Supplementary Material : Determination of Vector Error Correction Models in High Dimensions

Chong Liang Melanie Schienle

August 16, 2018

1 Technical Lemmas

Lemma 1. *Let $A \in \mathbb{R}^{m\times k}, B \in \mathbb{R}^{k\times n}$ with $\sigma_1(A), \sigma_1(B)$ denoting the largest singular value. $C \in \mathbb{R}^{m\times m}$ is non-singular with largest/smallest singular value denoted by $\sigma_1(C)/\sigma_m(C)$ Given T observations the estimators for A, B, C are denoted as $\widetilde{A}, \widetilde{B}, \widetilde{C}$ and satisfy*

$$||\widetilde{A} - A||_F = O(q(T)), \qquad ||\widetilde{B} - B||_F = O(q(T)), \qquad ||\widetilde{C} - C||_F = O(q(T))$$

with $q(T) \to 0$ as $T \to \infty$, then

$$\begin{aligned} ||\widetilde{A}\widetilde{B} - AB||_F &= O(\max(\sigma_1(A), \sigma_1(B))q(T)) \\ ||\widetilde{C}^{-1} - C^{-1}||_F &= O(\sigma_m^{-2}(C)q(T)) \end{aligned}$$

Proof. By the Mirsky version of matrix perturbation theory (see Theorem 4.11 of Stewart and Sun (1990)), the singular values of the estimated matrix are consistent for those of the true matrix, i.e.,

$$|\sigma_j(\widetilde{A}) - \sigma_j(A)| = O(q(T))$$

Therefore,

$$\begin{aligned} ||\widetilde{A}\widetilde{B} - AB||_F &= ||(\widetilde{A} - A)\widetilde{B} + A(\widetilde{B} - B)||_F \\ &\leqslant ||(\widetilde{A} - A)||_F ||\widetilde{B}||_2 + ||A||_2 ||(\widetilde{B} - B)||_F \\ &= O(\sigma_1(B)q(T) + \sigma_1(A)q(T)) \end{aligned}$$

The argument can be proved by showing that

$$\begin{aligned} ||\widetilde{C}^{-1} - C^{-1}||_F &= ||\widetilde{C}^{-1}(C - \widetilde{C})C^{-1}||_F \\ &\leqslant ||\widetilde{C}^{-1}||_2 ||(\widetilde{C} - C)||_F ||C^{-1}||_2 \\ &= O(\sigma_m(C)^{-2}q(T)) \end{aligned}$$

□

Lemma 2. *Under Assumptions 2.1, 2.2 and 2.3, and $m = (T^{1/4-\varepsilon})$ with $\varepsilon \in (0, 1/4]$ the following results hold:*

1. $||\frac{1}{T}\sum_{t=1}^{T} \Delta Y_t \Delta X'_{t-1}||_2 = O_P(1)$ *and* $||\frac{1}{T}\sum_{t=1}^{T} \beta' Y_{t-1}\Delta X'_{t-1}||_2 = O_P(1)$.
2. $||\frac{1}{T}\sum_{t=1}^{T} \Delta X_{t-1}\Delta X'_{t-1}||_2 = O_P(1)$ *and* $||(\frac{1}{T}\sum_{t=1}^{T} \Delta X_{t-1}\Delta X'_{t-1})^{-1}||_2 = O_P(1)$.
3. $||\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t \Delta X'_{t-1}||_2 = O_P(1)$ *and thus* $||\frac{1}{T}\sum_{t=1}^{T} \widetilde{w}_t\widetilde{w}'_t - \Sigma_w||_F = O_p(\frac{m}{\sqrt{T}})$.

Proof. To simplify the analysis, we rewrite the general VAR process in (2) as a VAR(1) process by defining:

$$\begin{aligned} F_t^1 &= [Y'_t, \Delta Y'_t, \ldots, \Delta Y'_{t-p+1}]' \\ F_t^0 &= [Z'_{1t}, \Delta Y'_t, \ldots, \Delta Y'_{t-p+1}]' \end{aligned}$$

Then we get from (2) and the stationary components after Q-transformation of (2) that

$$F_t^1 = \begin{pmatrix} \Pi + I_m & B_1 & \ldots & B_{p-1} & B_p \\ \Pi & B_1 & \ldots & B_{p-1} & B_p \\ 0 & I_m & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & I_m & 0 \end{pmatrix} F_{t-1}^1 + \begin{pmatrix} w_t \\ w_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{1}$$

$$F_t^0 = \begin{pmatrix} \beta'\alpha + I_r & \beta' B_1 & \ldots & \beta' B_{p-1} & \beta' B_p \\ \alpha & B_1 & \ldots & B_{p-1} & B_p \\ 0 & I_m & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & I_m & 0 \end{pmatrix} F_{t-1}^0 + \begin{pmatrix} \beta' w_t \\ w_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{2}$$

Setting the matrix in (1) as Φ_1, the cointegrated process F_t^1 has the compact VAR(1) representation

$$F_t^1 = \Phi_1 F_{t-1}^1 + [w'_t, w'_t, 0'_{m(p-1)}]' = \sum_{j=0}^{\infty} \Phi_1^j [w'_{t-j}, w'_{t-j}, 0'_{m(p-1)}]'.$$

The VMA(∞) representation holds as the $m(p+1)$-dimensional square matrix Φ_1 has $m - r$ eigenvalues on the unit circle and all the others within the unit circle due to Assumptions 2.2 and 2.3. In a similar way, denoting by Φ_0 the matrix in (2) we get

$$F_t^0 = \Phi_0 F_{t-1}^0 + [v'_{1t}, w'_t, 0'_{m(p-1)}]' = \sum_{j=0}^{\infty} \Phi_0^j [v'_{1,t-j}, w'_{t-j}, 0'_{m(p-1)}]'$$

with $\lambda_1(\Phi_0) < 1$ and $||\Phi_0||_2 < \infty$. Define $\tilde{v}_t = [v'_{1,t-j}, w'_{t-j}, 0'_{m(p-1)}]'$ with covariance matrix $\Sigma_{\tilde{v}}$, then with $m = (T^{1/4-\varepsilon})$ for some finite K large enough: according to Chen, Xu and Wu (2013), we have

$$||\frac{1}{T}\sum_{t=1}^{T} F_t^0 F_t^{0\prime} - \mathbf{E}(F_t^0 F_t^{0\prime})||_2 = O_P(m/\sqrt{T})$$

while

$$\begin{aligned}||\mathbf{E}(F_t^0 F_t^{0\prime})||_2 &= ||\sum_{j=0}^{\infty}\Phi_0^j\Sigma_{\tilde{v}}\Phi_0^{j\prime}||_2\\ &\leqslant \sum_{j=0}^{\infty}||\Phi_0^j||_2^2||\Sigma_{\tilde{v}}||_2\\ &\leqslant ||\Sigma_{\tilde{v}}||_2\left(\sum_{j=0}^{K}||\Phi_0^j||_2^2+\sum_{j=K+1}^{\infty}||\Phi_0^j||_2^2\right)\end{aligned}$$

where $||\Sigma_{\tilde{v}}||_2$ is bounded due to $\lambda_1(\Sigma_w) < \infty$ by Assumption 2.1. Moreover, $\sum_{j=0}^{K}||\Phi_0^j||_2^2$ is bounded for finite K and $\sum_{j=K+1}^{\infty}||\Phi_0^j||_2^2$ is bounded due to Gelfand's formula, since $||\Phi_0^K||^{1/K} \leqslant \lambda_1(\Phi_0) + \epsilon(K) < 1$ for sufficiently large K. Thus $||\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}||_2 = O_P(1)$ which implies points 1. in the Lemma. The first result in point 2. can also be derived from this conclusion. To show the second result in point 2., we can rewrite ΔX_t as

$$\Delta X_t = A\begin{pmatrix} Z_{1,t-1}\\ \Delta X_{t-1}\\ \vdots\\ Z_{1,t-p-1}\\ \Delta X_{t-p-1}\end{pmatrix} + \begin{pmatrix} w_t\\ w_{t-1}\\ \vdots\\ w_{t-p}\end{pmatrix}$$

where A is a block-diagonal matrix with p blocks of $[\alpha, B]$ on the diagonal. Thus

$$\mathbf{E}(\Delta X_t\Delta X_t') = Cov(\begin{pmatrix} w_t\\ w_{t-1}\\ \vdots\\ w_{t-p}\end{pmatrix}) + RT_1$$

Because the smallest eigenvalue of $Cov(\begin{pmatrix} w_t\\ w_{t-1}\\ \vdots\\ w_{t-p}\end{pmatrix})$ is bounded away from zero by $\tau_w > 0$ and other terms contained RT_1 is symmetric matrix, thus all the eigenvalues are non-negative. Therefore, the smallest eigenvalue of $\mathbf{E}(\Delta X_t\Delta X_t')$ is larger than τ_w.
To show the last point, we can show that

$$\begin{aligned}&||\frac{1}{T}\sum_{t=1}^{T}\widetilde{w}_t\widetilde{w}_t' - \Sigma_{ww}||_F\\ =\;& ||\frac{1}{T}\sum_{t=1}^{T}w_tw_t' - \Sigma_{ww} - (\frac{1}{T}\sum_{t=1}^{T}w_t\Delta X_{t-1}')(\frac{1}{T}\Delta X_{t-1}\Delta X_{t-1}')^{-1}(\frac{1}{T}\sum_{t=1}^{T}\Delta X_{t-1}w_t')||_F\\ \leqslant\;& ||\frac{1}{T}\sum_{t=1}^{T}w_tw_t' - \Sigma_{ww}||_F + \frac{1}{T}||(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w_t\Delta X_{t-1}')(\frac{1}{T}\Delta X_{t-1}\Delta X_{t-1}')^{-1}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta X_{t-1}w_t')||_F\\ =\;& O_p(\frac{m}{\sqrt{T}}) + O_p(\frac{m}{T}) = O_p(\frac{m}{\sqrt{T}})\end{aligned}$$

as the first term in the second to last line $||\frac{1}{T}\sum_{t=1}^{T} w_t w_t' - \Sigma_{ww}||_F = O_P(\frac{m}{\sqrt{T}})$ due to a standard law of large numbers for stationary time series. For the last term in that line, note that the l_2-norm of the expression inside the norm is $O_P(1)$, which implies that the stated Frobenius-norm is at most $O_P(m)$. □

Lemma 3. *Let the assumptions of Theorem 3.1 hold and $\breve{Y}, \breve{X}$, and $\breve{w}$ are as defined in* (9)*. Then*

$$||\frac{1}{T}\sum_{t=1}^{T} \Delta\breve{Y}_t \Delta\breve{Y}_t' - \Sigma_{\Delta y.z1}||_F = O_p(\frac{m}{\sqrt{T}})$$
$$||\frac{1}{T}\sum_{t=1}^{T} \Delta\breve{X}_{t-1} \Delta\breve{X}_{t-1}' - \Sigma_{\Delta x.z1}||_F = O_p(\frac{m}{\sqrt{T}})$$
$$||\frac{1}{T}\sum_{t=1}^{T} \breve{w}_t \breve{w}_t' - \Sigma_{w}||_F = O_p(\frac{m}{\sqrt{T}})$$

Proof. Let $D_{0T} = diag\{\sqrt{T}I_r, TI_{m-r}\}$ and the matrix Q as defined for (5). Then we can write the transformation for lag selection C from (9) as

$$\begin{aligned} C &= I_T - Y_{-1}'(Y_{-1}Y_{-1}')^{-1}Y_{-1} = I_T - Y_{-1}'Q'D_{0T}^{-1}(D_{0T}^{-1}QY_{-1}Y_{-1}'Q'D_{0T}^{-1})^{-1}D_{0T}^{-1}QY_{-1} \\ &= I_T - Z_{-1}'D_{0T}^{-1}(D_{0T}^{-1}Z_{-1}Z_{-1}'D_{0T}^{-1})^{-1}D_{0T}^{-1}Z_{-1} \end{aligned}$$

We therefore obtain

$$\Delta\breve{Y}_t = \Delta Y_t - (\textstyle\sum_{t=1}^{T} \Delta Y_t Z_{t-1}' D_{0T}^{-1})(D_{0T}^{-1}Z_{-1}Z_{-1}'D_{0T}^{-1})^{-1}D_{0T}^{-1}Z_{t-1}$$

$$\Delta\breve{X}_{t-1} = \Delta X_t - (\textstyle\sum_{t=1}^{T} \Delta X_{t-1} Z_{t-1}' D_{0T}^{-1})(D_{0T}^{-1}Z_{-1}Z_{-1}'D_{0T}^{-1})^{-1}D_{0T}^{-1}Z_{t-1}$$

$$\breve{w}_t = w_t - (\textstyle\sum_{t=1}^{T} w_t Z_{t-1}' D_{0T}^{-1})(D_{0T}^{-1}Z_{-1}Z_{-1}'D_{0T}^{-1})^{-1}D_{0T}^{-1}Z_{t-1}$$

Denote $S^z = D_{0T}^{-1}Z_{-1}Z_{-1}'D_{0T}^{-1}$ and

$$\begin{aligned} S_{11}^z &= \frac{1}{T}\sum_{t=1}^{T} Z_{1,t-1}Z_{1,t-1}' \\ S_{12}^z &= \frac{1}{T^{3/2}}\sum_{t=1}^{T} Z_{1,t-1}Z_{2,t-1}' \\ S_{22}^z &= \frac{1}{T^2}\sum_{t=1}^{T} Z_{2,t-1}Z_{2,t-1}' \end{aligned}$$

Then $||S_{11}^z||_2 = O_P(1)$ by Lemma 2. With χ_{12} of block 6 in the proof of Theorem 2.1, we have $||S_{12}^z||_2 = ||\chi_{12}/\sqrt{T}||_2 + o_P(1)$ due to Lemma 2. Hence (A.12) implies that $||S_{12}^z||_2 = O_p(\frac{r^{\tau_1}}{\sqrt{T}})$. In the same way, Lemma 2 yields $||S_{22}^z||_2 = O_p(1)$.

The inverse $S^{z,-1}$ of S^z has the following blockwise form

$$S^{z,-1} = \begin{bmatrix} (S_{11}^z - S_{12}^z S_{22}^{z,-1} S_{21}^z)^{-1} & -(S_{11}^z - S_{12}^z S_{22}^{z,-1} S_{21}^z)^{-1} S_{12}^z S_{22}^{z,-1} \\ -(S_{22}^z - S_{21}^z S_{11}^{z,-1} S_{12}^z)^{-1} S_{21}^z S_{11}^{z,-1} & (S_{22}^z - S_{21}^z S_{11}^{z,-1} S_{12}^z)^{-1} \end{bmatrix} \quad (3)$$

where $||S^z_{ij}S^{z,-1}_{jj}S^z_{ji}||_2 \leqslant ||S^z_{ij}||_2||S^{z,-1}_{jj}||_2||S^z_{ji}||_2 = O_p(\frac{r^{2\tau_1}}{T})$ for $1 \leqslant i \neq j \leqslant 2$ by the considerations above. We get

$$
\begin{aligned}
||S^{z,-1}_{11} - \Sigma^{-1}_{z1}||_F &= O_p(\frac{r}{\sqrt{T}}) \\
||S^{z,-1}_{12}||_2 &= O_p(\frac{r^{\tau_1}}{\sqrt{T}})
\end{aligned}
$$

where the first equation is analogous to (A.3) and the second follows from above. Then we get

$$
\begin{aligned}
&(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta Y_t Z'_{t-1}D^{-1}_{0T})(D^{-1}_{0T}Z_{-1}Z'_{-1}D^{-1}_{0T})^{-1}(\frac{1}{\sqrt{T}}D^{-1}_{0T}\sum_{t=1}^{T}Z_{t-1}\Delta Y'_t) \\
&= \frac{1}{T}\sum_{t=1}^{T}\Delta Y_t Z'_{1,t-1}S^{z,-1}_{11}\frac{1}{T}\sum_{t=1}^{T}Z_{1,t-1}\Delta Y'_t + \frac{1}{T^{3/2}}\sum_{t=1}^{T}\Delta Y_t Z'_{2,t-1}S^{z,-1}_{21}\frac{1}{T}\sum_{t=1}^{T}Z_{1,t-1}\Delta Y'_t \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\Delta Y_t Z'_{1,t-1}S^{z,-1}_{21}\frac{1}{T^{3/2}}\sum_{t=1}^{T}Z_{2,t-1}\Delta Y'_t + \frac{1}{T^{3/2}}\sum_{t=1}^{T}\Delta Y_t Z'_{2,t-1}S^{z,-1}_{22}\frac{1}{T^{3/2}}\sum_{t=1}^{T}Z_{2,t-1}\Delta Y'_t
\end{aligned}
$$

For the first term we get

$$
||\frac{1}{T}\sum_{t=1}^{T}\Delta Y_t Z'_{1,t-1}S^{z,-1}_{11}\frac{1}{T}\sum_{t=1}^{T}Z_{1,t-1}\Delta Y'_t - \Sigma_{\Delta y.z1}||_F = O_p(\frac{m}{\sqrt{T}})
$$

where $\Sigma_{\Delta y.z1} = \mathbf{E}(\Delta Y_t Z'_{1,t-1})\mathbf{E}(Z_{1,t-1}Z'_{1,t-1})^{-1}\mathbf{E}(Z_{1,t-1}\Delta Y'_t)$. For the second one, we have

$$
\frac{1}{T^{3/2}}\sum_{t=1}^{T}\Delta Y_t Z'_{2,t-1} = \frac{1}{T^{3/2}}\sum_{t=1}^{T}\left(\alpha Z_{1,t-1}Z'_{2,t-1} + B\Delta X_{t-1}Z'_{2,t-1} + w_t Z'_{2,t-1}\right),
$$

which implies that $||\frac{1}{T^{3/2}}\sum_{t=1}^{T}\Delta Y_t Z'_{2,t-1}||_2 = O_p(\frac{r^{\tau_1}}{\sqrt{T}})$, as well as $||\frac{1}{T^{3/2}}\sum_{t=1}^{T}Z_{1,t-1}Z'_{2,t-1}||_2 = O_p(\frac{r^{\tau_1}}{\sqrt{T}})$ due to (A.12) and Lemma 2 from the first part of the expression. The l_2 norms for the other two terms are negligible with faster rate $O_P(1/\sqrt{T})$ and Lemma 2. Therefore,

$$
\begin{aligned}
||\frac{1}{T^{3/2}}\sum_{t=1}^{T}\Delta Y_t Z'_{2,t-1}S^{z,-1}_{21}\frac{1}{T}\sum_{t=1}^{T}Z_{1,t-1}\Delta Y'_t||_F &\leqslant ||\frac{1}{T^{3/2}}\sum_{t=1}^{T}\Delta Y_t Z'_{2,t-1}||_2||S^{z,-1}_{21}||_2||\frac{1}{T}\sum_{t=1}^{T}Z_{1,t-1}\Delta Y'_t||_2\sqrt{r} \\
&= O_p(\frac{r^{2\tau_1+1/2}}{T}).
\end{aligned}
$$

Thus in total,

$$
||(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta Y_t Z'_{t-1}D^{-1}_{0T})(D^{-1}_{0T}Z_{-1}Z'_{-1}D^{-1}_{0T})^{-1}(\frac{1}{\sqrt{T}}D^{-1}_{0T}\sum_{t=1}^{T}Z_{t-1}\Delta Y'_t) - \Sigma_{\Delta y.z1}||_F = O_p(\frac{m}{\sqrt{T}})
$$

Therefore, we conclude that $||\frac{1}{T}\sum_{t=1}^{T}\Delta\breve{Y}_t\Delta\breve{Y}'_t - \Sigma_{\Delta y.z1}||_F = O_p(\frac{m}{\sqrt{T}})$. In the same way, the results for $\Delta\breve{X}_{t-1}$ are obtained. The performance of $\sum_{t=1}^{T}\breve{w}_t\breve{w}'_t$ can be directly inferred from Lemma 2. □

2 Proof of Theorem 3.2

Proof. With the Lasso-estimator $\hat{B}$ from (12) for $(B_1, B_2, \ldots, B_P)$ define $\delta_B = \hat{B} - B$. Then we get for the first part of the Lasso criterion function (12) that

$$
\begin{aligned}
& \sum_{t=1}^{T} ||\Delta\breve{Y}_t - \sum_{j=1}^{P} B_j \Delta\breve{Y}_{t-j}||^2 \\
= & \sum_{t=1}^{T} (B\Delta\breve{X}_{t-1} + \breve{w}_t - \hat{B}\Delta\breve{X}_{t-1})'(B\Delta\breve{X}_{t-1} + \breve{w}_t - \hat{B}\Delta\breve{X}_{t-1}) \\
= & \sum_{t=1}^{T} (\breve{w}_t - \delta_B \breve{X}_{t-1})'(\breve{w}_t - \delta_B \breve{X}_{t-1}) \\
= & \sum_{t=1}^{T} \breve{w}_t'\breve{w}_t - 2\frac{1}{\sqrt{T}} w_t'(\breve{X}_{t-1}' \otimes I_m) vec(\sqrt{T}\delta_B) + \frac{1}{T} vec(\sqrt{T}\delta_B)'(\breve{X}_{t-1}\breve{X}_{t-1}' \otimes I_m) vec(\sqrt{T}\delta_B) \, .
\end{aligned}
$$

Taking first order conditions w.r.t. $vec(\sqrt{T}\delta_B)$ yields

$$
\sum_{t=1}^{T} -\frac{2}{\sqrt{T}} (\Delta\breve{X}_{t-1} \otimes I_m) w_t' + \frac{2}{T} (\Delta\breve{X}_{t-1}\Delta\breve{X}_{t-1}' \otimes I_m) vec(\sqrt{T}\delta_B)
$$

For the rest of the proof, we assume for ease of notation that all $B_1, B_2, \ldots, B_p$ are non-zero not just B_p as Assumption 2.3 implies. Denote by δ_B^0 the first mp columns of δ_B. Consistent lag selection requires that each $m \times m$ block in δ_B^0 contains a non-zero element but the last $m(P-p)$ columns are zero, which can be ensured by the following KKT conditions:

$$
\begin{aligned}
\sum_{t=1}^{T} -\frac{1}{\sqrt{T}} w_t \Delta\breve{X}_{t-1}^{0\prime} + \frac{1}{T}(\sqrt{T}\delta_B^0)(\Delta\breve{X}_{t-1}^0 \Delta\breve{X}_{t-1}^{0\prime}) &= -\frac{\lambda_T^{lag}}{2\sqrt{T}} \Big[\frac{B_1}{||B_1||_F ||vec(\hat{B}_1)||_\infty^\gamma}, \\
& \quad \ldots, \frac{B_p}{||B_p||_F ||vec(\hat{B}_p)||_\infty^\gamma} \Big] \qquad (4) \\
|| \sum_{t=1}^{T} -\frac{1}{\sqrt{T}} w_t \Delta\breve{X}_{t-1}^{k\prime} + \frac{1}{T}(\sqrt{T}\delta_B^0)(\Delta\breve{X}_{t-1}^0 \Delta\breve{X}_{t-1}^{k\prime}) ||_F &< \frac{\lambda_T^{lag}}{2\sqrt{T}} ||vec(\hat{B}_k)||_\infty^{-\gamma} \qquad (5)
\end{aligned}
$$

for $k = p+1, \ldots, P$. $\Delta\breve{X}_{t-1}^k$ denotes the k-th m-dimensional block in $\Delta\breve{X}_{t-1}$ and $\Delta\breve{X}_{t-1}^0$ for the first p blocks. In the same way as for rank selection, define

$$
\begin{aligned}
V_{B_0} &= \Big[\frac{B_1}{||B_1||_F ||vec(\hat{B}_1)||_\infty^\gamma}, \ldots, \frac{B_p}{||B_p||_F ||vec(\hat{B}_p)||_\infty^\gamma} \Big] \\
S_{wx0} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} w_t \Delta\breve{X}_{t-1}^{0\prime} \qquad S_{x0} = \frac{1}{T} \sum_{t=1}^{T} \Delta\breve{X}_{t-1}^0 \Delta\breve{X}_{t-1}^{0\prime} \\
S_{wx1} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} w_t \Delta\breve{X}_{t-1}^{k\prime} \qquad S_{x0x1} = \frac{1}{T} \sum_{t=1}^{T} \Delta\breve{X}_{t-1}^0 \Delta\breve{X}_{t-1}^{k\prime}
\end{aligned}
$$

Then from the first KKT condition (4) we get

$$\sqrt{T}\delta_B^0 = S_{wx0}S_{x0}^{-1} - \frac{\lambda_T^{lag}}{2\sqrt{T}}V_{B_0}S_{x0}^{-1} \ . \tag{6}$$

With Theorem 3.1 for V_{B_0} and Lemma 3 for $||S_{x0}^{-1}||_2 = O_P(1)$ it holds that

$$\frac{\lambda_T^{lag}}{\sqrt{T}}||V_{B_0}S_{x0}^{-1}||_F \leqslant \frac{\lambda_T^{lag}}{\sqrt{T}}||V_{B_0}||_F||S_{x0}^{-1}||_2 = O_p(\frac{\lambda_T^{lag}}{\sqrt{T}})$$

where the assumption on λ_T^{lag}, implies that $O_p(\frac{\lambda_T^{lag}}{\sqrt{T}}) = o_P(1)$. Thus on the true active set in the lags, the effect of the penalty vanishes.

From the second part of the KKT conditions (5), we obtain by plugging in (6) the following sufficient condition for exclusion of lags larger than p, i.e. of lags which are not in the true active set

$$||vec(-S_{wx1} + S_{wx0}S_{x0}^{-1}S_{x0x1})_j||_2 < \frac{\lambda_T^{lag}}{2\sqrt{T}}\Big(||vec(\hat{B}_{p+j})||_\infty^{-\gamma} - ||(V_{B_0}S_{x0}^{-1}S_{x0x1})_j||_F\Big) \tag{7}$$

where the subscript j of a vector denotes the j-th m^2 block. With Theorem 3.1 for V_{B_0} together with Lemma 3 for stationary $||S_{x0x1}||_2 = O_P(1)$ we find

$$||(V_{B_0}S_{x0}^{-1}S_{x0x1})_j||_F \leqslant ||V_{B_0}S_{x0}^{-1}S_{x0x1}||_F \leqslant ||V_{B_0}||_F||S_{x0}^{-1}||_2||S_{x0x1}||_2 = O_p(1)$$

and similarly on the LHS that $||(S_{wx0}S_{x0}^{-1}S_{x0x1})_j||_F = O_P(1/\sqrt{T})$. So both terms are negligible in (7). We use that by Theorem 3.1 it holds that $||vec(\hat{B}_{p+j})||_\infty^{-\gamma} = O_p\Big((\sqrt{\frac{\log m}{T}})^{-\gamma}\Big)$. Then by (7) and setting $N_k = vec(-S_{wx1})_k$, this yields

$$\begin{aligned}
\mathbb{P}\Big(\sqrt{\sum_{k=1}^{m^2} N_k^2} > \frac{\lambda_T^{lag}}{2\sqrt{T}}(\frac{\sqrt{\log m}}{\sqrt{T}})^{-\gamma}\Big) &= \mathbb{P}\Big(\sum_{k=1}^{m^2} N_k^2 > (\frac{\lambda_T^{lag}}{2\sqrt{T}}(\frac{\sqrt{\log m}}{\sqrt{T}})^{-\gamma})^2\Big) \\
&\leqslant \sum_{k=1}^{m^2}\mathbb{P}\Big(|N_k| > \frac{\lambda_T^{lag}}{2m\sqrt{T}}(\frac{\sqrt{\log m}}{\sqrt{T}})^{-\gamma}\Big) \\
&\leqslant \Big(\frac{m^2(\log m)^{\gamma/2}C_1}{\lambda_T^{lag}T^{1/2(\gamma-1)}}\Big)^2 \text{ for } 0 < C_1 < \infty
\end{aligned}$$

where Chebyshev's inequality was applied in the last line. For the second moment bound C_1, we use that due to Assumption 2.1. The bound then follows from Lemma 3. Hence for P fixed and due to $\frac{\lambda_T^{lag}T^{1/2(\gamma-1)}}{m^2(\log m)^{\gamma/2}} \to \infty$, the last line implies that with probability tending to one, irrelevant lags are excluded by the proposed Lasso procedure. □

3 Proof of Theorem 4.1

Proof. We first show that $\mathbf{E}(|u_t^k|^{4+\delta})$ is bounded for all $k = 1, \ldots, m$.

Define $\widetilde{A}_l = A_l \Sigma_w^{1/2}$ with A_l from Assumption 4.1. Then $\sum_{l=1}^{\infty} j||\widetilde{A}_l||_F < \infty$. Denote $\tilde{a}_{l,kj}$ as the k,j-th element in $\widetilde{A}_l$. Not that the assumption $\sum_{j=1}^{\infty} j||\tilde{A}_j||_F < \infty$ implies that $\sum_{j=1}^{\infty} ||\tilde{A}_j||_2 < \infty$. Thus for every $\varepsilon > 0$ close enough to zero, there exists an N such that for all $n > N$, $||\tilde{A}_n||_2 < \varepsilon$. Therefore, for all $1 < \zeta < \infty$, we have

$$\sum_{j=1}^{N} ||\tilde{A}_j||_2^{\zeta} + \sum_{j=N+1}^{\infty} ||\tilde{A}_j||_2^{\zeta} < \sum_{j=1}^{N} ||\tilde{A}_j||_2^{\zeta} + \sum_{j=N+1}^{\infty} \varepsilon^{\zeta} < \infty$$

We use this to bound the $4+\delta$-th moment of u_t^k for $k = 1, \dots, m$ split up as follows

$$u_t^k = \sum_{j=1}^{m} \tilde{a}_{0,kj} e_{t,j} + \sum_{l=1}^{\infty} \sum_{j=1}^{m} \tilde{a}_{l,kj} e_{t-l,j}$$

Define the sequence $X_l = \sum_{j=1}^{m} \tilde{a}_{l,kj} e_{t-l,j}$, then applying Rosenthal's inequality yields for the fourth moment of X_l

$$\begin{aligned} E(|X_l|^{4+\delta}) &\leqslant C_X \Big(\sum_{j=1}^{m} |\tilde{a}_{l,kj}|^{4+\delta} E(|e_{t-l,j}|^{4+\delta}) + (\sum_{j=1}^{m} \tilde{a}_{l,kj}^2 E(e_{t-l,j}^2))^{2+\delta/2} \Big) \\ &= O_{a.s.} \Big(\sum_{j=1}^{m} |\tilde{a}_{l,kj}|^{4+\delta} + (\sum_{j=1}^{m} \tilde{a}_{l,kj}^2)^{2+\delta/2} \Big) \\ &= O_{a.s.} \Big(||\tilde{A}_l||_2^{4+\delta} + ||\tilde{A}_l||_2^{4+\delta} \Big) = O_{a.s.}(1). \end{aligned}$$

With this, we get for the partial sum $\sum_{l=0}^{L} X_l$ that

$$\begin{aligned} E(|\sum_{l=0}^{L} X_l|^{4+\delta}) &\leqslant C_X^L \Big(\sum_{l=0}^{L} \sum_{j=1}^{m} |\tilde{a}_{l,kj}|^{4+\delta} E(|e_{t-l,j}|^{4+\delta}) + (\sum_{l=0}^{L} \sum_{j=1}^{m} \tilde{a}_{l,kj}^2 E(e_{t-l,j}^2))^{2+\delta/2} \Big) \\ &= O_{a.s.} \Big(\sum_{l=0}^{L} \sum_{j=1}^{m} |\tilde{a}_{l,kj}|^{4+\delta} + (\sum_{l=0}^{L} \sum_{j=1}^{m} \tilde{a}_{l,kj}^2)^{2+\delta/2} \Big) \\ &= O_{a.s.} \Big(\sum_{l=0}^{L} ||\tilde{A}_l||_2^{4+\delta} + (\sum_{l=0}^{L} ||\tilde{A}_l||_2^2)^{2+\delta/2} \Big) = O_{a.s.}(1) \end{aligned}$$

For the $L_{4+\delta}$-convergence of $\sum_{l=0}^{L} X_l$, only remains to show that the partial sum $\sum_{l=0}^{L} X_l$ is an $L_{4+\delta}$-Cauchy sequence. Define $\xi_j = \sum_{l=0}^{j} X_l$, then for $i < j$, as i goes to infinity,

$$\begin{aligned} E(|\xi_i - \xi_j|^{4+\delta}) &= E(|\sum_{l=i+1}^{j} X_l|^{4+\delta}) \\ &\leqslant C_\xi \Big(\sum_{l=i+1}^{j} E|X_l|^{4+\delta} + (\sum_{l=i+1}^{j} E(X_l^2))^{2+\delta/2} \Big) \\ &= O_{a.s.} \Big(\sum_{l=i+1}^{\infty} ||\tilde{A}_j||_2^{4+\delta} + (\sum_{l=i+1}^{\infty} ||\tilde{A}_l||_2^2)^{2+\delta/2} \Big) = o_{a.s.}(1) \end{aligned}$$

Therefore, ξ_j constitutes an $L_{4+\delta}$-Cauchy sequence and thus ξ_j is $L_{4+\delta}$ convergent. Therefore with dominated convergence,

$$\begin{aligned} E(|u_t^k|^{4+\delta}) &= \lim_{L\to\infty} E(|\sum_{l=0}^{L} X_l + \sum_{l=L+1}^{\infty} X_l|^{4+\delta}) \\ &\leqslant \lim_{L\to\infty} C\left(E(|\sum_{l=0}^{L} X_l|^{4+\delta}) + E(|\sum_{l=L+1}^{\infty} X_l|^{4+\delta})\right) < \infty \end{aligned}$$

which is the first claim of the theorem.

If the iid innovation e_0 in w_t is replaced by an i.i.d. copy $\dot{e}_0$, its impact at time t is $A_t(w_0 - \dot{w}_0) = A_t\Sigma_w^{1/2}(e_0 - \dot{e}_0) = \widetilde{A}_t(e_0 - \dot{e}_0)$. Denote the kth row of $\widetilde{A}_t$ by $\widetilde{a}_{tk}$, then it holds for $\hat{e}_0 = e_0 - e_0'$ that

$$\begin{aligned} \left(E(|\widetilde{a}_{tk}\hat{e}_0|^4)\right)^{1/4} &\leqslant \left(C_4 E(\sum_{j=1}^{m} |\widetilde{a}_{tk,j}\hat{e}_{0j}|^2)^{4/2}\right)^{1/4} \\ &\leqslant C(\sum_{j=1}^{m} \widetilde{a}_{tk,j}^2)^{1/2} = ||A_t\Sigma_w^{1/2}||_F \leqslant ||A_t||_F||\Sigma_w^{1/2}||_2 \end{aligned} \tag{8}$$

by Marcinkiewicz-Zygmund inequality since each element in e_t has bounded 4-th moment. Then according to Subsection 3.1 in Wu (2007), (8) bounds the physical dependence measure γ_{tk} of Chen, Xu and Wu (2013) elementwise. Thus we get for each element k by Assumption 4.1 that

$$\sum_{t=0}^{\infty} t\gamma_{tk} \leqslant ||\Sigma_w^{1/2}||_2 \sum_{t=0}^{\infty} t||A_t||_F < \infty$$

which is the sufficient condition for the elementwise strong invariance principle in Corollary 4 in Wu (2007). This implies the claim of the theorem. Moreover, the covariance matrix of $\mathbf{M}(s)$ is obtained as $\sum_{j=0}^{\infty} A_j\Sigma_w A_j'$ by elementary calculations.

□

4 Proof for Theorem 4.2

Proof. To derive the results in Theorem 4.2, we first show the following block-wise convergence results as in Theorem 2.1:

$$||\frac{1}{T}\sum_{t=1}^{T} \Delta Z_t Z_{t-1}' - \begin{bmatrix} (\beta'\alpha)\Sigma_{z1} + \Gamma_{v1z1}^1 & -(\beta'\alpha + I_r)\Gamma_{v2z1}^{1\prime} - \Sigma_{v1v2} \\ \Gamma_{v2z1}^1 & \int_0^1 d\mathbf{M}_2(s)\mathbf{M}_2(s)' + \Gamma_{22}^0 \end{bmatrix}||_F = O_P(a_n)$$

$$||D_T^{-1}\frac{1}{T}\sum_{t=1}^{T} Z_{t-1}Z_{t-1}' - \begin{bmatrix} \Sigma_{z1} & -(\beta'\alpha)^{-1}\big((\beta'\alpha + I_r)\Gamma_{v2z1}^{1,\prime} + \Sigma_{v1v2} + \Gamma_{12}^0 + \int_0^1 d\mathbf{M}_1(s)\mathbf{M}_2'(s)\big) \\ 0 & \int_0^1 \mathbf{M}_2(s)\mathbf{M}_2'(s)ds \end{bmatrix}||_F = O_P(a_n)$$

with $a_n = \sqrt{\frac{r^2}{T}} + \sqrt{\frac{mr}{T}} + \sqrt{\frac{m^2(\log T)^{3/2}(\log\log T)}{T^{1/2}}}$.

In a similar way as in the proof for Theorem 2.1, we proceed with the eight blocks and highlight the differences.

1.+2. purely stationary blocks $\bar{b}_{11} = \frac{1}{T}\Delta Z_1 M Z'_{1,-1}$ *and* $\bar{\chi}_{11} = \frac{1}{T} Z_{1,-1} M Z'_{1,-1}$
For the second block, it follows from Lemma 2 that

$$||\frac{1}{T}\sum_{t=1}^{T} Z_{1,t-1}Z'_{1,t-1} - \Sigma_{z1}||_F = O_p(\frac{r}{\sqrt{T}}) \ . \tag{9}$$

For the first term we get from (15)

$$\frac{1}{T}\sum_{t=1}^{T}\Delta Z_{1,t}Z'_{1,t-1} = (\beta'\alpha)\frac{1}{T}\sum_{t=1}^{T} Z_{1,t-1}Z'_{1,t-1} + \frac{1}{T}\sum_{t=1}^{T} v_{1,t}Z'_{1,t-1}$$

which implies with (9) that

$$||\frac{1}{T}\sum_{t=1}^{T}\Delta Z_{1,t}Z_{1,t-1} - (\beta'\alpha)\Sigma_{z1} - \Gamma^1_{v1z1}||_F = O_p(\frac{r}{\sqrt{T}}) \tag{10}$$

3. mixed stationary/nonstationary block $\bar{b}_{12} = \frac{1}{T}\Delta Z_1 Z'_{2,-1}$
From (15) we have that $Z_{2,t} = \sum_{s=1}^{t} v_{2,s}$ which yields

$$\begin{aligned}
\frac{1}{T}\sum_{t=0}^{T}\Delta Z_{1,t}Z'_{2,t-1} &= -\frac{1}{T}\sum_{t=1}^{T-1} Z_{1,t}\Delta Z'_{2,t} + \frac{1}{T}(Z_{1,T}Z'_{2,T} - Z_{1,0}Z'_{2,0}) \\
&= -\frac{1}{T}(\beta'\alpha + I_r)\sum_{t=1}^{T} Z_{1,t-1}v'_{2,t} - \frac{1}{T}\sum_{t=1}^{T} v_{1,t}v'_{2,t} + R_8
\end{aligned}$$

with $||R_8||_F = O_P(\sqrt{mr}/\sqrt{T})$. Hence

$$||\frac{1}{T}\sum_{t=0}^{T}\Delta Z_{1,t}Z'_{2,t-1} + (\beta'\alpha + I_r)\Gamma^{1\prime}_{v2z1} + \Sigma_{v1v2}||_F = O_p(\frac{\sqrt{mr}}{\sqrt{T}}) \tag{11}$$

4. mixed stationary/nonstationary block $\bar{b}_{12} = \frac{1}{T}\Delta Z_2 Z'_{1,-1}$
With $Z_{2,t} = \sum_{s=1}^{t} v_{2,s}$ from (15) we get that $\frac{1}{T}\sum_{t=1}^{T}\Delta Z_{2,t}Z'_{1,t-1} = \frac{1}{T}\sum_{t=1}^{T} v_{2,t}Z'_{1,t-1}$ which leads to

$$||\frac{1}{T}\sum_{t=1}^{T}\Delta Z_{2,t}Z'_{1,t-1} - \Gamma^1_{v2z1}||_F = O_p(\frac{\sqrt{mr}}{\sqrt{T}}) \tag{12}$$

5. purely nonstationary block $\bar{b}_{22} = \frac{1}{T}\Delta Z_2 Z'_{2,-1}$
Different from the block b_{22} in the proof for Theorem 2.1,the increment of $Z_{2,t}$ is no longer independent of $\mathcal{F}_{t-1}$ due to the weak dependence of $v_t = Qu_t$. Therefore, the standard discrete approximation of the stochastic integral (see, e.g. Section 2.5 of Chung and Williams (1990)) can not be directly applied.

With $Z_{2,t} = \sum_{s=1}^{t} v_{2,s}$ from (15) we get that

$$\frac{1}{T}\sum_{t=1}^{T} \Delta Z_{2,t} Z'_{2,t-1} = \frac{1}{T}\sum_{t=1}^{T} v_{2,t} Z'_{2,t-1} = \frac{1}{T}\sum_{t=1}^{T} v_{2,t} \sum_{s=0}^{t-1} v'_{2,s}$$

Define $\Upsilon_t = \sum_{s=0}^{t} u_s$, due to the assumption that $\sum_{j=1}^{\infty} j||A_j||_F < \infty$, for some $K > 0$, it holds that

$$\begin{aligned} u_t \Upsilon'_{t-1} = w_t \Upsilon'_{t-1} &+ A_1(w_{t-1}u'_{t-1} + w_{t-1}\Upsilon'_{t-2}) \\ &+ A_2\Big(w_{t-2}(u'_{t-1} + u'_{t-2}) + w_{t-2}\Upsilon'_{t-3}\Big) \\ &+ \dots \\ &+ A_K\Big(w_{t-K}(\sum_{j=1}^{K} u'_{t-j}) + w_{t-K}\Upsilon'_{t-K-1}\Big) + o(1) \end{aligned}$$

By summing up $u_t\Upsilon'_{t-1}$ over t and dividing the sum by T, the term $\frac{1}{T}\sum_{t=1}^{T} A_k w_{t-k}(\sum_{j=1}^{k} u'_{t-j})$ (for $1 \leqslant k \leqslant K$) satisfies

$$||\frac{1}{T}\sum_{t=1}^{T} A_k w_{t-k}(\sum_{j=1}^{k} u'_{t-j}) - A_k\Sigma_w(\sum_{j=1}^{k} A'_{k-j})||_F = O_p(||A_k||_F \frac{m}{\sqrt{T}})$$

We leave the $||A_k||_F$ in the convergence rate so that the sequence still converge at the rate of $\frac{m}{\sqrt{T}}$ after summing over k.

Sum up $A_k\Sigma_w(\sum_{j=1}^{k} A'_{k-j})$ over $k = 1, \dots, K$, i.e.,

$$\begin{aligned} \sum_{k=1}^{K} A_k\Sigma_w(\sum_{j=1}^{k} A'_{k-j}) &= A_1\Sigma_w A'_0 \\ &+ A_2\Sigma_w A'_1 + A_2\Sigma_w A'_0 \\ &+ \dots \\ &+ A_K\Sigma_w A'_{K-1} + A_K\Sigma_w A'_{K-2} + \dots + A_K\Sigma_w A'_0 \\ &\rightarrow \Gamma_u(1) + \Gamma_u(2) + \dots + \Gamma_u(K) \end{aligned}$$

where $\Gamma_u(k) = \mathbf{E}(u_t u'_{t-k})$. The left terms can be summed up over t and expressed as

$$\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{j=0}^{K} A_j w_{t-j}\Upsilon'_{t-j-1}\Big) = \sum_{j=0}^{K} A_j\Big(\frac{1}{T}\sum_{t=1}^{T} w_{t-j}\Upsilon'_{t-j-1}\Big)$$

By the same argument as in proof for Theorem 2.1, we conclude that

$$||\frac{1}{T}\sum_{t=1}^{T} w_{t-j}\Upsilon'_{t-j-1} - \int_0^1 d\mathbf{M}_w\mathbf{M}'||_F = O_p(\frac{m(\log T)^{3/4}(\log\log T)^{1/2}}{T^{1/4}})$$

where $\mathbf{M}_w$ denotes the m-dimensional Brownian motion with the same covarianc matrix as w_t. According to Theorem 4.1, we can conclude that

$$||\frac{1}{T}\sum_{t=1}^{T} u_t\Upsilon'_{t-1} - \int_0^1 d\mathbf{M}\mathbf{M}' - \sum_{k=1}^{\infty}\Gamma_u(k)||_F = O_p(\frac{m(\log T)^{3/4}(\log\log T)^{1/2}}{T^{1/4}}) \qquad (13)$$

The convergence rate of those terms to $\sum_{k=1}^{\infty}\Gamma_u(k)$ is $\frac{m}{\sqrt{T}}$, dominated by the rate of strong invariance principle and thus ignored here.

The desired result can be achieved by pre- and post-multiplying $u_t\Upsilon'_{t-1}$ by β' or $\alpha'_{\perp}$.

6. mixed stationary/nonstationary block $\bar{\chi}_{12} = \frac{1}{T}Z_{1,-1}Z'_{2,-1}$

From (15) we get that

$$\frac{1}{T}\sum_{t=1}^{T}\Delta Z_{1,t}Z'_{2,t-1} = \frac{1}{T}\sum_{t=1}^{T}(\beta'\alpha)Z_{1,t-1}Z'_{2,t-1} + \frac{1}{T}\sum_{t=1}^{T}v_{1,t}Z'_{2,t-1}$$

which we rearrange as

$$\frac{1}{T}\sum_{t=1}^{T}Z_{1,t-1}Z'_{2,t-1} = (\beta'\alpha)^{-1}(\frac{1}{T}\sum_{t=1}^{T}\Delta Z_{1,t}Z'_{2,t-1} - \frac{1}{T}\sum_{t=1}^{T}v_{1,t}Z'_{2,t-1}) \ .$$

Using (11) for the first term on the right, the strong invariance principle of Theorem 4.1 for the second term as above and $||(\beta'\alpha)^{-1}||_2 = O(r^{\tau_1})$ by Assumption 4.2 it holds that

$$\begin{aligned} &||\frac{1}{T}\sum_{t=1}^{T}Z_{1,t-1}Z'_{2,t-1} + (\beta'\alpha)^{-1}\Big((\beta'\alpha + I_r)\Gamma^{1\prime}_{v2z1} + \Sigma_{v1v2} + \Gamma^0_{12} + \int_0^1 d\mathbf{M}_1\mathbf{M}'_2\Big)||_F \\ = \; & O\Big(r^{\tau_1}\sqrt{\frac{mr}{T} + \frac{mr(\log T)^{3/2}(\log\log T)}{\sqrt{T}}}\Big) \end{aligned} \tag{14}$$

7. mixed stationary/nonstationary block $\bar{\chi}_{21} = \frac{1}{T}\left(\frac{1}{T}Z_{2,-1}Z'_{1,-1}\right)$

By similar argument as in the independent case for block χ_{21}, it is sufficient to work here withe conservative upper bound $O_P(r^{\tau_1})$ from (14) for each element in the inner bracket. We thus obtain

$$||\frac{1}{T^2}\sum_{t=1}^{T}Z_{2,t-1}Z'_{1,t-1}||_F = O_p(\frac{\sqrt{mr}r^{\tau_1}}{T}) \tag{15}$$

8. purely nonstationary block $\bar{\chi}_{22} = \frac{1}{T}\left(\frac{1}{T}Z_{2,-1}Z'_{2,-1}\right)$

As before, we show the distance between $\frac{1}{T^2}\sum_{t=1}^{T}Z_{2,t-1}Z'_{2,t-1}$ and $\int_0^1 d\mathbf{M}_2\mathbf{M}_2$. Element-wise, we have

$$\begin{aligned} &\frac{1}{T^2}\sum_{t=1}^{T}Z^i_{2,t-1}Z^j_{2,t-1} - \int_0^1 \mathbf{M}_{2,i}(s)\mathbf{M}_{2,j}(s)ds \\ = \; & \sum_{t=1}^{T}\int_{\frac{t-1}{T}}^{\frac{t}{T}}(\frac{1}{\sqrt{T}}Z^i_{2,t-1})\Big(\frac{1}{\sqrt{T}}Z^j_{2,t-1} - \mathbf{M}_{2,j}(s)\Big)ds \\ + \; & \sum_{t=1}^{T}\int_{\frac{t-1}{T}}^{\frac{t}{T}}\Big(\frac{1}{\sqrt{T}}Z^i_{2,t-1} - \mathbf{M}_{2,i}(s)\Big)(\mathbf{M}_{2,j}(s))ds \end{aligned}$$

We have shown that $|\frac{1}{\sqrt{T}}Z^i_{2,t-1} - \mathbf{M}_{2,i}(\frac{t-1}{T})| = O_p(\frac{(\log T)^{3/4}(\log\log T)^{1/2}}{T^{1/4}})$ and for any Brownian motion element in $\mathbf{M}(s)$, $\max_{\frac{t-1}{T}\leqslant s\leqslant \frac{t}{T}} |\mathbf{M}_j(s) - \mathbf{M}_j(\frac{t-1}{T})| = O_p(\sqrt{\frac{\log T}{T}})$. Thus in total, we get

$$||\frac{1}{T^2}Z_{2,t-1}Z'_{2,t-1} - \int_0^1 \mathbf{M}_2\mathbf{M}'_2||_F \quad = \quad O_p(m\frac{(\log T)^{3/4}(\log\log T)^{1/2}}{T^{1/4}}) \tag{16}$$

Now the first part of the initial claim follows from (10)-(13) and the second part from (9) and (14)-(16). In the same manner as the proof of Theorem 2.1, we can define $\bar{\chi}$ composed of the blocks $\bar{\chi}_{11} - \bar{\chi}_{22}$ and $\xi = \bar{\chi}^{-1}$. Then the final result for $\widetilde{\Psi}$ follows from direct calculations. □

5 Proof of Theorem 4.3

Proof. Define $\beta_0 = \begin{bmatrix} \beta' \\ \beta'_\perp \end{bmatrix}$. We thus obtain for $\beta_0\widetilde{\Pi}'$

$$\begin{aligned} \begin{pmatrix} \beta'\widetilde{\Pi}' \\ \beta'_\perp\widetilde{\Pi}' \end{pmatrix} &= \begin{pmatrix} I_r & \frac{1}{T}\beta'\alpha_\perp \\ 0 & \frac{1}{T}\beta'_\perp\alpha_\perp \end{pmatrix} \left(Q^{-1}\widetilde{\Psi}\right)' \\ &= \begin{pmatrix} I_r & \frac{1}{T}\beta'\alpha_\perp \\ 0 & \frac{1}{T}\beta'_\perp\alpha_\perp \end{pmatrix} \begin{pmatrix} \alpha(\beta'\alpha)^{-1}\widetilde{\Psi}_{11} + \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\widetilde{\Psi}_{21} & \alpha(\beta'\alpha)^{-1}\widetilde{\Psi}_{12} + \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\widetilde{\Psi}_{22} \end{pmatrix}' \end{aligned} \tag{17}$$

From (17) and Theorem 4.2, we get

$$\begin{aligned} ||\beta'\widetilde{\Pi}' - \alpha'_\star||_F &= O_p(\sqrt{\frac{mr}{T}}) \\ ||\beta'_\perp\widetilde{\Pi}'||_2 &= O_p(\frac{r^{\tau_1}}{T}) \end{aligned}$$

The l_2 norms of $\widetilde{\Psi}_{21}$ and $\widetilde{\Psi}_{22}$ may increase with $r^{\tau_1}\sqrt{m}$, which slows down the converging rate of the irrelevant basis.

Due to the unitary invariant property of singular values, we have

$$\sigma_j(\beta_0\widetilde{\Pi}') = \sigma_j(S\widetilde{\Pi}') = \sigma_j(\widetilde{R})$$

which implies that

$$|\sigma_j(\widetilde{R}) - \sigma_j(\alpha_\star)| = O_p(\sqrt{\frac{mr}{T}}) \quad \text{for } j = 1, \dots, r \tag{18}$$

by matrix perturbation theory (Mirsky version, Theorem 4.11 of Stewart and Sun (1990)) and (**??**). The column-pivoting step in QR decomposition makes the $\widetilde{R}_{11}$ a well-conditioned matrix, thus the largest r singular values in $\widetilde{R}$ are contributed by $\widetilde{R}_1$, the first r-columns. Besides, the upper-triangular structure of $\widetilde{R}$ excludes linear dependence between any two rows. Therefore, we can conclude that

$$\sigma_r(\widetilde{R}) \leqslant \sqrt{\sum_{j=k}^{m} \widetilde{R}(k,j)^2} \leqslant \sigma_1(\widetilde{R}) \quad \text{for } k = 1, \dots, r \tag{19}$$

The matrix perturbation theory result (18) provides further bounds for l_2 norm of each row in $\widetilde{R}_1$, i.e.,

$$\sigma_r(\widetilde{R}) \geqslant \sigma_r(\alpha_\star) - O_p(\sqrt{\frac{mr}{T}})$$
$$\sigma_1(\widetilde{R}) \leqslant \sigma_1(\alpha_\star) + O_p(\sqrt{\frac{mr}{T}})$$

Also by the upper-triangular structure and column-pivoting, we can derive that

$$\sqrt{\sum_{j=k}^{m} \widetilde{R}(k,j)^2} = O_p(\frac{r^{\tau_1}}{T}) \quad \text{for } k = r+1, \ldots, m \tag{20}$$

Moreover, (20) leads to the conclusion that

$$||\widetilde{R}_{22}||_F = O_p(\frac{\sqrt{m}}{T} r^{\tau_1}) \tag{21}$$

The difference between $\widetilde{R}_1$ and $\widetilde{R}$ is $\widetilde{R}_{22}$. Therefore, we can also conclude

$$|\sigma_j(\widetilde{R}_1) - \sigma_j(\widetilde{R})| = O_p(\frac{\sqrt{m} r^{\tau_1}}{T}), \qquad j = 1, .., r$$

and thus

$$|\sigma_j(\widetilde{R}_1) - \sigma_j(\alpha_\star)| = O_p(\sqrt{\frac{mr}{T}}) \quad \text{for } j = 1, \ldots, r \tag{22}$$

(17) can be further written as

$$\begin{pmatrix} \beta' \widetilde{\Pi}' \\ \beta_\perp' \widetilde{\Pi}' \end{pmatrix} = \begin{pmatrix} \beta' \widetilde{S}_1 \widetilde{R}_{11} & \beta' \widetilde{S}_1 \widetilde{R}_{12} + \beta' \widetilde{S}_2 \widetilde{R}_{22} \\ \beta_\perp' \widetilde{S}_1 \widetilde{R}_{11} & \beta_\perp' \widetilde{S}_1 \widetilde{R}_{12} + \beta_\perp' \widetilde{S}_2 \widetilde{R}_{22} \end{pmatrix} \tag{23}$$

with the after QR-decomposition components.

By equating the (17) and (23) we also have

$$\begin{pmatrix} \beta_\perp' \widetilde{S}_1 \widetilde{R}_{11} & \beta_\perp' \widetilde{S}_1 \widetilde{R}_{12} + \beta_\perp' \widetilde{S}_2 \widetilde{R}_{22} \end{pmatrix} = \frac{1}{T} (\beta_\perp' \alpha_\perp) \Big(\alpha (\beta' \alpha)^{-1} \widetilde{\Psi}_{12} + \beta_\perp (\alpha_\perp' \beta_\perp)^{-1} \widetilde{\Psi}_{22} \Big)'$$

which is equivalent to

$$\begin{aligned} \beta_\perp' \widetilde{S}_1 &= -\Big[\, 0 \quad \beta_\perp' \widetilde{S}_2 \widetilde{R}_{22} \, \Big] \widetilde{R}_1' (\widetilde{R}_1 \widetilde{R}_1')^{-1} \\ &+ \frac{1}{T} (\beta_\perp' \alpha_\perp) \Big(\alpha (\beta' \alpha)^{-1} \widetilde{\Psi}_{12} + \beta_\perp (\alpha_\perp' \beta_\perp)^{-1} \widetilde{\Psi}_{22} \Big)' \widetilde{R}_1' (\widetilde{R}_1 \widetilde{R}_1')^{-1} \end{aligned}$$

The singular values of $\widetilde{R}_1$ can be approximated by those of α. Therefore we conclude that $||\beta_\perp' \widetilde{S}_1||_F = O_p(\frac{m r^{\tau_1 + 2\tau_2}}{T})$.

□

6 Proof of Theorem 4.4

Proof. Denote $\widetilde{S}'Y_{t-1} = \begin{bmatrix} \breve{Z}_{1,t-1} \\ \breve{Z}_{2,t-1} \end{bmatrix}$ where $\breve{Z}_{1,t-1}$ is the projection of Y_{t-1} onto the subspace generated by $\widetilde{S}_1$. Because the distance between $\widetilde{S}_1$ and β converges at the rate of T, faster than other error terms mentioned above, we use $Z_{1,t-1}$ instead of $\breve{Z}_{1,t-1}$ in this proof. While both $\breve{Z}_{2,t-1}$ and $Z_{2,t-1}$ are non-stationary process, we can also use $Z_{2,t-1}$ instead of $\breve{Z}_{2,t-1}$ to keep the proof easier to read. Besides, we do not distinguish different matrix representations of $\hat{\alpha}$.
Define $\alpha_0 = [\alpha, 0], \hat{\alpha} = \alpha + \Sigma_{uz1}\Sigma_{z1}^{-1}, \hat{\alpha}_0 = [\hat{\alpha}, 0_{m\times m-r}], \quad \hat{u}_t = u_t - (\hat{\alpha} - \alpha)Z_{1,t-1}$. Then $\mathbf{E}(Z_{1,t-1}\hat{u}_t') = 0$. δ_R and δ_{R1} are defined as before. We have the same Lasso criterion function as in Theorem 2.3 which leads to the identical KKT optimality conditions Thus the Karush-Kuhn-Tucker (KKT) condition for group-wise variable selection from (A.26) is

$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t\widetilde{Z}_{1,t-1}' + \sqrt{T}\delta_{R1}\frac{1}{T}\sum_{t=1}^{T}\widetilde{Z}_{1,t-1}\widetilde{Z}_{1,t-1}' = -[\frac{\bar{\lambda}_{1,T}}{2\sqrt{T}}\frac{\bar{\alpha}(,1)}{||\bar{\alpha}(,1)||_2},\ldots,\frac{\bar{\lambda}_{r,T}}{2\sqrt{T}}\frac{\bar{\alpha}(,r)}{||\bar{\alpha}(,r)||_2}] \quad (24)$$

$$||\Big(\sum_{t=1}^{T} -2\frac{1}{T}w_t\widetilde{Z}_{2,t-1}' + 2\sqrt{T}\delta_{R1}\frac{1}{T^{3/2}}\widetilde{Z}_{1,t-1}\widetilde{Z}_{2,t-1}'\Big)_j||_2 < \frac{\bar{\lambda}_{r+j,T}}{T} \quad (25)$$

where $\bar{\lambda}_{j,T} = \frac{\lambda_T^{rank}}{\tilde{\mu}_j^{\gamma}}$ and the subscript j denotes the jth column.
According to the definition of $\hat{u}_t$, $||\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t Z_{1,t-1}'||_F = O_p(\sqrt{\frac{mr}{T}})$.

Rewrite

$$[\frac{\bar{\lambda}_{1,T}}{2\sqrt{T}}\frac{\hat{\alpha}(,1)}{||\hat{\alpha}(,1)||_2},\ldots,\frac{\bar{\lambda}_{r,T}}{2\sqrt{T}}\frac{\hat{\alpha}(,r)}{||\hat{\alpha}(,r)||_2}] = \frac{\lambda_T^{rank}}{2\sqrt{T}}[\frac{\hat{\alpha}(,1)}{||\hat{\alpha}(,1)||_2\tilde{\mu}_1^{\gamma}},\ldots,\frac{\hat{\alpha}(,r)}{||\hat{\alpha}(,r)||_2\tilde{\mu}_r^{\gamma}}] = \frac{\lambda_T^{rank}}{2\sqrt{T}}V_\alpha$$

Define

$$S_{z1z1} = \frac{1}{T}\sum_{t=1}^{T} Z_{1,t-1}Z_{1,t-1}' \qquad S_{uz1} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{u}_t Z_{1,t-1}'$$

$$S_{z1z2} = \frac{1}{T^{3/2}}\sum_{t=1}^{T} Z_{1,t-1}Z_{2,t-1}' \qquad S_{uz2} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t Z_{2,t-1}'$$

Then we can derive that

$$\sqrt{T}\delta_{R1} = -\frac{\lambda_T^{rank}}{2\sqrt{T}}V_\alpha S_{z1z1}^{-1} + S_{uz1}S_{z1z1}^{-1}$$

as $\frac{\lambda_T^{rank} r^{\tau_2\gamma+\frac{1}{2}}}{\sqrt{T}} \to 0$, $||\delta_{R1}||_F = O_p(\sqrt{\frac{mr}{T}})$.
To study the tail properties of elements in S_{uz2}, we need the following results based

on Beveridge-Nelson decomposition.

$$
\begin{aligned}
u_{t-q}u_t' &= (\sum_{j=0}^{\infty} A_j w_{t-q-j})(\sum_{k=0}^{\infty} A_k w_{t-k})' = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} A_j w_{t-q-j} w_{t-k}' A_k' \\
&= \sum_{j=q}^{\infty}\sum_{k=0}^{q-1} A_{j-q} w_{t-j} w_{t-k}' A_k' + \sum_{j=q}^{\infty}\sum_{k=q}^{\infty} A_{j-q} w_{t-j} w_{t-k}' A_k' \\
&= \sum_{j=q}^{\infty}\sum_{k=0}^{q-1} A_{j-q} w_{t-j} w_{t-k}' A_k' + \sum_{k=q}^{\infty} A_{k-q} w_{t-k} w_{t-k}' A_k' \\
&+ \sum_{j=q}^{\infty}\sum_{i=1}^{\infty} A_{j+i-q} w_{t-j-i} w_{t-j}' A_j' + \sum_{j=q}^{\infty}\sum_{i=1}^{\infty} A_{j-q} w_{t-j} w_{t-j-i}' A_{j+i}'
\end{aligned}
$$

The term $\sum_{k=q}^{\infty} A_{k-q} w_{t-k} w_{t-k}' A_k'$ can be further decomposed as

$$
\begin{aligned}
& \sum_{k=q}^{\infty} A_{k-q} w_{t-k} w_{t-k}' A_k' \\
= & \sum_{k=q}^{\infty} A_{k-q} w_{t-q} w_{t-q}' A_k' - \sum_{k=q+1}^{\infty} A_{k-q} w_{t-q} w_{t-q}' A_k' \\
+ & \sum_{k=q+1}^{\infty} A_{k-q} w_{t-q-1} w_{t-q-1}' A_k' - \sum_{k=q+2}^{\infty} A_{k-q} w_{t-q-1} w_{t-q-1}' A_k' \\
+ & \ldots \\
+ & \sum_{k=q+K}^{\infty} A_{k-q} w_{t-q-K} w_{t-q-K}' A_k' - \sum_{k=q+K+1}^{\infty} A_{k-q} w_{t-q-K} w_{t-q-K}' A_k' \\
+ & \ldots \\
= & \sum_{k=q}^{\infty} A_{k-q} w_{t-q} w_{t-q}' A_k' \\
- & (\sum_{k=q+1}^{\infty} A_{k-q} w_{t-q} w_{t-q}' A_k' - \sum_{k=q+1}^{\infty} A_{k-q} w_{t-q-1} w_{t-q-1}' A_k') \\
- & \ldots \\
- & (\sum_{k=q+K+1}^{\infty} A_{k-q} w_{t-q-K} w_{t-q-K}' A_k' - \sum_{k=q+K+1}^{\infty} A_{k-q} w_{t-q-K-1} w_{t-q-K-1}' A_k') \\
- & \ldots
\end{aligned}
$$

Therefore, if we sum up $\sum_{k=q}^{\infty} A_{k-q} w_{t-k} w_{t-k}' A_k'$ over t, only $\sum_{k=q}^{\infty} A_{k-q} w_{t-q} w_{t-q}' A_k'$ remains and the other terms get deleted.

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\sum_{k=q}^{\infty} A_{k-q} w_{t-k} w_{t-k}' A_k' \\
= \; & \frac{1}{T}\sum_{t=1}^{T}\sum_{k=q}^{\infty} A_{k-q} w_{t-q} w_{t-q}' A_k' + O_p(\frac{1}{T}) \\
= \; & \sum_{k=q}^{\infty} A_{k-q} (\frac{1}{T}\sum_{t=1}^{T} w_{t-q} w_{t-q}') A_k' + O_p(\frac{1}{T})
\end{aligned}$$

The expectation of $\frac{1}{T}\sum_{t=1}^{T}\sum_{k=q}^{\infty} A_{k-q} w_{t-k} w_{t-k}' A_k'$ is thus $\sum_{k=q}^{\infty} A_{k-q}\Sigma_w A_k'$. Each element in $\sum_{t=1}^{T}(w_t w_t' - \Sigma_w)$ constitute a martingale process due to the independence of w_t. The bounded second moment of elements in $w_t w_t' - \Sigma_w$ is ensured by the $(4+\delta)$-th moment condition for w_t or e_t. The martingale property of the other terms in $\sum_{t=1}^{T} u_{t-q} u_t'$ can be proved similarly since given $i \neq j$, $\frac{1}{T}\sum_{t=1}^{T} w_{t-i} w_{t-j}'$ converges to zero due to the independent w_t. To sum up, we have shown that each element in

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(u_{t-q}u_t' - \mathbf{E}(u_{t-q}u_t')\Big) \tag{26}$$

has bounded second moment. Because this result holds for a general $q \geqslant 0$, after linear combination, we can also show that each element in

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(Z_{1,t-1}u_t' - \mathbf{E}(Z_{1,t-1}u_t')\Big)$$

has bounded second moment. This is also true if we sum up (26) over q. However, to ensure the convergence, we must divide the new result by $\sqrt{T}$, i.e., each element in

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(\frac{1}{\sqrt{T}}\sum_{q=1}^{t} u_{t-q}u_t' - \frac{1}{\sqrt{T}}\sum_{q=1}^{t}\mathbf{E}(u_{t-q}u_t')\Big) \tag{27}$$

has bounded second moment. Summing up over q is well-defined due the fast convergence rate of A_j according to our assumption. Therefore, it holds also for

$$\frac{1}{T}\sum_{t=1}^{T}\Big(Z_{2,t-1}v_t' - \mathbf{E}(Z_{2,t-1}v_t')\Big) \tag{28}$$

From the block 6 in the proof of Theorem 4.3, the l_2 norm of $\frac{1}{T}\sum_{t=1}^{T} Z_{1,t-1}Z_{2,t-1}'$ is inflated by r^{τ_1} approximating l_2 norm of $(\beta'\alpha)^{-1}$, which make it more difficult to exclude the irrelevant groups compared to the i.i.d case.

To exclude the irrelevant part, KKT condition is satisfied if

$$||(S_{uz1}S_{z1z1}^{-1}S_{z1z2} - S_{uz2})_k||_2 < \frac{\lambda_T^{rank}}{2T}\tilde{\mu}_{r+k}^{-\gamma} - \frac{\lambda_T^{rank}}{2\sqrt{T}}||(V_\alpha S_{z1z1}^{-1}S_{z1z2})_k||_2 \tag{29}$$

$$
\begin{aligned}
& \frac{\lambda_T^{rank}}{2\sqrt{T}}||(V_\alpha S_{z1z1}^{-1}S_{z1z2})_j||_2) \leqslant \frac{\lambda_T^{rank}}{2\sqrt{T}}||V_\alpha S_{z1z1}^{-1}S_{z1z2}||_F \\
\leqslant \; & \frac{\lambda_T^{rank}}{2\sqrt{T}}||V_\alpha||_F||S_{z1z1}^{-1}||_2||S_{z1z2}||_2 \\
= \; & O_p\Big(\frac{\lambda_T^{rank} r^{\tau_1+\tau_2\gamma+\frac{1}{2}}}{T}\Big)
\end{aligned}
$$

Thus the RHS of (29) is dominated by the first term. The LHS of (29) is dominated by S_{uz2} since the l_2 norm of $S_{uz1}S_{z1z1}^{-1}S_{z1z2}$ converges to zero at the rate of $\frac{r^{\tau_1}}{\sqrt{T}}$ as S_{z1z2}. Denoting N_i as element in S_{uz2} and $\widetilde{N}_i$ as the perturbation of $\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t Z_{2,t-1}$ from the expectation. By the same argument as above, we have

$$
\begin{aligned}
& \mathbb{P}(\sqrt{\sum_{i=1}^{m} N_i^2} > \frac{\lambda_T^{rank}}{2T}\tilde{\mu}_{r+k}^{-\gamma}) \\
\leqslant \; & \mathbb{P}(\sum_{i=1}^{m} N_i^2 > \left(\frac{\lambda_T^{rank}}{2T}\tilde{\mu}_{r+k}^{-\gamma}\right)^2) \\
\leqslant \; & \sum_{i=1}^{m}\mathbb{P}(|N_i| > \frac{\lambda_T^{rank}}{2T\sqrt{m}}\tilde{\mu}_{r+k}^{-\gamma}) \\
\leqslant \; & \sum_{i=1}^{m}\mathbb{P}(|\widetilde{N}_i| + |c| > \frac{\lambda_T^{rank}}{2T\sqrt{m}}\tilde{\mu}_{r+k}^{-\gamma}) \\
\leqslant \; & \bar{C}_0 r^{2\tau_1} m \left(\frac{\sqrt{m} r^{\tau_1\gamma}}{\lambda_T^{rank}T^{\gamma-1}}\right)^2 = \bar{C}_0\left(\frac{m r^{\tau_1(\gamma+1)}}{\lambda_T^{rank}T^{\gamma-1}}\right)^2
\end{aligned}
$$

To make all the non-stationary parts excluded from the final estimator, we require that

$$
\frac{\lambda_T^{rank}T^{\gamma-1}}{m^{3/2}r^{\tau_1(\gamma+1)}} \to \infty
$$

□

References

Chen, X., Xu, M., Wu, W.B., 2013. Covariance and precision matrix estimation for high-dimensional time series. Ann. Statist. 41, 2994–3021.

Chung, K., Williams, R., 1990. Introduction to Stochastic Integration. Probability and Its Applications, Springer New York.

Stewart, G.W., Sun, J., 1990. Matrix Perturbation Theory. Academic Press.

Wu, W.B., 2007. Strong invariance principles for dependent random variables. Ann. Probab. 35, 2294–2320.

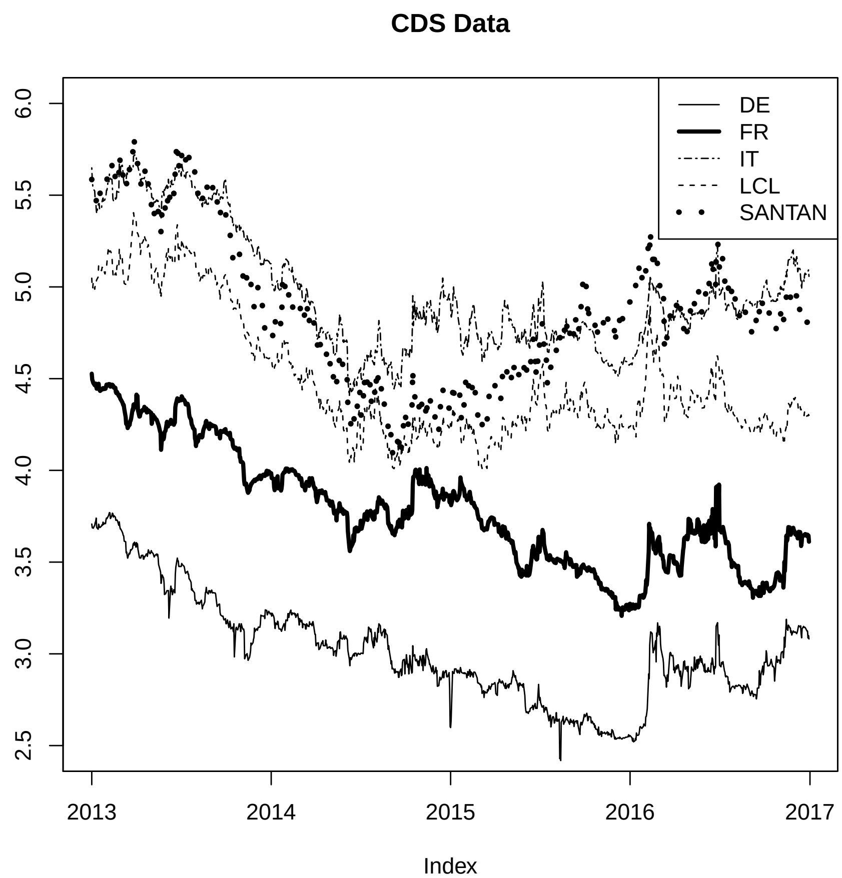

Figure 1: CDS data of Germany, France and Italy

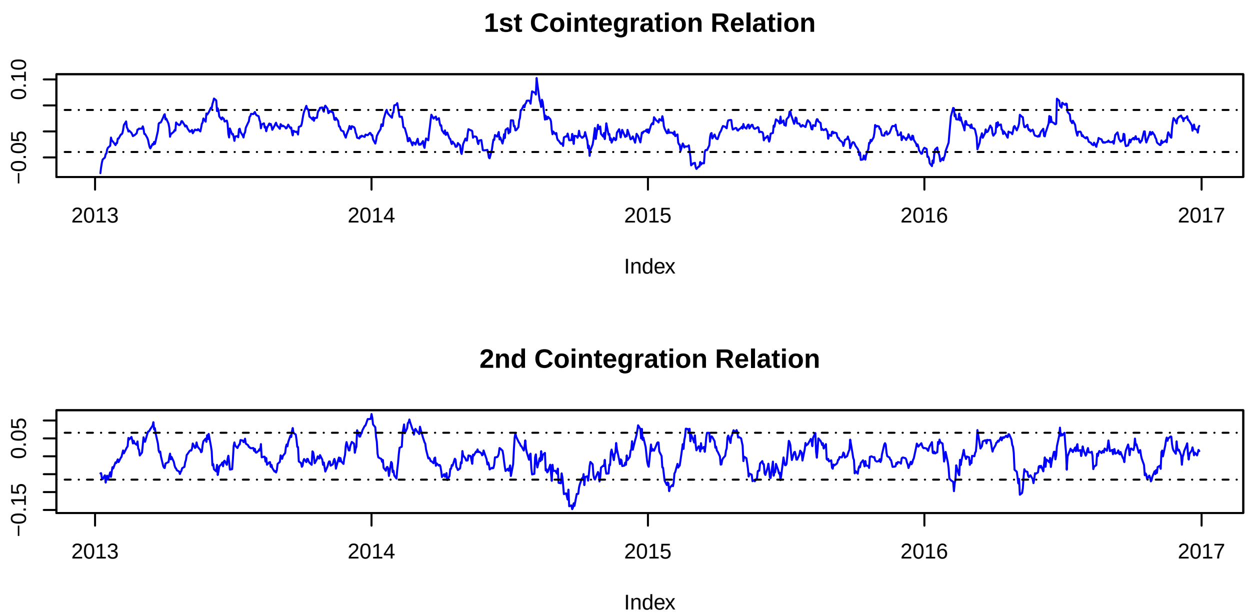

Figure 2: Significant cointegration relations

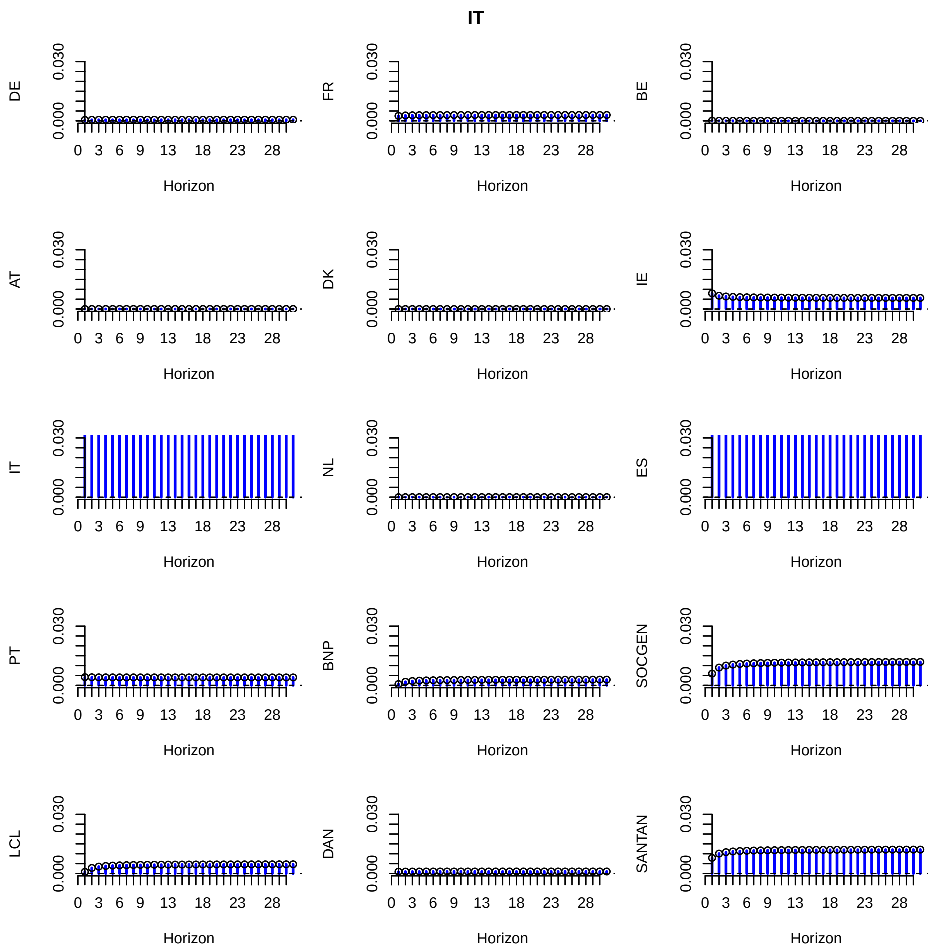

Figure 3: FEVD from Italy. The FEVD of Italy to itself is plotted as zero to highlight its contribution to others.