$

Now consider $\mathbb{E}\left(T^{-1}\sum_{t=1}^{T}b_{ij,t}\right)^2$ and note that

$$\begin{aligned}\mathbb{E}\left(T^{-1}\sum_{t=1}^{T}b_{ij,t}\right)^2 &= T^{-2}\sum_{t=1}^{T}\mathbb{E}\left(b_{ij,t}^2\right) + 2T^{-2}\sum_{t=2}^{T}\sum_{t'=1}^{t}\mathbb{E}\left(b_{ij,t}b_{ij,t'}\right) \\ &= T^{-2}\sum_{t=1}^{T}\mathbb{E}\left(b_{ij,t}^2\right) + 2T^{-2}\sum_{t=2}^{T}\sum_{t'=1}^{t}\mathbb{E}\left[b_{ij,t'}\mathbb{E}\left(b_{ij,t}|\mathcal{F}_{t-1}\right)\right].\end{aligned}$$

But, by Assumptions 1, 3, and 6,

$$\begin{aligned}\mathbb{E}\left(b_{ij,t}|\mathcal{F}_{t-1}\right) &= \mathbb{E}\left(x_{it}x_{jt}|\mathcal{F}_{t-1}\right)\mathbb{E}\left[(\beta_{it} - \bar{\beta}_{iT})(\beta_{jt} - \bar{\beta}_{jT})|\mathcal{F}_{t-1}\right] - \mathbb{E}\left(x_{it}x_{jt}\right)\mathbb{E}\left[(\beta_{it} - \bar{\beta}_{iT})(\beta_{jt} - \bar{\beta}_{jT})\right] \\ &= \mathbb{E}\left(x_{it}x_{jt}\right)\mathbb{E}\left[(\beta_{it} - \bar{\beta}_{iT})(\beta_{jt} - \bar{\beta}_{jT})\right] - \mathbb{E}\left(x_{it}x_{jt}\right)\mathbb{E}\left[(\beta_{it} - \bar{\beta}_{iT})(\beta_{jt} - \bar{\beta}_{jT})\right] = 0.\end{aligned}$$

Therefore,

$$\mathbb{E}\left(T^{-1}\sum_{t=1}^{T}b_{ij,t}\right)^2 = T^{-2}\sum_{t=1}^{T}\mathbb{E}\left(b_{ij,t}^2\right) = O\left(\frac{1}{T}\right),$$

and by Lemma S.20 we conclude that

$$\left|T^{-1}\sum_{t=1}^{T}b_{ij,t}\right| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Since by Assumption 3, k is a finite fixed integer, we can further conclude that

$$T^{-1}\left[\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau} - \mathbb{E}\left(\boldsymbol{\tau}'\mathbf{H}'\mathbf{H}\boldsymbol{\tau}\right)\right] = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}b_{ij,t}\right) = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{S.16}$$

Now, consider, $T^{-1}\boldsymbol{\tau}'\mathbf{H}'\mathbf{u}$. Note that

$$T^{-1}\boldsymbol{\tau}'\mathbf{H}'\mathbf{u} = T^{-1}\boldsymbol{\tau}'\left(\sum_{t=1}^{T}\mathbf{h}_t u_t\right) = T^{-1}\sum_{t=1}^{T}\boldsymbol{\tau}'\mathbf{h}_t u_t = T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{k} h_{it}u_t = \sum_{i=1}^{k}\left(T^{-1}\sum_{t=1}^{T} h_{it}u_t\right).$$

We have

$$\mathbb{E}\left(T^{-1}\sum_{t=1}^{T} h_{it}u_t\right)^2 = T^{-2}\sum_{t=1}^{T}\mathbb{E}\left[(h_{it}u_t)^2\right] + T^{-2}\sum_{t=1}^{T}\sum_{t'\neq t}\mathbb{E}\left(h_{it}h_{it'}u_t u_{t'}\right).$$

Since $h_{it} = x_{it}(\beta_{it} - \bar{\beta}_{iT})$, and β_{it} for $i = 1, 2, \cdots, k$ are distributed independently of x_{js}, $j = 1, 2, \cdots, N$, and u_s for all t and s, we can further write for any $t' \neq t$

$$\mathbb{E}\left(h_{it}h_{it'}u_t u_{t'}\right) = \mathbb{E}\left(x_{it}u_t x_{it'}u_{t'}\right)\mathbb{E}\left[(\beta_{it} - \bar{\beta}_{iT})(\beta_{it'} - \bar{\beta}_{iT})\right]$$

But, by Assumption 1, $\mathbb{E}\left[x_{it}u_t - \mathbb{E}(x_{it}u_t)|\mathcal{F}_{t-1}\right] = 0$ and we also have $\mathbb{E}(x_{it}u_t) = 0$ for $i = 1, 2, \cdots, k$ and thus for any $t' \neq t$ we have

$$\mathbb{E}\left(x_{it}u_t x_{it'}u_{t'}\right) = \mathbb{E}\left(x_{it}u_t\right)\mathbb{E}\left(x_{it'}u_{t'}\right) = 0.$$

Therefore,

$$\mathbb{E}\left(T^{-1}\sum_{t=1}^{T} h_{it}u_t\right)^2 = T^{-2}\sum_{t=1}^{T}\mathbb{E}\left[(h_{it}u_t)^2\right] = O\left(\frac{1}{T}\right).$$

Hence, by Lemma S.20, $\left|T^{-1}\sum_{t=1}^{T} h_{it}u_t\right| = O_p\left(\frac{1}{\sqrt{T}}\right)$. Since, by Assumption 3, k is a finite fixed integer, we conclude that

$$T^{-1}\boldsymbol{\tau}'\mathbf{H}'\mathbf{u} = \sum_{i=1}^{k}\left(T^{-1}\sum_{t=1}^{T} h_{it}u_t\right) = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{S.17}$$

By substituting (S.16) and (S.17) into (S.15), and noting that $\left\|T^{-1}\mathbf{W}'\left(\mathbf{H}\boldsymbol{\tau} + \mathbf{u}\right)\right\|^2 = O_p(l_T/T)$, $\left\|(T^{-1}\mathbf{W}'\mathbf{W})^{-1} - \left[\mathbb{E}\left(T^{-1}\mathbf{W}'\mathbf{W}\right)\right]^{-1}\right\|_F = O_p(l_T/\sqrt{T})$, and $T^{-1}\left[\mathbf{u}'\mathbf{u} - \mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)\right] = O_p(1/\sqrt{T})$, we conclude that

$$T^{-1}\hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(T^{-1}\sum_{t=1}^{T}\sigma_{ijt,x}\sigma^*_{ijt,\beta}\right) + \bar{\sigma}^2_{u,T} + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{l_T}{T}\right),$$

where $\sigma^*_{ijt,\beta} = \mathbb{E}\left[(\beta_{it} - \bar{\beta}_{i,T})(\beta_{jt} - \bar{\beta}_{j,T})\right]$, $\bar{\beta}_{iT} = T^{-1}\sum_{t=1}^{T}\mathbb{E}(\beta_{it})$, and $\bar{\sigma}^2_{u,T} = T^{-1}\mathbb{E}\left(\mathbf{u}'\mathbf{u}\right)$. ■

Complementary Lemmas

Lemma S.8 *Let z_t be a martingale difference process with respect to $\mathcal{F}_{t-1}^z = \sigma(z_{t-1}, z_{t-2}, \cdots)$, and suppose that there exist some finite positive constants C_0 and C_1, and $s > 0$ such that*

$$\sup_t \Pr(|z_t| > \alpha) \leq C_0 \exp(-C_1 \alpha^s), \quad \text{for all } \alpha > 0.$$

Let also $\sigma_{zt}^2 = \mathbb{E}(z_t^2 | \mathcal{F}_{t-1}^z)$ and $\bar{\sigma}_{z,T}^2 = T^{-1} \sum_{t=1}^T \sigma_{zt}^2$. Suppose that $\zeta_T = \ominus(T^\lambda)$, for some $0 < \lambda \leq (s+1)/(s+2)$. Then for any π in the range $0 < \pi < 1$, we have,

$$\Pr\left(|\textstyle\sum_{t=1}^T z_t| > \zeta_T\right) \leq \exp\left[\frac{-(1-\pi)^2 \zeta_T^2}{2T\bar{\sigma}_{z,T}^2}\right].$$

if $\lambda > (s+1)/(s+2)$, then for some finite positive constant C_2,

$$\Pr\left(|\textstyle\sum_{t=1}^T z_t| > \zeta_T\right) \leq \exp\left(-C_2 \zeta_T^{s/(s+1)}\right).$$

Proof. The results follow from Lemma A3 of Chudik et al. (2018) Online Theory Supplement. ■

Lemma S.9 *Let*

$$c_p(n, \delta) = \Phi^{-1}\left(1 - \frac{p}{2f(n, \delta)}\right), \tag{S.18}$$

where $\Phi^{-1}(.)$ is the inverse of standard normal distribution function, p $(0 < p < 1)$ is the nominal size of a test, and $f(n, \delta) = cn^\delta$ for some positive constants δ and c. Moreover, let $a > 0$ and $0 < b < 1$. Then (I) $c_p(n, \delta) = O\left[\sqrt{\delta \ln(n)}\right]$ and (II) $n^a \exp\left[-bc_p^2(n, \delta)\right] = \ominus(n^{a-2b\delta})$.

Proof. The results follow from Lemma 3 of Bailey et al. (2019) Supplementary Appendix A. ■

Lemma S.10 *Let x_i, for $i = 1, 2, \cdots, n$, be random variables. Then for any constants π_i, for $i = 1, 2, \cdots, n$, satisfying $0 < \pi_i < 1$ and $\sum_{i=1}^n \pi_i = 1$, we have*

$$\Pr(\textstyle\sum_{i=1}^n |x_i| > C_0) \leq \sum_{i=1}^n \Pr(|x_i| > \pi_i C_0),$$

where C_0 is a finite positive constant.

Proof. The result follows from Lemma A11 of Chudik et al. (2018) Online Theory Supplement. ■

Lemma S.11 *Let x, y and z be random variables. Then for any finite positive constants C_0, C_1, and C_2, we have*

$$\Pr(|x| \times |y| > C_0) \leq \Pr(|x| > C_0/C_1) + \Pr(|y| > C_1),$$

and

$$\Pr(|x| \times |y| \times |z| > C_0) \leq \Pr(|x| > C_0/(C_1 C_2)) + \Pr(|y| > C_1) + \Pr(|z| > C_2).$$

Proof. The results follow from Lemma A11 of Chudik et al. (2018) Online Theory Supplement. ■

Lemma S.12 *Let x be a random variable. Then for some finite constants B, and C, with $|B| \geq C > 0$, we have*

$$\Pr(|x + B| \leq C) \leq \Pr(|x| > |B| - C).$$

Proof. The results follow from Lemma A12 of Chudik et al. (2018) Online Theory Supplement. ■

Lemma S.13 *Let x_T to be a random variable. Then for a deterministic sequence, $\alpha_T > 0$, with $\alpha_T \to 0$ as $T \to \infty$, there exists $T_0 > 0$ such that for all $T > T_0$ we have*

$$\Pr\left(\left|\frac{1}{\sqrt{x_T}} - 1\right| > \alpha_T\right) \leq \Pr(|x_T - 1| < \alpha_T).$$

Proof. The results follow from Lemma A13 of Chudik et al. (2018) Online Theory Supplement. ■

Lemma S.14 *Consider random variables x_t and z_t with the exponentially bounded probability tail distributions such that*

$$\sup_t \Pr(|x_t| > \alpha) \leq C_0 \exp(-C_1 \alpha^{s_x}), \text{ for all } \alpha > 0,$$
$$\sup_t \Pr(|z_t| > \alpha) \leq C_0 \exp(-C_1 \alpha^{s_z}), \text{ for all } \alpha > 0,$$

where C_0, and C_1 are some finite positive constants, $s_x > 0$, and $s_z > 0$. Then

$$\sup_t \Pr(|x_t z_t| > \alpha) \leq C_0 \exp(-C_1 \alpha^{s/2}), \text{ for all } \alpha > 0,$$

where $s = \min\{s_x, s_z\}$.

Proof. By using Lemma S.11, for all $\alpha > 0$,

$$\Pr(|x_t z_t| > \alpha) \leq \Pr(|x_t| > \alpha^{1/2}) + \Pr(|z_t| > \alpha^{1/2})$$

So,

$$\begin{aligned}\sup_t \Pr(|x_t z_t| > \alpha) &\leq \sup_t \Pr(|x_t| > \alpha^{1/2}) + \sup_t \Pr(|z_t| > \alpha^{1/2}) \\ &\leq C_0 \exp(-C_1 \alpha^{s_x/2}) + C_0 \exp(-C_1 \alpha^{s_z/2}) \\ &\leq C_0 \exp(-C_1 \alpha^{s/2})\end{aligned}$$

where $s = \min\{s_x, s_z\}$. ■

Lemma S.15 *Let x, y and z be random variables. Then for some finite positive constants C_0, and C_1, we have*

$$\Pr(|x| \times |y| < C_0) \leq \Pr(|x| < C_0/C_1) + \Pr(|y| < C_1),$$

Proof. Define events $\mathfrak{A} = \{|x| \times |y| < C_0\}$, $\mathfrak{B} = \{|x| < C_0/C_1\}$ and $\mathfrak{C} = \{|y| < C_1\}$. Then $\mathfrak{A} \in \mathfrak{B} \cup \mathfrak{C}$. Therefore, $\Pr(\mathfrak{A}) \leq \Pr(\mathfrak{B} \cup \mathfrak{C})$. But $\Pr(\mathfrak{B} \cup \mathfrak{C}) \leq \Pr(\mathfrak{B}) + \Pr(\mathfrak{C})$ and hence $\Pr(\mathfrak{A}) \leq \Pr(\mathfrak{B}) + \Pr(\mathfrak{C})$. ■

Lemma S.16 *Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times p$ and $p \times m$ matrices respectively, then*

$$\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_2, \text{ and } \|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_F. \tag{S.19}$$

Proof. $\|\mathbf{AB}\|_F^2 = \text{tr}(\mathbf{ABB}'\mathbf{A}') = \text{tr}[\mathbf{A}(\mathbf{BB}')\mathbf{A}']$, and by result (12) of Lütkepohl (1996, p.44),

$$\text{tr}\left[\mathbf{A}(\mathbf{BB}')\mathbf{A}'\right] \leq \lambda_{\max}(\mathbf{BB}')\text{tr}(\mathbf{AA}') = \|\mathbf{A}\|_F^2 \|\mathbf{B}\|_2^2,$$

where $\lambda_{\max}(\mathbf{BB}')$ is the largest eigenvalue of $\mathbf{BB}'$. Therefore, $\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_2$, as required. Similarly,

$$\|\mathbf{AB}\|_F^2 = \text{tr}(\mathbf{B}'\mathbf{A}'\mathbf{AB}) = \text{tr}[\mathbf{B}'(\mathbf{A}'\mathbf{A})\mathbf{B}] \leq \lambda_{\max}(\mathbf{A}'\mathbf{A})\text{tr}(\mathbf{B}'\mathbf{B}) = \|\mathbf{A}\|_2^2 \|\mathbf{B}\|_F^2,$$

and hence

$$\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_F.$$

■

Lemma S.17 *Let $\mathbf{A} = (a_{ij})_{n \times m}$ where $\sup_{ij} |a_{ij}| < C < \infty$, then*

$$\|\mathbf{A}\|_2 = O\left(\sqrt{nm}\right). \tag{S.20}$$

Proof. This result follows, since $\|\mathbf{A}\|_2 \leq \sqrt{\|\mathbf{A}\|_\infty \|\mathbf{A}\|_1}$, $\|\mathbf{A}\|_\infty = O(m)$ and $\|\mathbf{A}\|_1 = O(n)$. ■

Lemma S.18 *Consider two $N \times N$ nonsingular matrices $\mathbf{A}$ and $\mathbf{B}$ such that*

$$\|\mathbf{B}^{-1}\|_2 \|\mathbf{A} - \mathbf{B}\|_F < 1.$$

Then

$$\|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_F \leq \frac{\|\mathbf{B}^{-1}\|_2^2 \|\mathbf{A} - \mathbf{B}\|_F}{1 - \|\mathbf{B}^{-1}\|_2 \|\mathbf{A} - \mathbf{B}\|_F}.$$

Proof. By Lemma S.16,

$$\|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_F = \|\mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1}\|_F \leq \|\mathbf{A}^{-1}\|_2 \|\mathbf{B} - \mathbf{A}\|_F \|\mathbf{B}^{-1}\|_2$$

Note that

$$\begin{aligned}\|\mathbf{A}^{-1}\|_2 &= \|\mathbf{A}^{-1} - \mathbf{B}^{-1} + \mathbf{B}^{-1}\|_2 \leq \|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_2 + \|\mathbf{B}^{-1}\|_2 \\ &\leq \|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_F + \|\mathbf{B}^{-1}\|_2,\end{aligned}$$

and therefore,

$$\|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_F \leq (\|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_F + \|\mathbf{B}^{-1}\|_2)\|\mathbf{B} - \mathbf{A}\|_F \|\mathbf{B}^{-1}\|_2.$$

Hence,

$$\|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_F (1 - \|\mathbf{B}^{-1}\|_2 \|\mathbf{B} - \mathbf{A}\|_F) \leq \|\mathbf{B}^{-1}\|_2^2 \|\mathbf{B} - \mathbf{A}\|_F.$$

Since $\|\mathbf{B}^{-1}\|_2 \|\mathbf{B} - \mathbf{A}\|_F < 1$, we can further write,

$$\|\mathbf{A}^{-1} - \mathbf{B}^{-1}\|_F \leq \frac{\|\mathbf{B}^{-1}\|_2^2 \|\mathbf{A} - \mathbf{B}\|_F}{1 - \|\mathbf{B}^{-1}\|_2 \|\mathbf{A} - \mathbf{B}\|_F}.$$

■

Lemma S.19 *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be* $T \times N_x$ *and* $T \times N_y$ *matrices of observations on random variables* x_{it} *and* y_{jt}*, for* $i = 1, 2, \cdots, N_x$*,* $j = 1, 2, \cdots, N_y$ *and* $t = 1, 2, \cdots, T$*, respectively. Denote*

$$w_{ij,t} = x_{it} y_{jt} - \mathbb{E}(x_{it} y_{jt}), \text{ for all } i, j \text{ and } t.$$

Suppose that

(i) $\sup_{i,t} \mathbb{E} |x_{it}|^4 < C$, $\sup_{j,t} \mathbb{E} |y_{jt}|^4 < C$, *and*

(ii) $\sup_{i,j} \left[\sum_{t=1}^{T} \sum_{t'=1}^{T} \mathbb{E}(w_{ij,t} w_{ij,t'}) \right] = O(T)$.

Then,

$$\mathbb{E} \left\| T^{-1} \left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y}) \right] \right\|_F^2 = O\left(\frac{N_x N_y}{T} \right). \tag{S.21}$$

Proof. The results follow from Lemma A18 of Chudik et al. (2018) Online Theory Supplement. ■

Lemma S.20 *Let* $\mathbf{X} = (x_{ij})_{T \times N_x}$ *and* $\mathbf{Y} = (y_{ij})_{T \times N_y}$ *be matrices of random variables, respectively. Suppose that,*

$$\mathbb{E} \left\| T^{-1} \left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y}) \right] \right\|_F^2 = O(a_T), \tag{S.22}$$

where $a_T > 0$. Then

$$\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F = O_p(\sqrt{a_T}). \tag{S.23}$$

Proof. For any $B > 0$, by the Markov's inequality

$$\Pr\left(\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F > B\sqrt{a_T}\right) \leq \frac{\mathbb{E}\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F^2}{a_T B^2}$$

Since $\mathbb{E}\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F^2 = O(a_T)$, there exist C and T_0 such that for all $T > T_0$

$$\mathbb{E}\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F^2 \leq C a_T.$$

Hence, for any $\varepsilon > 0$, there exist $B_\varepsilon = \sqrt{\frac{C}{\varepsilon}}$ and $T_\varepsilon = T_0$, such that for all $T > T_\varepsilon$

$$\Pr\left(\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F > B_\varepsilon\sqrt{a_T}\right) \leq \varepsilon.$$

Therefore,

$$\left\|T^{-1}\left[\mathbf{X}'\mathbf{Y} - \mathbb{E}(\mathbf{X}'\mathbf{Y})\right]\right\|_F = O_p\left(\sqrt{a_T}\right).$$

■

Lemma S.21 *Let $\mathbf{\Sigma}_T$ be a positive definite matrix and $\hat{\mathbf{\Sigma}}_T$ be its corresponding estimator. Suppose that $\lambda_{\min}(\mathbf{\Sigma}_T) > c > 0$, and*

$$\mathbb{E}\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F^2 = O(a_T) \tag{S.24}$$

where $a_T > 0$, and $a_T = o(1)$. Then

$$\left\|\hat{\mathbf{\Sigma}}_T^{-1} - \mathbf{\Sigma}_T^{-1}\right\|_F = O_p(\sqrt{a_T}) \tag{S.25}$$

Proof. Let $\mathcal{A}_T = \left\{\left\|\mathbf{\Sigma}_T^{-1}\right\|_2\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F < 1\right\}$, $\mathcal{B}_T = \left\{\left\|\hat{\mathbf{\Sigma}}_T^{-1} - \mathbf{\Sigma}_T^{-1}\right\|_F > B\sqrt{a_T}\right\}$ and $\mathcal{D}_T = \left\{\frac{\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F}{\left(1 - \left\|\mathbf{\Sigma}_T^{-1}\right\|_2\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F\right)} > B\sqrt{a_T}\right\}$ where $B > 0$ is an arbitrary constant. If $\mathcal{A}_T$ holds, by Lemma S.18,

$$\left\|\hat{\mathbf{\Sigma}}_T^{-1} - \mathbf{\Sigma}_T^{-1}\right\|_F \leq \frac{\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F}{1 - \left\|\mathbf{\Sigma}_T^{-1}\right\|_2\left\|\hat{\mathbf{\Sigma}}_T - \mathbf{\Sigma}_T\right\|_F}.$$

Hence $\mathcal{B}_T \cap \mathcal{A}_T \subseteq \mathcal{D}_T$. Therefore

$$\Pr(\mathcal{B}_T \cap \mathcal{A}_T) \leq \Pr\left(\frac{\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2 \left\|\hat{\boldsymbol{\Sigma}}_T - \boldsymbol{\Sigma}_T\right\|_F}{\left(1 - \left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2 \left\|\hat{\boldsymbol{\Sigma}}_T - \boldsymbol{\Sigma}_T\right\|_F\right)} > B\sqrt{a_T} \right)$$
$$= \Pr\left(\left\|\hat{\boldsymbol{\Sigma}}_T - \boldsymbol{\Sigma}_T\right\|_F > \frac{B\sqrt{a_T}}{\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2 \left(\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2 + B\sqrt{a_T}\right)} \right)$$

By the Markov's inequality, we can further conclude that

$$\Pr(\mathcal{B}_T \cap \mathcal{A}_T) \leq \frac{\mathbb{E}\left\|\hat{\boldsymbol{\Sigma}}_T - \boldsymbol{\Sigma}_T\right\|_F^2}{a_T} \times \frac{\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2 \left(\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2 + B\sqrt{a_T}\right)^2}{B^2}.$$

Since by assumption $\mathbb{E}\left\|\hat{\boldsymbol{\Sigma}}_T - \boldsymbol{\Sigma}_T\right\|_F^2 = O(a_T)$, there exist C and $T_0 > 0$ such that for all $T > T_0$,

$$\mathbb{E}\left\|\hat{\boldsymbol{\Sigma}}_T - \boldsymbol{\Sigma}_T\right\|_F^2 \leq C a_T.$$

Therefore, for all $T > T_0$,

$$\Pr(\mathcal{B}_T \cap \mathcal{A}_T) \leq \frac{C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2 \left(\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2 + B\sqrt{a_T}\right)^2}{B^2}.$$

Moreover,

$$\Pr(\mathcal{A}_T^c) = \Pr\left(\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2 \left\|\hat{\boldsymbol{\Sigma}}_T - \boldsymbol{\Sigma}_T\right\|_F \geq 1\right) = \Pr\left(\left\|\hat{\boldsymbol{\Sigma}}_T - \boldsymbol{\Sigma}_T\right\|_F \geq \frac{1}{\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2} \right).$$

By the Markov's inequality, we can further write

$$\Pr(\mathcal{A}_T^c) \leq \left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2 \times \mathbb{E}\left\|\hat{\boldsymbol{\Sigma}}_T - \boldsymbol{\Sigma}_T\right\|_F^2,$$

and hence, for all $T > T_0$,

$$\Pr(\mathcal{A}_T^c) \leq C\left\|\boldsymbol{\Sigma}_T^{-1}\right\|_2^2 a_T.$$

Note that

$$\Pr(\mathcal{B}_T) = \Pr\left(\mathcal{B}_T \cap \mathcal{A}_T\right) + \Pr(\mathcal{B}_T|\mathcal{A}_T^c)\Pr(\mathcal{A}_T^c),$$

and since $\Pr(\mathcal{B}_T \cap \mathcal{A}_T) \leq \Pr(\mathcal{D}_T)$ and $\Pr(\mathcal{B}_T|\mathcal{A}_T^c) \leq 1$, we have

$$\Pr(\mathcal{B}_T) \leq \Pr(\mathcal{B}_T \cap \mathcal{A}_T) + \Pr(\mathcal{A}_T^c).$$

Therefore, for all $T > T_0$,

$$\Pr\left(\left\|\hat{\mathbf{\Sigma}}_T^{-1} - \mathbf{\Sigma}_T^{-1}\right\|_F > B\sqrt{a_T}\right) \leq \frac{C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2\left(\left\|\mathbf{\Sigma}_T^{-1}\right\|_2 + B\sqrt{a_T}\right)^2}{B^2} + C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2 a_T.$$

Now, for a given $\varepsilon > 0$, we are interested to find $B_\varepsilon > 0$ and $T_\varepsilon > 0$ such that for all $T > T_\varepsilon$,

$$\Pr\left(\left\|\hat{\mathbf{\Sigma}}_T^{-1} - \mathbf{\Sigma}_T^{-1}\right\|_F > B_\varepsilon\sqrt{a_T}\right) \leq \varepsilon.$$

To do so, we first find a value of B such that

$$\frac{C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2\left(\left\|\mathbf{\Sigma}_T^{-1}\right\|_2 + B\sqrt{a_T}\right)^2}{B^2} + C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2 a_T = \varepsilon.$$

By multiplying both sides of the above equality by B^2 and bringing all the equations to the left hand side we have

$$\left(\varepsilon - 2C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2 a_T\right)B^2 - 2C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^3\sqrt{a_T}B - C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^4 = 0.$$

By solving the above quadratic equation of B we have

$$\begin{aligned} B^* &= \frac{2C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^3\sqrt{a_T} \pm \sqrt{4C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^4\varepsilon - 4C^2\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^6 a_T}}{2\left(\varepsilon - 2C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2 a_T\right)} \\ &= \frac{\left\|\mathbf{\Sigma}_T^{-1}\right\|_2\left(\sqrt{a_T} \pm \sqrt{\frac{\varepsilon}{C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2} - a_T}\right)}{\frac{\varepsilon}{C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2} - 2a_T} \end{aligned}$$

Notice that $a_T \to 0$ as $T \to \infty$, therefore for large enough T^* we have both $\frac{\varepsilon}{C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2} - 2a_T$ and $\frac{\varepsilon}{C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2} - a_T$ being greater than zero for all $T > T^*$. Now, by setting $T_\varepsilon = \max\{T^*, T_0\}$ and

$$B_\varepsilon = \frac{\left\|\mathbf{\Sigma}_T^{-1}\right\|_2\left(\sqrt{a_T} + \sqrt{\frac{\varepsilon}{C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2} - a_T}\right)}{\frac{\varepsilon}{C\left\|\mathbf{\Sigma}_T^{-1}\right\|_2^2} - 2a_T} > 0,$$

we achieve our goal that for all $T > T_\varepsilon$,

$$\Pr\left(\left\|\hat{\mathbf{\Sigma}}_T^{-1} - \mathbf{\Sigma}_T^{-1}\right\|_F > B_\varepsilon\sqrt{a_T}\right) \leq \varepsilon.$$

■

Remark 7 *By using Lemma S.18 we achieve the probability convergence order for* $\left\|\hat{\mathbf{\Sigma}}_T^{-1} - \mathbf{\Sigma}_T^{-1}\right\|_F$ *that is sharper than the one shown in the proof Lemma A21 of Chudik et al. (2018) (see equations (B.103) and (B.105) of Chudik et al. (2018) Online Theory Supplement).*

Lemma S.22 *Let z_{ij} be a random variable for $i = 1, 2, \cdots, N$, and $j = 1, 2, \cdots, N$. Then, for any $d_T > 0$,*

$$\Pr(N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}|z_{ij}| > d_T) \leq N^2 \sup_{i,j}\Pr(|z_{ij}| > d_T)$$

Proof. We know that $N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}|z_{ij}| \leq \sup_{i,j}|z_{ij}|$. Therefore,

$$\begin{aligned}
&\Pr(N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}|z_{ij}| > d_T) \leq \Pr(\sup_{i,j}|z_{ij}| > d_T) \\
&\quad \leq \Pr[\cup_{i=1}^{N}\cup_{j=1}^{N}(|z_{ij}| > d_T)] \leq \sum_{i=1}^{N}\sum_{j=1}^{N}\Pr(|z_{ij}| > d_T) \\
&\quad \leq N^2 \sup_{i,j}\Pr(|z_{ij}| > d_T).
\end{aligned}$$

■

Lemma S.23 *Let $\hat{\mathbf{\Sigma}}$ be an estimator of a $N \times N$ symmetric invertible matrix $\mathbf{\Sigma}$. Suppose that there exits a finite positive constant C_0, such that*

$$\sup_{i,j}\Pr(|\hat{\sigma}_{ij} - \sigma_{ij}| > d_T) \leq \exp(-C_0 T d_T^2), \text{ for any } d_T > 0,$$

where σ_{ij} and $\hat{\sigma}_{ij}$ are the elements of $\mathbf{\Sigma}$ and $\hat{\mathbf{\Sigma}}$ respectively. Then, for any $b_T > 0$,

$$\begin{aligned}
\Pr(\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_F > b_T) \leq N^2 \exp\left[-C_0 \frac{T b_T^2}{N^2\|\mathbf{\Sigma}^{-1}\|_2^2(\|\mathbf{\Sigma}^{-1}\|_2 + b_T)^2}\right] + \\
N^2 \exp\left(-C_0 \frac{T}{N^2\|\mathbf{\Sigma}^{-1}\|_2^2}\right).
\end{aligned}$$

Proof. Let $\mathcal{A}_N = \{\|\mathbf{\Sigma}^{-1}\|_2 \| \hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_F \leq 1\}$ and $\mathcal{B}_N = \{\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_F > b_T\}$, and note that by Lemma S.18 if $\mathcal{A}_N$ holds we have

$$\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_F \leq \frac{\|\mathbf{\Sigma}^{-1}\|_2^2\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_F}{1 - \|\mathbf{\Sigma}^{-1}\|_2\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_F}.$$

Hence

$$\begin{aligned}
\Pr(\mathcal{B}_N|\mathcal{A}_N) &\leq \Pr\left(\frac{\|\mathbf{\Sigma}^{-1}\|_2^2\|\hat{\mathbf{\Sigma}} - \ \mathbf{\Sigma}\|_F}{1 - \|\mathbf{\Sigma}^{-1}\|_2\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_F} > b_T\right) \\
&= \Pr\left[\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_F > \frac{b_T}{\|\mathbf{\Sigma}^{-1}\|_2(\|\mathbf{\Sigma}^{-1}\|_2 + b_T)}\right]
\end{aligned}$$

Note that $\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_F = \left(\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\sigma}_{ij} - \sigma_{ij})^2\right)^{1/2}$. Therefore,

$$\begin{aligned}
\Pr(\mathcal{B}_N|\mathcal{A}_N) &\leq \Pr\left[\left(\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\sigma}_{ij} - \sigma_{ij})^2\right)^{1/2} > \frac{b_T}{\|\mathbf{\Sigma}^{-1}\|_2(\|\mathbf{\Sigma}^{-1}\|_2 + b_T)}\right] \\
&= \Pr\left[\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\sigma}_{ij} - \sigma_{ij})^2 > \frac{b_T^2}{\|\mathbf{\Sigma}^{-1}\|_2^2(\|\mathbf{\Sigma}^{-1}\|_2 + b_T)^2}\right]
\end{aligned}$$

By Lemma S.22, we can further write,

$$\begin{aligned}
\Pr(\mathcal{B}_N|\mathcal{A}_N) &\leq N^2 \sup_{i,j} \Pr\left[(\hat{\sigma}_{ij}-\sigma_{ij})^2 > \frac{b_T^2}{N^2\|\mathbf{\Sigma}^{-1}\|_2^2(\|\mathbf{\Sigma}^{-1}\|_2+b_T)^2}\right] \\
&= N^2 \sup_{i,j} \Pr\left[|\hat{\sigma}_{ij}-\sigma_{ij}| > \frac{b_T}{N\|\mathbf{\Sigma}^{-1}\|_2(\|\mathbf{\Sigma}^{-1}\|_2+b_T)}\right] \\
&\leq N^2 \exp\left[-C_0 \frac{Tb_T^2}{N^2\|\mathbf{\Sigma}^{-1}\|_2^2(\|\mathbf{\Sigma}^{-1}\|_2+b_T)^2}\right]
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\Pr(\mathcal{A}_N^c) &= \Pr(\|\mathbf{\Sigma}^{-1}\|_2\|\hat{\mathbf{\Sigma}}-\mathbf{\Sigma}\|_F > 1) \\
&= \Pr(\|\hat{\mathbf{\Sigma}}-\mathbf{\Sigma}\|_F > \|\mathbf{\Sigma}^{-1}\|_2^{-1}) \\
&= \Pr\left[\left(\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\sigma}_{ij}-\sigma_{ij})^2\right)^{1/2} > \|\mathbf{\Sigma}^{-1}\|_2^{-1}\right] \\
&= \Pr\left[\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\sigma}_{ij}-\sigma_{ij})^2 > \|\mathbf{\Sigma}^{-1}\|_2^{-2}\right] \\
&\leq N^2 \sup_{i,j} \Pr\left[(\hat{\sigma}_{ij}-\sigma_{ij})^2 > \frac{1}{N^2\|\mathbf{\Sigma}^{-1}\|_2^2}\right] \\
&\leq N^2 \sup_{i,j} \Pr\left[|\hat{\sigma}_{ij}-\sigma_{ij}| > \frac{1}{N\|\mathbf{\Sigma}^{-1}\|_2}\right] \\
&\leq N^2 \exp\left[-C_0 \frac{T}{N^2\| \ \mathbf{\Sigma}^{-1}\|_2^2}\right]
\end{aligned}$$

Note that

$$\Pr(\mathcal{B}_N) = \Pr(\mathcal{B}_N|\mathcal{A}_N)\Pr(\mathcal{A}_N) + \Pr(\mathcal{B}_N|\mathcal{A}_N^c)\Pr(\mathcal{A}_N^c),$$

and since $\Pr(\mathcal{A}_N)$ and $\Pr(\mathcal{B}_N|\mathcal{A}_N^c)$ are less than equal to one, we have

$$\Pr(\mathcal{B}_N) \leq \Pr(\mathcal{B}_N|\mathcal{A}_N) + \Pr(\mathcal{A}_N^c).$$

Therefore,

$$\Pr(\mathcal{B}_{NT}) \leq N^2 \exp\left[-C_0 \frac{Tb_T^2}{N^2\|\mathbf{\Sigma}^{-1}\|_2^2(\|\mathbf{\Sigma}^{-1}\|_2+b_T)^2}\right] + N^2 \exp\left[-C_0 \frac{T}{N^2\|\mathbf{\Sigma}^{-1}\|_2^2}\right].$$

■

Lemma S.24 *Let $\hat{\mathbf{\Sigma}}$ be an estimator of a $N \times N$ symmetric invertible matrix $\mathbf{\Sigma}$. Suppose that there exits a finite positive constant C_0, such that*

$$\sup_{i,j} \Pr(|\hat{\sigma}_{ij}-\sigma_{ij}| > d_T) \leq \exp\left[-C_0(Td_T)^{s/s+2}\right], \ \textit{for any } d_T > 0,$$

where σ_{ij} and $\hat{\sigma}_{ij}$ are the elements of $\mathbf{\Sigma}$ and $\hat{\mathbf{\Sigma}}$ respectively. Then, for any $b_T > 0$,

$$\Pr(\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_F > b_T) \leq N^2 \exp\left[-C_0 \frac{(Tb_T)^{s/s+2}}{N^{s/s+2}\|\mathbf{\Sigma}^{-1}\|_2^{s/s+2}(\| \mathbf{\Sigma}^{-1}\|_2 + b_T)^{s/s+2}}\right] + N^2 \exp\left(-C_0 \frac{T^{s/s+2}}{N^{s/s+2}\|\mathbf{\Sigma}^{-1}\|_2^{s/s+2}}\right).$$

Proof. The proof is similar to the proof of Lemma S.23. ■

Lemma S.25 *Let $\{x_{it}\}_{t=1}^T$ for $i = 1, 2, \cdots, N$ and $\{z_{jt}\}_{t=1}^T$ for $j = 1, 2, \cdots, m$ be time-series processes. Also let $\mathcal{F}_{it}^x = \sigma(x_{it}, x_{i,t-1}, \cdots)$ for $i = 1, 2, \cdots, N$, $\mathcal{F}_{jt}^z = \sigma(z_{jt}, z_{j,t-1}, \cdots)$ for $j = 1, 2, \cdots, m$, $\mathcal{F}_t^x = \cup_{i=1}^N \mathcal{F}_{it}^x$, $\mathcal{F}_t^z = \cup_{j=1}^m \mathcal{F}_{jt}^z$, and $\mathcal{F}_t = \mathcal{F}_t^x \cup \mathcal{F}_t^z$. Define the projection regression of x_{it} on $\mathbf{z}_t = (z_{1t}, z_{2t}, \cdots, z_{m,t})'$ as*

$$x_{it} = \mathbf{z}_t' \boldsymbol{\psi}_{i,T} + \nu_{it}$$

where $\boldsymbol{\psi}_{i,T} = (\psi_{1i,T}, \psi_{2i,T}, \cdots, \psi_{mi,T})'$ is the $m \times 1$ vector of projection coefficients which is equal to $\left[T^{-1} \sum_{t=1}^T \mathbb{E}(\mathbf{z}_t \mathbf{z}_t')\right]^{-1} [T^{-1} \sum_{t=1}^T \mathbb{E}(\mathbf{z}_t x_{it})]$. Suppose, $\mathbb{E}[x_{it}x_{i't} - \mathbb{E}(x_{it}x_{i't})|\mathcal{F}_{t-1}] = 0$ for all $i, i' = 1, 2, \cdots, N$, $\mathbb{E}[z_{jt}z_{j't} - \mathbb{E}(z_{jt}z_{j't})|\mathcal{F}_{t-1}] = 0$ for all $j, j' = 1, 2, \cdots, m$, and $\mathbb{E}[z_{jt}x_{it} - \mathbb{E}(z_{jt}x_{it})|\mathcal{F}_{t-1}] = 0$ for all $j = 1, 2, \cdots, m$ and for all $i = 1, 2, \cdots, N$. Then

$$\mathbb{E}\left[\nu_{it}\nu_{i't} - \mathbb{E}(\nu_{it}\nu_{i't})|\mathcal{F}_{t-1}\right] = 0,$$

for all $j, j' = 1, 2, \cdots, N$,

$$\mathbb{E}\left[\nu_{it}z_{jt} - \mathbb{E}(\nu_{it}z_{jt})|\mathcal{F}_{t-1}\right] = 0,$$

for all $i = 1, 2, \cdots, N$ and $j = 1, 2, \cdots, m$, and

$$T^{-1} \sum_{t=1}^T \mathbb{E}(\nu_{it}z_{jt}) = 0,$$

for all $i = 1, 2, \cdots, N$ and $j = 1, 2, \cdots, m$.

Proof.

$$\begin{aligned}
\mathbb{E}(\nu_{it}\nu_{i't}|\mathcal{F}_{t-1}) &= \mathbb{E}(x_{it}x_{i't}|\mathcal{F}_{t-1}) - \mathbb{E}(x_{it}\mathbf{z}_t'|\mathcal{F}_{t-1})\boldsymbol{\psi}_{i',T} - \\
&\quad \mathbb{E}(x_{i't}\mathbf{z}_t'|\mathcal{F}_{t-1})\boldsymbol{\psi}_{i,T} + \boldsymbol{\psi}_{i,T}'\mathbb{E}(\mathbf{z}_t\mathbf{z}_t'|\mathcal{F}_{t-1})\boldsymbol{\psi}_{i',T} \\
&= \mathbb{E}(x_{it}x_{i't}) - \mathbb{E}(x_{it}\mathbf{z}_t')\boldsymbol{\psi}_{i',T} - \mathbb{E}(x_{i't}\mathbf{z}_t')\boldsymbol{\psi}_{i,T} + \\
&\quad \boldsymbol{\psi}_{i,T}'\mathbb{E}(\mathbf{z}_t\mathbf{z}_t')\boldsymbol{\psi}_{i',T} = \mathbb{E}(\nu_{it}\nu_{i't}).
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(\nu_{it}z_{jt}|\mathcal{F}_{t-1}) &= \mathbb{E}(x_{it}z_{jt}|\mathcal{F}_{t-1}) - \mathbb{E}(\mathbf{z}_t'z_{jt}|\mathcal{F}_{t-1})\boldsymbol{\psi}_{i,T} \\
&= \mathbb{E}(x_{it}z_{jt}) - \mathbb{E}(\mathbf{z}_t'z_{jt})\boldsymbol{\psi}_{i,T} = \mathbb{E}(\nu_{it}z_{it}).
\end{aligned}$$

$$T^{-1}\sum_{t=1}^{T}\mathbb{E}(\nu_{it}\mathbf{z}_t) = T^{-1}\sum_{t=1}^{T}\mathbb{E}(x_{it}\mathbf{z}_t) - \boldsymbol{\psi}_{i,T}'^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t\mathbf{z}_t')]$$
$$= T^{-1}\sum_{t=1}^{T}\mathbb{E}(x_{it}\mathbf{z}_t) - T^{-1}\sum_{t=1}^{T}\mathbb{E}(x_{it}\mathbf{z}_t) = \mathbf{0}.$$

■

Lemma S.26 *Let $\{x_{it}\}_{t=1}^{T}$ for $i = 1, 2, \cdots, N$ and $\{z_{jt}\}_{t=1}^{T}$ for $j = 1, 2, \cdots, m$ be time-series processes. Define the projection regression of x_{it} on $\mathbf{z}_t = (z_{1t}, z_{2t}, \cdots, z_{m,t})'$ as*

$$x_{it} = \mathbf{z}_t'\boldsymbol{\psi}_{i,T} + \nu_{it}$$

where $\boldsymbol{\psi}_{i,T} = (\psi_{1i,T}, \psi_{2i,T}, \cdots, \psi_{mi,T})'$ is the $m \times 1$ vector of projection coefficients which is equal to $\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{z}_t\mathbf{z}_t'\right)\right]^{-1}[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t x_{it})]$. Suppose that only a finite number of elements in $\boldsymbol{\psi}_{i,T}$ is different from zero for all $i = 1, 2, \cdots, N$ and there exist sufficiently large positive constants C_0 and C_1, and $s > 0$ such that

(i) $\sup_{j,t}\Pr(|z_{jt}| > \alpha) \leq C_0\exp(-C_1\alpha^s)$, *for all* $\alpha > 0$, *and*

(ii) $\sup_{i,t}\Pr(|x_{it}| > \alpha) \leq C_0\exp(-C_1\alpha^s)$, *for all* $\alpha > 0$.

Then, there exist sufficiently large positive constants C_0 and C_1, and $s > 0$ such that

$$\sup_{i,t}\Pr(|\nu_{it}| > \alpha) \leq C_0\exp(-C_1\alpha^s), \text{ for all } \alpha > 0.$$

Proof. Without loss of generality assume that the first finite ℓ elements of $\psi_{i,T}$ are different from zero and write

$$x_{it} = \sum_{j=1}^{\ell}\psi_{ji,T}z_{jt} + \nu_{it}.$$

Now, note that

$$\Pr(|\nu_{it}| > \alpha) \leq \Pr\left(|x_{it}| + \sum_{j=1}^{\ell}|\psi_{ji,T}z_{jt}| > \alpha\right),$$

and hence by Lemma S.10, for any $0 < \pi_j < 1$, $j = 1, 2, \cdots, \ell + 1$ we have,

$$\begin{aligned}\Pr(|\nu_{it}| > \alpha) &\leq \sum_{j=1}^{\ell}\Pr\left(|\psi_{ji,T}z_{jt}| > \pi_j\alpha\right) + \Pr\left(|x_{it}| > \pi_{\ell+1}\alpha\right)\\ &= \sum_{j=1}^{\ell}\Pr\left(|z_{jt}| > |\psi_{ji,T}|^{-1}\pi_j\alpha\right) + \Pr\left(|x_{it}| > \pi_{\ell+1}\alpha\right)\\ &\leq \ell\sup_{j,t}\Pr\left(|z_{jt}| > |\psi_T^*|^{-1}\pi^*\alpha\right) + \sup_{i,t}\Pr\left(|x_{it}| > \pi^*\alpha\right).\end{aligned}$$

where $\psi_T^* = sup_{i,j}\{\psi_{ji,T}\}$ and $\pi^* = \inf_{j\in 1,2,\cdots,\ell+1}\{\pi_j\}$. Therefore, by the exponential decaying probability tail assumptions for x_{it} and z_{jt} we have

$$\Pr(|\nu_{it}| > \alpha) \leq \ell C_0\exp(-C_1\alpha^s) + C_0\exp(-C_1\alpha^s),$$

and hence there exist sufficiently large positive constants C_0 and C_1 , and $s > 0$ such that

$$\sup_{i,t}\Pr(|\nu_{it}| > \alpha) \leq C_0\exp(-C_1\alpha^s), \text{ for all } \alpha > 0.$$

■

Lemma S.27 *Let $\{x_{it}\}_{t=1}^{T}$ for $i = 1, 2, \cdots, N$ and $\{z_{\ell t}\}_{t=1}^{T}$ for $\ell = 1, 2, \cdots, m$ be time-series processes and $m = \ominus(T^d)$. Also let $\mathcal{F}_{it}^{x} = \sigma(x_{it}, x_{i,t-1}, \cdots)$ for $i = 1, 2, \cdots, N$, $\mathcal{F}_{\ell t}^{z} = \sigma(z_{\ell t}, z_{\ell,t-1}, \cdots)$ for $\ell = 1, 2, \cdots, m$, $\mathcal{F}_{t}^{x} = \cup_{i=1}^{N}\mathcal{F}_{it}^{x}$, $\mathcal{F}_{t}^{z} = \cup_{\ell=1}^{m}\mathcal{F}_{\ell t}^{z}$, and $\mathcal{F}_t = \mathcal{F}_t^x \cup \mathcal{F}_t^z$. Define the projection regression of x_{it} on $\mathbf{z}_t = (z_{1t}, z_{2t}, \cdots, z_{m,t})'$ as*

$$x_{it} = \mathbf{z}_t'\boldsymbol{\psi}_{i,T} + \nu_{it}$$

where $\boldsymbol{\psi}_{i,T} = (\psi_{1i,T}, \psi_{2i,T}, \cdots, \psi_{mi,T})'$ is the $m \times 1$ vector of projection coefficients which is equal to $\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{z}_t\mathbf{z}_t'\right)\right]^{-1}[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t x_{it})]$. Suppose, $\mathbb{E}[x_{it}x_{jt} - \mathbb{E}(x_{it}x_{jt})|\mathcal{F}_{t-1}] = 0$ for all $i, j = 1, 2, \cdots, N$, $\mathbb{E}[z_{\ell t}z_{\ell' t} - \mathbb{E}(z_{\ell t}z_{\ell t})|\mathcal{F}_{t-1}] = 0$ for all $\ell, \ell' = 1, 2, \cdots, m$, and $\mathbb{E}[z_{\ell t}x_{it} - \mathbb{E}(z_{\ell t}x_{it})|\mathcal{F}_{t-1}] = 0$ for all $\ell = 1, 2, \cdots, m$ and for all $i = 1, 2, \cdots, N$. Additionally, assume that only a finite number of elements in $\boldsymbol{\psi}_{i,T}$ is different from zero for all $i = 1, 2, \cdots, N$ and there exist sufficiently large positive constants C_0 and C_1, and $s > 0$ such that

(i) $\sup_{j,t}\Pr(|z_{\ell t}| > \alpha) \leq C_0\exp(-C_1\alpha^s)$, *for all* $\alpha > 0$, *and*

(ii) $\sup_{i,t}\Pr(|x_{\ell t}| > \alpha) \leq C_0\exp(-C_1\alpha^s)$, *for all* $\alpha > 0$.

Then, there exist some finite positive constants C_0, C_1 and C_2 such that if $d < \lambda \leq (s+2)/(s+4)$,

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j - \mathbb{E}(\boldsymbol{\nu}_i'\boldsymbol{\nu}_j)| > \zeta_T) \leq \exp(-C_0T^{-1}\zeta_T^2) + \exp(-C_1T^{C_2})$$

and if $\lambda > (s+2)/(s+4)$

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j - \mathbb{E}(\boldsymbol{\nu}_i'\boldsymbol{\nu}_j)| > \zeta_T) \leq \exp(-C_0\zeta_T^{s/(s+1)}) + \exp(-C_1T^{C_2})$$

for all $i, j = 1, 2, \cdots, N$, where $\boldsymbol{\nu}_i = (\nu_{i1}, \nu_{i2}, \cdots, \nu_{iT})'$, $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iT})'$, and $\mathbf{M}_z = \mathbf{I} - T^{-1}\mathbf{Z}\hat{\boldsymbol{\Sigma}}_{zz}^{-1}\mathbf{Z}'$ with $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_T)'$ and $\hat{\boldsymbol{\Sigma}}_{zz} = T^{-1}\sum_{t=1}^{T}(\mathbf{z}_t\mathbf{z}_t')$.

Proof.

$$\begin{aligned}\Pr\left[|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j - \mathbb{E}(\boldsymbol{\nu}_i'\boldsymbol{\nu}_j)| > \zeta_T\right] &= \Pr\left[|\boldsymbol{\nu}_i'\mathbf{M}_z\boldsymbol{\nu}_j - \mathbb{E}(\boldsymbol{\nu}_i'\boldsymbol{\nu}_j)| > \zeta_T\right] \\ &= \Pr\left[|\boldsymbol{\nu}_i'\boldsymbol{\nu}_j - \mathbb{E}(\boldsymbol{\nu}_i'\boldsymbol{\nu}_j) - T^{-1}\boldsymbol{\nu}_i'\mathbf{Z}\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{Z}'\boldsymbol{\nu}_j - T^{-1}\boldsymbol{\nu}_i'\mathbf{Z}(\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1})\mathbf{Z}'\boldsymbol{\nu}_j| > \zeta_T\right]\end{aligned}$$

where $\boldsymbol{\Sigma}_{zz} = \mathbb{E}[T^{-1}\sum_{t=1}^{T}(\mathbf{z}_t\mathbf{z}_t')]$. By Lemma S.10, we can further write

$$\begin{aligned}&\Pr\left[|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j - \mathbb{E}(\boldsymbol{\nu}_i'\boldsymbol{\nu}_j)| > \zeta_T\right] \\ &\quad\leq \Pr\left[|\boldsymbol{\nu}_i'\boldsymbol{\nu}_j - \mathbb{E}(\boldsymbol{\nu}_i'\boldsymbol{\nu}_j)| > \pi_1\zeta_T\right] + \Pr(|T^{-1}\boldsymbol{\nu}_i'\mathbf{Z}\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{Z}'\boldsymbol{\nu}_j| > \pi_2\zeta_T) + \\ &\qquad\Pr\left[|T^{-1}\boldsymbol{\nu}_i'\mathbf{Z}(\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1})\mathbf{Z}'\boldsymbol{\nu}_j|) > \pi_3\zeta_T\right].\end{aligned}$$

where $0 < \pi_i < 1$ and $\sum_{i=1}^{3} \pi_i = 1$. By Lemma S.16,

$$\Pr(|T^{-1}\boldsymbol{\nu}_i'\mathbf{Z}\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{Z}'\boldsymbol{\nu}_j| > \pi_2\zeta_T) \leq \Pr(\|\boldsymbol{\nu}_i'\mathbf{Z}\|_F\|\boldsymbol{\Sigma}_{zz}^{-1}\|_2\|\mathbf{Z}'\boldsymbol{\nu}_j\|_F > \pi_2\zeta_T T),$$

and again by Lemma S.11, we have

$$\begin{aligned}&\Pr(|T^{-1}\boldsymbol{\nu}_i'\mathbf{Z}\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{Z}'\boldsymbol{\nu}_j| > \pi_2\zeta_T)\\ &\quad\leq \Pr(\|\boldsymbol{\nu}_i'\mathbf{Z}\|_F > \|\boldsymbol{\Sigma}_{zz}^{-1}\|_2^{-1/2}\pi_2^{1/2}\zeta_T^{1/2}T^{1/2}) + \Pr(\|\mathbf{Z}'\boldsymbol{\nu}_j\|_F > \|\boldsymbol{\Sigma}_{zz}^{-1}\|_2^{-1/2}\pi_2^{1/2}\zeta_T^{1/2}T^{1/2}).\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}&\Pr(|T^{-1}\boldsymbol{\nu}_i'\mathbf{Z}(\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1})\mathbf{Z}'\boldsymbol{\nu}_j| > \pi_3\zeta_T)\\ &\quad\leq \Pr(\|\boldsymbol{\nu}_i'\mathbf{Z}\|_F\|\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1}\|_F\|\mathbf{Z}'\boldsymbol{\nu}_j\|_F > \pi_3\zeta_T T)\\ &\quad\leq \Pr(\|\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1}\zeta_T) + \Pr(\|\boldsymbol{\nu}_i'\mathbf{Z}\|_F > \pi_3^{1/2}\delta_T^{1/2}T^{1/2})\\ &\qquad+ \Pr(\|\mathbf{Z}'\boldsymbol{\nu}_j\|_F > \pi_3^{1/2}\delta_T^{1/2}T^{1/2})\end{aligned}$$

where $\delta_T = \ominus(T^{\alpha})$ with $0 < \alpha < \lambda$.

Note that $\Pr(\|\mathbf{Z}'\boldsymbol{\nu}_i\|_F > c) = \Pr(\|\mathbf{Z}'\boldsymbol{\nu}_i\|_F^2 > c^2) = \Pr[\sum_{\ell=1}^{m}(\sum_{t=1}^{T}\nu_{it}z_{\ell t})^2 > c^2]$, where c is a positive constant. So, by Lemma S.10, we have

$$\Pr(\|\mathbf{Z}'\boldsymbol{\nu}_i\|_F > c) \leq \sum_{\ell=1}^{m}\Pr[(\sum_{t=1}^{T}\nu_{it}z_{\ell t})^2 > m^{-1}c^2]$$

Hence, $\Pr(\|\mathbf{Z}'\boldsymbol{\nu}_i\|_F > c) \leq \sum_{\ell=1}^{m}\Pr(|\sum_{t=1}^{T}\nu_{it}z_{\ell t}| > m^{-1/2}c)$. Also, by Lemma S.25 we have $\sum_{t=1}^{T}\mathbb{E}(\nu_{it}z_{\ell t}) = 0$ and hence we can further write

$$\Pr(\|\mathbf{Z}'\boldsymbol{\nu}_i\|_F > c) \leq \sum_{\ell=1}^{m}\Pr\{|\sum_{t=1}^{T}[\nu_{it}z_{\ell t} - \mathbb{E}(\nu_{it}z_{\ell t})]| > m^{-1/2}c\}.$$

Note that $\|\boldsymbol{\Sigma}_{zz}^{-1}\|_2$ is equal to the largest eigenvalue of $\boldsymbol{\Sigma}_{zz}^{-1}$ and it is a finite positive constant. So, there exists a positive constant $C > 0$ such that,

$$\begin{aligned}&\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j - \mathbb{E}(\boldsymbol{\nu}_i'\boldsymbol{\nu}_j)| > \zeta_T)\\ &\quad\leq \Pr\{|\sum_{t=1}^{T}[\nu_{it}\nu_{jt} - \mathbb{E}(\nu_{it}\nu_{jt})]| > CT^{\lambda}\}+\\ &\qquad\sum_{\ell=1}^{m}\Pr\{|\sum_{t=1}^{T}[\nu_{it}z_{\ell t} - \mathbb{E}(\nu_{it}z_{\ell t})]| > CT^{1/2+\lambda/2-d/2}\}+\\ &\qquad\sum_{\ell=1}^{m}\Pr\{|\sum_{t=1}^{T}[\nu_{jt}z_{\ell t} - \mathbb{E}(\nu_{jt}z_{\ell t})]| > CT^{1/2+\lambda/2-d/2}\}+\\ &\qquad\sum_{\ell=1}^{m}\Pr\{|\sum_{t=1}^{T}[\nu_{it}z_{\ell t} - \mathbb{E}(\nu_{it}z_{\ell t})]| > CT^{1/2+\alpha/2-d/2}\}+\\ &\qquad\sum_{\ell=1}^{m}\Pr\{|\sum_{t=1}^{T}[\nu_{jt}z_{\ell t} - \mathbb{E}(\nu_{jt}z_{\ell t})]| > CT^{1/2+\alpha/2-d/2}\}+\\ &\qquad\Pr(\|\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1}\zeta_T)\end{aligned}$$

Let

$$\kappa_{T,i}(h,d) = \sum_{\ell=1}^{m}\Pr\{|\sum_{t=1}^{T}[\nu_{it}z_{\ell t} - \mathbb{E}(\nu_{it}z_{\ell t})]| > CT^{1/2+\kappa/2-d/2}\}, \text{ for } h = \lambda, \alpha,$$

and $i = 1, 2, ..., N$. By Lemmas S.14, S.25, and S.26, we have $\nu_{it}\nu_{jt} - \mathbb{E}(\nu_{it}\nu_{jt})$ and $\nu_{it}z_{\ell t} - \mathbb{E}(\nu_{it}z_{\ell t})$ are martingale difference processes with exponentially bounded probability tail, $\frac{s}{2}$. So, depending on the value of exponentially bounded probability tail parameter, from Lemma S.8, we know that either

$$\kappa_{T,i}(h, d) \leq m \exp\left[-\ominus\left(T^{h-d}\right)\right]$$

or

$$\kappa_{T,i}(h, d) \leq m \exp\left[-\ominus\left(T^{s(1/2+h/2-d/2)/(s+2)}\right)\right],$$

for $h = \lambda, \alpha$. Also, depending on the value of exponentially bounded probability tail parameter, from Lemmas S.23 and S.24 we have,

$$\Pr(\|\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1}\zeta_T) \leq m^2 \exp\left[-C_0 \frac{T\delta_T^{-2}\zeta_T^2}{m^2\|\boldsymbol{\Sigma}_{zz}^{-1}\|_2^2(\|\boldsymbol{\Sigma}_{zz}^{-1}\|_2 + \delta_T^{-1}\zeta_T)^2}\right] + m^2 \exp\left(-C_0 \frac{T}{m^2\|\boldsymbol{\Sigma}_{zz}^{-1}\|_2^2}\right),$$

or

$$\Pr(\|\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1}\zeta_T) \leq m^2 \exp\left[-C_0 \frac{(T\delta_T^{-1}\zeta_T)^{s/s+2}}{m^{s/s+2}\|\ \boldsymbol{\Sigma}_{zz}^{-1}\|_2^{s/s+2}(\|\boldsymbol{\Sigma}_{zz}^{-1}\|_2 + \delta_T^{-1}\zeta_T)^{s/s+2}}\right] + m^2 \exp\left(-C_0 \frac{T^{s/s+2}}{m^{s/s+2}\|\boldsymbol{\Sigma}_{zz}^{-1}\|_2^{s/s+2}}\right).$$

Therefore,

$$\begin{aligned}
&\Pr(\|\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1}\zeta_T) \\
&\quad \leq m \exp[-\ominus(T^{\max\{1-2d+2(\lambda-\alpha),1-2d+\lambda-\alpha,1-2d\}})] + \\
&\quad\ \ m \exp[-\ominus(T^{1-2d})],
\end{aligned}$$

or,

$$\begin{aligned}
&\Pr(\|\hat{\boldsymbol{\Sigma}}_{zz}^{-1} - \boldsymbol{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1}\zeta_T) \\
&\quad \leq m \exp[-\ominus(T^{s(\max\{1-d+\lambda-\alpha,1-d\})/(s+2)})] + \\
&\quad\ \ m \exp[-\ominus(T^{s(1-d)/(s+2)})].
\end{aligned}$$

Setting $d < 1/2$, $\alpha = 1/2$, and $\lambda > d$, we have all the terms going to zero as $T \to \infty$ and there exist some finite positive constants C_1 and C_2 such that

$$\kappa_{T,i}(\lambda, d) \leq \exp\left(-C_1 T^{C_2}\right),\ \kappa_{T,i}(\alpha, d) \leq \exp\left(-C_1 T^{C_2}\right),$$

and

$$\Pr(\|\hat{\mathbf{\Sigma}}_{zz}^{-1} - \mathbf{\Sigma}_{zz}^{-1}\|_F > \delta_T^{-1}\zeta_T) \leq \exp(-C_1 T^{C_2}).$$

Hence, if $d < \lambda \leq (s+2)/(s+4)$,

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j - \mathbb{E}(\boldsymbol{\nu}_i'\boldsymbol{\nu}_j)| > \zeta_T) \leq \exp(-C_0 T^{-1}\zeta_T^2) + \exp(-C_1 T^{C_2}),$$

and if $\lambda > (s+2)/(s+4)$,

$$\Pr(|\mathbf{x}_i'\mathbf{M}_z\mathbf{x}_j - \mathbb{E}(\boldsymbol{\nu}_i'\boldsymbol{\nu}_j)| > \zeta_T) \leq \exp(-C_0 \zeta_T^{s/(s+1)}) + \exp(-C_1 T^{C_2}),$$

where C_0, C_1 and C_2 are some finite positive constants. ■

Lasso, Adaptive Lasso and Cross-validation algorithms

This section explains how Lasso, K-fold cross-validation and Adaptive Lasso are implemented in this paper. Let $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$ be a $T \times 1$ vector of target variable, and let $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_T)'$ be a $T \times m$ matrix of conditioning covariates where $\{\mathbf{z}_t : t = 1, 2, \cdots, T\}$ are $m \times 1$ vectors and let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T)'$ be a $T \times N$ matrix of covariates in the active set where $\{\mathbf{x}_t : t = 1, 2, \cdots, T\}$ are $N \times 1$ vectors.

Lasso Procedure

1. *Construct the filtered variables* $\tilde{\mathbf{y}} = \mathbf{M}_z\mathbf{y}$ *and* $\tilde{\mathbf{X}} = \mathbf{M}_z\mathbf{X} = (\tilde{\mathbf{x}}_{1\circ}, \tilde{\mathbf{x}}_{2\circ}, ..., \tilde{\mathbf{x}}_{N\circ})$, *where* $\mathbf{M}_z = \mathbf{I}_T - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, *and* $\tilde{\mathbf{x}}_{i\circ} = (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{iT})'$.

2. *Normalize each covariate* $\tilde{\mathbf{x}}_{i\circ} = (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{iT})'$ *by its* ℓ_2 *norm, such that*

   $$\tilde{\mathbf{x}}_{i\circ}^* = \tilde{\mathbf{x}}_{i\circ}/\|\tilde{\mathbf{x}}_{i\circ}\|_2,$$

   where $\|.\|_2$ *denotes the* ℓ_2 *norm of a vector. The corresponding matrix of normalized covariates in the active set is now denoted by* $\tilde{\mathbf{X}}^*$.

3. *For a given value of* $\varphi \geq 0$, *find* $\hat{\boldsymbol{\gamma}}_x^*(\varphi) \equiv [\hat{\gamma}_{1x}^*(\varphi), \hat{\gamma}_{2x}^*(\varphi), \cdots, \hat{\gamma}_{Nx}^*(\varphi)]'$ *such that*

   $$\hat{\boldsymbol{\gamma}}_x^*(\varphi) = \arg\min_{\boldsymbol{\gamma}_x^*} \left\{ \|\tilde{\mathbf{y}} - \tilde{\mathbf{X}}^*\boldsymbol{\gamma}_x^*\|_2^2 + \varphi\|\boldsymbol{\gamma}_x^*\|_1 \right\},$$

   where $\|.\|_1$ *denotes the* ℓ_1 *norm of a vector.*

4. *Divide* $\hat{\gamma}_{ix}^*(\varphi)$ *for* $i = 1, 2, \cdots, N$ *by* ℓ_2 *norm of the* $\tilde{\mathbf{x}}_{i\circ}$ *to match the original scale of* $\tilde{\mathbf{x}}_{i\circ}$, *namely set*

   $$\hat{\gamma}_{ix}(\varphi) = \hat{\gamma}_{ix}^*(\varphi)/\|\tilde{\mathbf{x}}_{i\circ}\|_2,$$

where $\hat{\boldsymbol{\gamma}}_x(\varphi) \equiv [\hat{\gamma}_{1x}(\varphi), \hat{\gamma}_{2x}(\varphi), \cdots, \hat{\gamma}_{Nx}(\varphi)]'$ *denotes the vector of scaled coefficients.*

5. *Compute* $\hat{\boldsymbol{\gamma}}_z(\varphi) \equiv [\hat{\gamma}_{1z}(\varphi), \hat{\gamma}_{2z}(\varphi), \cdots, \hat{\gamma}_{mz}(\varphi)]'$ *by* $\hat{\boldsymbol{\gamma}}_z(\varphi) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{e}}(\varphi)$ *where* $\hat{\mathbf{e}}(\varphi) = \tilde{\mathbf{y}} - \tilde{\mathbf{X}}\hat{\boldsymbol{\gamma}}_x(\varphi)$.

For a given set of values of φ's, say $\{\varphi_j : j = 1, 2, \cdots, h\}$, the optimal value of φ is chosen by K-fold cross-validation as described below.

K-fold Cross-validation

1. *Create a* $T \times 1$ *vector* $\mathbf{w} = (1, 2, \cdots, K, 1, 2, \cdots, K, \cdots)'$ *where* K *is the number of folds.*

2. *Let* $\mathbf{w}^* = (w_1^*, w_2^*, \cdots, w_T^*)'$ *be a* $T \times 1$ *vector generated by randomly permuting the elements of* $\mathbf{w}$.

3. *Group observations into* K *folds such that*

$$g_k = \{t : t \in \{1, 2, \cdots, T\} \text{ and } w_t^* = k\} \text{ for } k = 1, 2, \cdots, K.$$

4. *For a given value of* φ_j *and each fold* $k \in \{1, 2, \cdots, K\}$,

   (a) *Remove the observations related to fold* k *from the set of all observations.*

   (b) *Given the value of* φ_j, *use the remaining observations to estimate the coefficients of the model.*

   (c) *Use the estimated coefficients to compute predicted values of the target variable for the observations in fold* k *and hence compute mean square forecast error of fold* k *denoted by* $MSFE_k(\varphi_j)$.

5. *Compute the average mean square forecast error for a given value of* φ_j *by*

$$\overline{MSFE}(\varphi_j) = \sum_{k=1}^{K} MSFE_k(\varphi_j)/K.$$

6. *Repeat steps 1 to 5 for all values of* $\{\varphi_j : j = 1, 2, \cdots, h\}$.

7. *Select* φ_j *with the lowest corresponding average mean square forecast error as the optimal value of* φ.

In this study, following Friedman et al. (2010), we consider a sequence of 100 values of φ's decreasing from $\varphi_{\max}$ to $\varphi_{\min}$ on log scale where $\varphi_{\max} = \max_{i=1,2,\cdots,N}\left\{|\sum_{t=1}^{T}\tilde{x}_{it}^*\tilde{y}_t|\right\}$ and $\varphi_{\min} = 0.001\varphi_{\max}$. We use 10-fold cross-validation ($K = 10$) to find the optimal value of φ.

Denote $\hat{\boldsymbol{\gamma}}_x \equiv \hat{\boldsymbol{\gamma}}_x(\varphi_{op})$ where φ_{op} is the optimal value of φ obtained by the K-fold cross-validation. Given $\hat{\boldsymbol{\gamma}}_x$, we implement Adaptive Lasso as described below.

Adaptive Lasso Procedure

1. *Let* $\mathcal{S} = \{i : i \in \{1, 2, \cdots, N\}$ *and* $\hat{\gamma}_{ix} \neq 0\}$ *and* $\mathbf{X}_{\mathcal{S}}$ *be the* $T \times s$ *set of covariates in the active set with* $\hat{\gamma}_{ix} \neq 0$ *(from the Lasso step) where* $s = |\mathcal{S}|$*. Additionally, denote the corresponding* $s \times 1$ *vector of non-zero Lasso coefficients by* $\hat{\boldsymbol{\gamma}}_{x,\mathcal{S}} = (\hat{\gamma}_{1x,\mathcal{S}}, \hat{\gamma}_{2x,\mathcal{S}}, \cdots, \hat{\gamma}_{sx,\mathcal{S}})'$.

2. *For a given value of* $\psi \geq 0$*, find* $\hat{\boldsymbol{\delta}}^*_{x,\mathcal{S}}(\psi) \equiv [\hat{\delta}^*_{1x,\mathcal{S}}(\psi), \hat{\delta}^*_{2x,\mathcal{S}}(\psi), \cdots, \hat{\delta}^*_{sx,\mathcal{S}}(\psi)]'$ *such that*

$$\hat{\boldsymbol{\delta}}^*_{x,\mathcal{S}}(\psi) = \arg\min_{\boldsymbol{\delta}^*_{x,\mathcal{S}}} \left\{ \|\tilde{\mathbf{y}} - \tilde{\mathbf{X}}_{\mathcal{S}} diag(\hat{\boldsymbol{\gamma}}_{x,\mathcal{S}}) \boldsymbol{\delta}^*_{x,\mathcal{S}}\|_2^2 + \psi \|\boldsymbol{\delta}^*_{x,\mathcal{S}}\|_1 \right\},$$

   where $diag(\hat{\boldsymbol{\gamma}}_{x,\mathcal{S}})$ *is an* $s \times s$ *diagonal matrix with its diagonal elements given by the corresponding elements of* $\hat{\boldsymbol{\gamma}}_{x,\mathcal{S}}$.

3. *Post multiply* $\hat{\boldsymbol{\delta}}^*_{x,\mathcal{S}}(\psi)$ *by* $diag(\hat{\boldsymbol{\gamma}}_{x,\mathcal{S}})$ *to match the original scale of* $\tilde{\mathbf{X}}_{\mathcal{S}}$*, such that*

$$\hat{\boldsymbol{\delta}}_{x,\mathcal{S}}(\psi) = diag(\hat{\boldsymbol{\gamma}}_{x,\mathcal{S}}) \hat{\boldsymbol{\delta}}^*_{x,\mathcal{S}}(\psi).$$

   The coefficients of the covariates in the active set that belong to $\mathcal{S}^c$ *are set equal to zero. In other words,* $\hat{\boldsymbol{\delta}}_{x,\mathcal{S}^c}(\psi) = 0$ *for all* $\psi \geq 0$.

4. *Compute* $\hat{\boldsymbol{\delta}}_z(\psi) \equiv [\hat{\delta}_{1z}(\psi), \hat{\delta}_{2z}(\psi), \cdots, \hat{\delta}_{mz}(\psi)]'$ *by* $\hat{\boldsymbol{\delta}}_z(\psi) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{e}}(\psi)$ *where* $\hat{\mathbf{e}}(\psi) = \tilde{\mathbf{y}} - \tilde{\mathbf{X}}_{\mathcal{S}} \hat{\boldsymbol{\delta}}_{x,\mathcal{S}}(\psi)$.

As in the Lasso step, the optimal value ψ is set using 10-fold cross-validation as described before.[10]

[10] To implement Lasso, Adaptive Lasso and 10-fold cross-validation we take advantage of glmnet package (Matlab version) available at `http://web.stanford.edu/~hastie/glmnet_matlab/`

References

Bailey, N., Pesaran, M. H., and Smith, L. V. (2019). A multiple testing approach to the regularisation of large sample correlation matrices. *Journal of Econometrics*, 208(2): 507-534. https://doi.org/10.1016/j.jeconom.2018.10.006

Chudik, A., Kapetanios, G., and Pesaran, M. H. (2018). A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models. *Econometrica*, 86(4): 1479-1512. https://doi.org/10.3982/ECTA14176

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1-22. https://doi.org/10.18637/jss.v033.i01

Lütkepohl, H. (1996). *Handbook of Matrices.* John Wiley & Sons, West Sussex, UK. ISBN-10: 9780471970156