Active-Set Based Block Coordinate Descent Algorithm in Group LASSO for Self-Exciting Threshold Autoregressive Model: Supplementary Materials

Muhammad Jaffri Mohd Nasir[1*], Ramzan Nazim Khan[2†], Gopalan Nair[2†] and Darfiana Nur[3†]

[1*]Faculty of Entrepreneurship and Business, Universiti Malaysia Kelantan, Kota Bharu, 16100, Kelantan, Malaysia.
[2]Department of Mathematics and Statistics, The University of Western Australia, 35 Stirling Highway, Perth, 6009, Western Australia, Australia.
[3]School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Kent Street, Perth, 6102, Western Australia, Australia.

*Corresponding author(s). E-mail(s): jaffri.mn@umk.edu.mm; Contributing authors: nazim.khan@uwa.edu.au; gopalan.nair@uwa.edu.au; darfiana.nur@curtin.edu.au;
†These authors contributed equally to this work.

Algorithm A1: Single Line Search (*SLS*) for Group LASSO of the reformulated SETAR

Data: $\mathbf{y}_\pi \in \mathbb{R}^N$, $\mathbf{x}_{\pi_1} \in \mathbb{R}^{p+1}, \cdots, \mathbf{x}_{\pi_N} \in \mathbb{R}^{p+1}$, $\lambda_n \geq 0$, and $k_{\max} \geq 1$.

Result: $\widehat{\boldsymbol{\theta}}^N \leftarrow \boldsymbol{\theta}^N$ satisfying (9), and $\mathcal{B}$.

```
for j = 1, 2, ··· , N, do
    Obtain U_j and D_j, from ∑_{l=j}^N x_{π_l} x_{π_l}^T using SVD, such that
      ∑_{l=j}^N x_{π_l} x_{π_l}^T = U_j^T D_j U_j. Write D_j = diag(d_{j,1}, d_{j,2}, ··· , d_{j,p+1}).
Initialize: θ^N = (θ_{π_1}^T, θ_{π_2}^T, ··· , θ_{π_N}^T)^T ← 0 and B = ∅.
repeat
    for j = 1, 2, ··· , N, do
        if j = 1 then
            Compute θ_{π_1} = U_1^T D_1^{-1} U_1 f_1(B) and B = B ∪ {1}.
        else
            if (2 ||f_j||_2 / N) > λ_n then
                Compute U_j f_j(B) = (ν_{j,1}, ν_{j,2}, ··· , ν_{j,p+1})^T, where f_j(B)
                  is given in (27).
                Find the unique u_j > 0 satisfying (27). Then, compute
                  θ_{π_j} = U_j^T (D_j + (N λ_n)/(2 u_j) I_{p+1})^{-1} U_j f_j(B).
                Compute B = B ∪ {j}.
            else
                Set θ_{π_j} = 0.
                Compute B = B \ {j}.
until some convergence criterion of parameters is met.
```

Algorithm A2: Group Least Angle Regression (*gLAR*) Algorithm for the reformulated SETAR

Data: $\mathbf{y}_\pi \in \mathbb{R}^N$, $\mathbf{x}_{\pi_1} \in \mathbb{R}^{p+1}, \cdots, \mathbf{x}_{\pi_N} \in \mathbb{R}^{p+1}$, $\triangle_* \geq 0$ and $k_{\text{max}} \geq 1$.

Result: $\mathcal{B}$ and $\widehat{\boldsymbol{r}}$.

Initialize: **Compute** $j_\star = \arg\max_{j\in\{1,2,\cdots,N\}} \left\|\vec{g}_j(\boldsymbol{y}_\pi)\right\|_2$. **Update** $\widehat{\boldsymbol{\mu}} \leftarrow \mathbf{0}$ and $\boldsymbol{\upsilon} \leftarrow \boldsymbol{y}_\pi$, $\mathcal{B} \leftarrow \{j_*\}$ and $\mathcal{B}^* \leftarrow \{1, \cdots, N\} \backslash \{j_\star - \triangle_*, \cdots, j_\star + \triangle_*\}$.

repeat

- **Compute** least-squares *direction* estimate $\widehat{\boldsymbol{\theta}}_{\mathcal{B}}$ by $\widehat{\boldsymbol{\theta}}_{\mathcal{B}} = (\widetilde{X}_{\mathcal{B}}^T \widetilde{X}_{\mathcal{B}})^+ (\widetilde{X}_{\mathcal{B}}^T \boldsymbol{\upsilon}) \in \mathbb{R}^{k(p+1)}$, where $(.)^+$ is a pseudo-inverse operator.
- **Compute** *direction* fit $\widehat{\boldsymbol{y}}_{\mathcal{B}} = X_{\mathcal{B}} \widehat{\boldsymbol{\theta}}_{\mathcal{B}}$.
- **for** $j \in \mathcal{B}^*$ **do**
  - **Compute** $\alpha_j \in (0, 1]$ by solving $\left\|\vec{g}_j\left(\boldsymbol{\upsilon}\right) - \alpha_j \vec{g}_j\left(\widehat{\boldsymbol{y}}_{\mathcal{B}}\right)\right\|_2^2 = (1 - \alpha_j)^2 d$.
- **Compute** $\alpha = \min_{j\in\mathcal{B}^*}(\alpha_j)$ and $j^* = \arg\min_{j\in\mathcal{B}^*}(\alpha_j)$.
- **Update** $\mathcal{B} \leftarrow \mathcal{B} \cup \{j^*\}$, $\mathcal{B}^* \leftarrow \mathcal{B}^* \setminus \{j^* - \Delta_*, \cdots, j^* + \Delta_*\}$ and $\widehat{\boldsymbol{\mu}} \leftarrow \widehat{\boldsymbol{\mu}} + \alpha \widehat{\boldsymbol{y}}_{\mathcal{B}}$.
- **Compute** $\boldsymbol{\upsilon} = \boldsymbol{y}_\pi - \widehat{\boldsymbol{\mu}}$.

until $\text{card}(\mathcal{B}) = k_{\text{max}}$ or $\alpha = 1$.

Compute $b_{\text{min}} = \min_j(b_j \in \mathcal{B})$.

Update $\mathcal{B} \leftarrow \mathcal{B} \setminus \{b_{\text{min}}\}$.

Obtain $\widehat{\boldsymbol{r}} = \left\{y_{\pi_{l-1}} : l \in \mathcal{B}\right\} := \{\widehat{r}_1, \cdots, \widehat{r}_{\widehat{m}}\}$ where $\widehat{m} = \text{card}(\widehat{\boldsymbol{r}})$.