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

Research Department

https://doi.org/10.24149/gwp394supp

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 Southern California

July 23, 2020

This online theory supplement has two sections. First section provides the complementary lemmas needed for the proofs of the lemmas in Section A.2 of the paper. Second section explains the algorithms used for implementing Lasso, Adaptive Lasso and Cross-validation.

Complementary Lemmas

Lemma S.1 *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*

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

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.2 *Let*

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

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.3 *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.4 *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.5 *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.6 *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.7 *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.4, 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.8 *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.9 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $n \times p$ *and* $p \times m$ *matrices respectively, then*

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

where $\|.\|_F$ *denotes the Frobenius norm and* $\|.\|_2$ *denotes the spectral norm. Moreover,*

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

Proof.

$$\|\mathbf{AB}\|_F^2 = \text{tr}(\mathbf{ABB'A'}) = \text{tr}[\mathbf{A}(\mathbf{BB'})\mathbf{A'}]$$

By result (12) at page 44 of Lütkepohl (1996),

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

where $\lambda_{\max}(\mathbf{BB'})$ is the largest eigenvalue of $\mathbf{BB'}$. Therefore,

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

Similarly,

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

and hence

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

■

Lemma S.10 *Let* z_{ij} *be a random variable for* $i = 1, 2, \cdots, N$, *and* $j = 1, 2, \cdots, N$. *Then, for any* $d_T > 0$,

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

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

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

■

Lemma S.11 *Consider two $N \times N$ nonsingular matrices $\mathbf{A}$ and $\mathbf{B}$ such that*

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

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.9,

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

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,

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

Hence,

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

Since $\|\mathbf{B}^{-1}\|_2 \|B - A\|_F \leq 1$, 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.12 *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.11 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.10, 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,

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

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

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

Therefore,

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

■

Lemma S.13 *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*

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

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

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

Proof. The proof is similar to the proof of Lemma S.12. ■

Lemma S.14 *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* $\left[T^{-1} \sum_{t=1}^T \mathbb{E}\left(\mathbf{z}_t \mathbf{z}_t'\right)\right]^{-1} [T^{-1} \sum_{t=1}^T \mathbb{E}(\mathbf{z}_t x_{it})]$. *Suppose,* $\mathbb{E}[x_{it} x_{i't} - \mathbb{E}(x_{it} x_{i't}) | \mathcal{F}_{t-1}] = 0$ *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.

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

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

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

■

Lemma S.15 *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* $\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{z}_t\mathbf{z}_t'\right)\right]^{-1}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}(\mathbf{z}_t x_{it})\right]$*. 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*

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

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.3, 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

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

■

Lemma S.16 *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* $d < \lambda \leq (s+2)/(s+4)$*,*

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

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

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

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.

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

where $\boldsymbol{\Sigma}_{zz} = \mathbb{E}[T^{-1}\sum_{t=1}^T (\mathbf{z}_t \mathbf{z}_t')]$. By Lemma S.3, we can further write

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

where $0 < \pi_i < 1$ and $\sum_{i=1}^{3} \pi_i = 1$. By Lemma S.9,

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

and again by Lemma S.4, 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.3, 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.14 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.7, S.14, and S.15, 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.1, 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.12 and S.13 we have,

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

or

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

Therefore,

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

or,

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

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

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

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