Detecting Multiple Structural Breaks in Systems of Linear Regression Equations with Integrated and Stationary Regressors – Supplementary Material B

September 17, 2024

1 Second step estimator - Adaptive group LASSO

As outlined in the main text, to obtain a consistent estimator for the number of breaks, their timing and coefficient changes, we need to design a second step refinement reducing the number of superfluous breaks. In the following, we outline the use of an adaptive group LASSO estimator with group LASSO estimates as weights as a second step. To prove consistency of the adaptive group LASSO estimator, slightly different assumptions about the minimum break intervals are needed:

Assumption 3-b. (i) $I_{\min} = \min\limits_{1 \leq j \leq m_0+1} |t_j^0 - t_{j-1}^0| > \zeta T$ for some $\zeta > 0$, where $I_{\min}$ denotes the minimum break interval.
(ii) The break magnitudes are bounded to satisfy $m_\theta = \min\limits_{1 \leq j \leq m_0+1} \|\boldsymbol{\theta}^0_{t^0_{j-1}}\| \geq \nu > 0$ and $M_\theta = \max\limits_{1 \leq j \leq m_0+1} \|\boldsymbol{\theta}^0_{t^0_{j-1}}\| \leq \mathcal{V} < \infty$.

Note that part (i) of Assumption 3-b requires the minimum break interval to grow at rate T which is a slightly stronger assumption than part (i) of Assumption 3 in the main text. Consequently, Theorem 2 can be proven under these conditions as well.

According to Theorem 2, the group LASSO estimator slightly overselects breaks under the right tuning. The algorithm employed to estimate $\hat{\boldsymbol{\theta}}(T)$ allows to pre-specify the maximum number of breakpoint candidates M, i.e. the maximum number of non-zero groups in $\hat{\boldsymbol{\theta}}(T)$, and the minimum distance between breaks. In line with Assumption 3-b, the minimum distance needs to be specified such that $I_{\min} = O(T)$. Since the group LASSO overselects breaks in the first step, M should be set large enough to encompass all true breakpoints and some additional falsely selected non-zero groups. This condition guarantees that $\hat{\boldsymbol{\theta}}(T)$ always contains Md elements. In turn, $Td - Md$ columns of $\boldsymbol{Z}$ corresponding to zero coefficients are eliminated during the first step to result in the $Tq \times Md$ design matrix $\boldsymbol{Z}_S$. Hence, for given

$M \ll T$, the column size of the new design matrix is substantially smaller than the original size Td and does not longer depend on the sample size. This allows us to further assume that all eigenvalues of $\boldsymbol{\Sigma}_S = \boldsymbol{Z}_S'\boldsymbol{Z}_S/T$ are contained in the interval $[c_*, c^*]$, where c_* and c^* are two positive constants. This means that we can relate to a restricted eigenvalue condition similar to Bickel et al. (2009) for the second step estimation. While the restricted eigenvalue condition in general does not hold for change-point settings, the dimension reduction of the first step allows us to state this assumption for our reduced design matrix. It should be noted that this assumption for the second step estimation is further justified by the minimum subsample size needed to precisely estimate coefficient changes. Consequently, M should be chosen so that the average regime length in case of equidistantly-spaced breaks still guarantees enough observations per regime to estimate all coefficient changes. In principle, we can re-estimate the model with post-LASSO OLS (the algorithm ensures that there is no local collinearity) before the adaptive group LASSO estimator is applied to benefit from the higher convergence rate of the OLS estimator after $T - M$ break candidates are eliminated. This comes at very little computational costs after finding the preliminary model specification with M breakpoint candidates. The important assumptions for consistency of the adaptive group LASSO are that consistent weights are available and the design matrix fulfills the restricted eigenvalue condition.

We follow Wang and Leng (2008), Behrendt and Schweikert (2021), and Schweikert (2022) and define the adaptive group LASSO objective function

$$Q(\boldsymbol{\theta}_S) \quad = \quad \frac{1}{T}\left\|\boldsymbol{Y} - \boldsymbol{Z}_S\boldsymbol{\theta}_S\right\|^2 + \lambda_S \sum_{i=1}^{M} w_i \|\boldsymbol{\theta}_{S,i}\|, \tag{S.1}$$

where $\gamma > 0$ and w_i are the group-specific weights assigned as follows

$$w_i = \begin{cases} \|\hat{\boldsymbol{\theta}}^P_{S,i}\|^{-\gamma} & \text{if} \quad \hat{\boldsymbol{\theta}}_{S,i} \neq 0 \\ \infty & \text{if} \quad \hat{\boldsymbol{\theta}}_{S,i} = 0, \end{cases}$$

and set $0 \times \infty = 0$. $\hat{\boldsymbol{\theta}}^P_{S,i}$, $i = 1, \ldots, |\mathcal{A}_T| + 1$ denotes the post-LASSO OLS estimates for the non-zero groups obtained from optimizing the group LASSO objective function in the first step. The remaining $M - |\mathcal{A}_T| - 1$ group elements of $\hat{\boldsymbol{\theta}}_S$ can be filled with zero groups as long as their selected indices lead to $\boldsymbol{\Sigma}_S$ being a positive definite matrix for all T.

We denote the estimator minimizing $Q(\boldsymbol{\theta}_S)$ with $\hat{\hat{\boldsymbol{\theta}}}_S = \arg\min_{\boldsymbol{\theta}_S} Q$. Eliminating columns from the initial design matrix requires a mapping of our second step indices to recover the original indices. For notational convenience, we use the mapping $g : \mathbb{N} \to \mathbb{N},\ i \mapsto g(i) = t_i$, where t_i is the breakpoint corresponding to the index i, for this purpose and define the index set $\bar{\mathcal{A}}^*$ ($\mathcal{A}^*$) to pick out the elements that correspond to truly non-zero coefficients (coefficient changes).

The next asymptotic result shows that the adaptive group LASSO estimator provides consistent parameter estimation and model selection

Theorem 5. If Assumptions 1 - 3-b hold, $\lambda_S \to 0$, $\lambda_S^2 T^\gamma \to \infty$ for $\gamma > 0$, then
(a) $\|\hat{\hat{\boldsymbol{\theta}}}_S - \boldsymbol{\theta}_S^0\| = O_p(T^{-\frac{1}{2}})$
(b) $P(g(\{j \geq 2 : \|\hat{\hat{\boldsymbol{\theta}}}_{S,j}\| \neq 0\}) = \mathcal{A}) \to 1$.

Remark 1. Part (a) of Theorem 5 gives a uniform convergence rate for all coefficients of the model which is, of course, accomplished by the scaling factors applied to the integrated regressors and the linear trend to ensure that all regressors have the same order of integration. Post-LASSO OLS without scaling factors can be used to benefit from higher convergence rates of each component. Essentially, part (a) is a necessary byproduct of the more important result about model selection consistency in part (b).

Turning to the finite sample case, we need a strategy to select the tuning parameter λ_T. If we apply the block coordinate decent algorithm to solve the second step adaptive group LASSO estimation, we need to evaluate a set of tuning parameters to find the optimal value with respect to a pre-specified criterion. We consider two data-dependent selection rules often applied in the literature, namely conventional information criteria like the BIC or MDL-based criteria. Our asymptotic results show that it is crucial to let the tuning parameter grow at the right rate. However, this rate provides only limited practical guidance towards the choice of the tuning parameter. First, we follow Kock (2016), Qian and Su (2016) and Schweikert (2022) and select λ_T by minimizing an information criterion in the form of

$$IC^*(\lambda_T) = \log(|\hat{\Sigma}|) + \rho_T \hat{D}, \tag{S.2}$$

where $\hat{\Sigma}$ denotes the estimated covariance matrix of the residuals $\hat{u}_t$ resulting from the adaptive group LASSO estimation of Equation (S.1) and $\hat{D}$ denotes the number of estimated coefficients. The penalty function ρ_T allows for different choices. While Kock (2016) suggests to use the BIC for potentially nonstationary autoregressive models which corresponds to $\rho_T = \log(T)/T$, Qian and Su (2016) propose to use $\rho_T = 1/\sqrt{T}$ for the estimation of structural breaks in stationary time series regressions. Here, we follow Kock (2016) and apply the BIC.

Second, we select λ_T by minimizing the minimum description length (Hansen and Yu, 2001; Davis et al., 2006):

$$MDL^*(\lambda_T) = \log_2(m) + m\log_2(T) + \sum_{j=1}^{m} \frac{d + q(q+1)/2}{2} \log_2(T_j) + \frac{T}{2}\log(|\hat{\Sigma}|) - \frac{1}{2}\sum_{j=1}^{T} \mathrm{tr}(\hat{\Sigma}^{-1}\hat{u}_t\hat{u}_t'). \tag{S.3}$$

The principle of MDL is to find the best fitting model from set of candidate models as the one that produces the shortest code length. Here, we need approximately $\log_2(m)$ bits to store the information about the number of breaks, at most $m \log_2(T)$ bits to store the location of the breaks, and at most $\sum_{j=1}^{m} \frac{qd+q(q+1)/2}{2} \log_2(T_j)$ to store the total number of parameters (coefficients and the residual covariance matrix). The remaining terms relate to the negative loglikelihood of the fitted model and represent the code length of the residuals $\hat{u}_t$ (Rissanen, 1989; Park, 1993).

2 Simulation results

Table 1: Estimation of (multiple) structural breaks in the full model with adaptive group LASSO (BIC)

	SB1: ($\tau = 0.5$)				
T	*pce*	τ			
100	98.0	0.503 (0.015)			
200	99.3	0.500 (0.004)			
400	100	0.500 (0.003)			
800	99.9	0.500 (0.001)			
	SB2: ($\tau_1 = 0.33$, $\tau_2 = 0.67$)				
T	*pce*	τ_1	τ_2		
150	97.7	0.326 (0.002)	0.667 (0.001)		
300	98.1	0.330 (0.001)	0.670 (0.000)		
600	98.5	0.330 (0.001)	0.670 (0.000)		
1200	98.5	0.330 (0.001)	0.670 (0.001)		
	SB4: ($\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$, $\tau_4 = 0.8$)				
T	*pce*	τ_1	τ_2	τ_3	τ_4
250	57.8	0.205 (0.022)	0.400 (0.008)	0.600 (0.004)	0.800 (0.002)
500	74.4	0.201 (0.006)	0.400 (0.002)	0.600 (0.002)	0.800 (0.001)
1000	87.5	0.200 (0.002)	0.400 (0.001)	0.600 (0.001)	0.800 (0.001)
2000	89.3	0.200 (0.001)	0.400 (0.001)	0.600 (0.001)	0.800 (0.000)

Note: We use 1,000 replications of the data-generating process given in Equation (10) in the main text. The variance of the error terms is $\sigma_\xi^2 = 1$ and $\sigma_u^2 = 2$, respectively. The first panel reports the results for one active breakpoint at $\tau = 0.5$, the second panel considers two active breakpoints at $\tau_1 = 0.33$ and $\tau_2 = 0.67$ and the third panel has four active breakpoints at $\tau_1 = 0.2$, $\tau_2 = 0.4$, $\tau_3 = 0.6$, and $\tau_4 = 0.8$. Standard deviations are given in parentheses.

3 Proofs

Lemma 5. *We define the $Tq \times d$ matrix $\boldsymbol{Z}_{S,i}$ selecting all columns from $\boldsymbol{Z}_S$ belonging to breakpoint candidate i. Then, a necessary and sufficient condition for the estimator $\hat{\hat{\boldsymbol{\theta}}}_S$ to be* *a solution to the adaptive group LASSO objective function* $Q(\boldsymbol{\theta})$ *is*

$$\boldsymbol{Z}_{S,i}'\left(\boldsymbol{Y}-\boldsymbol{Z}_S\hat{\hat{\boldsymbol{\theta}}}_S\right)-\frac{1}{2}T\lambda_S\|\hat{\boldsymbol{\theta}}_{S,i}\|^{-\gamma}\frac{\hat{\hat{\boldsymbol{\theta}}}_{S,i}}{\|\hat{\hat{\boldsymbol{\theta}}}_{S,i}\|}=0,\qquad \forall\hat{\hat{\boldsymbol{\theta}}}_{S,i}\neq 0,$$

and

$$\|\boldsymbol{Z}_{S,i}'\left(\boldsymbol{Y}-\boldsymbol{Z}_S\hat{\hat{\boldsymbol{\theta}}}_S\right)\|\leq\frac{1}{2}T\lambda_S\|\hat{\boldsymbol{\theta}}_{S,i}\|^{-\gamma},\qquad \forall\hat{\hat{\boldsymbol{\theta}}}_{S,i}=0,$$

where $\|\hat{\boldsymbol{\theta}}_{S,i}\|^{-\gamma}$ *are the group-specific weights.*

Proof of Lemma 5. This lemma is a direct consequence of the Karush-Kuhn-Tucker (KKT) conditions for adaptive group LASSO estimators.

□

Proof of Theorem 5. The adaptive group LASSO objective function $Q(\boldsymbol{\theta}_S)$ is a strictly convex function. We show that there is a local minimizer that is consistent and by global convexity of $Q(\boldsymbol{\theta}_S)$, it follows that such a local minimizer must be $\hat{\hat{\boldsymbol{\theta}}}_S$. Similar as in Fan and Li (2001), the existence of an above-described local minimizer is implied by the fact that for any $\epsilon>0$, there is a sufficiently large constant $C>0$, such that

$$\liminf_T P\left(\inf_{\boldsymbol{v}:=(\boldsymbol{v}_1,\dots,\boldsymbol{v}_M)\in\mathbb{R}^{Md}:\|\boldsymbol{v}\|=C} Q(\boldsymbol{\theta}_S^0+T^{-\frac{1}{2}}\boldsymbol{v})>Q(\boldsymbol{\theta}_S^0)\right)>1-\epsilon. \tag{S.4}$$

It holds that

$$\begin{aligned}
& Q(\boldsymbol{\theta}_S^0+T^{-\frac{1}{2}}\boldsymbol{v})-Q(\boldsymbol{\theta}_S^0)\\
= & \frac{1}{T}\|\boldsymbol{Y}-\boldsymbol{Z}_S(\boldsymbol{\theta}_S^0+T^{-\frac{1}{2}}\boldsymbol{v})\|^2+\lambda_S\sum_{i=1}^M w_i\|(\boldsymbol{\theta}_{S,i}^0+T^{-\frac{1}{2}}\boldsymbol{v}_i)\|\\
& -\frac{1}{T}\|\boldsymbol{Y}-\boldsymbol{Z}_S\boldsymbol{\theta}_S^0\|^2-\lambda_S\sum_{i=1}^M w_i\|\boldsymbol{\theta}_{S,i}^0\|\\
= & \frac{1}{T}\boldsymbol{v}'\left(\frac{1}{T}\boldsymbol{Z}_S'\boldsymbol{Z}_S\right)\boldsymbol{v}-\frac{2}{T^{\frac{3}{2}}}\boldsymbol{v}'\boldsymbol{Z}_S'\boldsymbol{U} \qquad\qquad (\text{S.5})\\
& +\lambda_S\sum_{i=1}^M w_i\|(\boldsymbol{\theta}_{S,i}^0+T^{-\frac{1}{2}}\boldsymbol{v}_i)\|-\lambda_S\sum_{i=1}^M w_i\|\boldsymbol{\theta}_{S,i}^0\|\\
\geq & \frac{1}{T}\boldsymbol{v}'\left(\frac{1}{T}\boldsymbol{Z}_S'\boldsymbol{Z}_S\right)\boldsymbol{v}-\frac{2}{T^{\frac{3}{2}}}\boldsymbol{v}'\boldsymbol{Z}_S'\boldsymbol{U}\\
& +\lambda_S\sum_{g(i)\in\mathcal{A}_T\cap\mathcal{A}}\|\hat{\boldsymbol{\theta}}_{S,i}^P\|^{-\gamma}\left(\|(\boldsymbol{\theta}_{S,i}^0+T^{-\frac{1}{2}}\boldsymbol{v}_i)\|-\|\boldsymbol{\theta}_{S,i}^0\|\right)\\
\geq & \frac{1}{T}\boldsymbol{v}'\left(\frac{1}{T}\boldsymbol{Z}_S'\boldsymbol{Z}_S\right)\boldsymbol{v}-\frac{2}{T^{\frac{3}{2}}}\boldsymbol{v}'\boldsymbol{Z}_S'\boldsymbol{U}-\frac{1}{T}\lambda_S\sum_{g(i)\in\mathcal{A}_T\cap\mathcal{A}}\|\hat{\boldsymbol{\theta}}_{S,i}^P\|^{-\gamma}\|\boldsymbol{v}_i\|\\
= & I_1-I_2-I_3.
\end{aligned}$$

Since the restricted eigenvalue condition holds for $\boldsymbol{\Sigma}_S = \boldsymbol{Z}'_S \boldsymbol{Z}_S / T$, i.e., its eigenvalues are positive for all T, and $\boldsymbol{\Sigma}_S$ thus converges to a positive definite random matrix, we have $I_1 = O_p(T^{-1})\|\boldsymbol{v}\|^2$. Further, it follows from Cauchy-Schwarz inequality and arguments similar to those used in Lemma 2 that

$$\begin{aligned} E|I_2|^2 &= \frac{4}{T^3} E\left(\boldsymbol{v}'\boldsymbol{Z}'_S\boldsymbol{U}\right)^2 \\ &\leq \frac{4}{T^3}\|\boldsymbol{v}\|^2 E\|\boldsymbol{Z}'_S\boldsymbol{U}\|^2 \\ &= \frac{1}{T^2}\|\boldsymbol{v}\|^2 O_p(1), \end{aligned} \tag{S.6}$$

and consequently $I_2 = O_p(T^{-1})\|\boldsymbol{v}\|$. Finally, using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} I_3 &\leq \frac{1}{T}\lambda_S \left(\sum_{g(i)\in\mathcal{A}_T\cap\mathcal{A}} \|\hat{\boldsymbol{\theta}}_{S,i}\|^{-2\gamma} \right)^{1/2} \|\boldsymbol{v}\| \\ &\leq \frac{1}{T}\lambda_S m_0^{1/2} \min_{g(i)\in\mathcal{A}_T\cap\mathcal{A}} \|\hat{\boldsymbol{\theta}}^P_{S,i}\|^{-\gamma}\|\boldsymbol{v}\|. \end{aligned} \tag{S.7}$$

We note that $\min_{g(i)\in\mathcal{A}_T\cap\mathcal{A}} \|\hat{\boldsymbol{\theta}}^P_{S,i}\|^{-\gamma} = O_p(1)$ since $\hat{\boldsymbol{\theta}}^P_{S,i}$ is a consistent post-LASSO OLS estimator and the first step estimation does not ignore relevant breakpoints asymptotically according to Theorem 2. Using the condition $\lambda_S \to 0$, we know that I_3 is bounded by $O_p(T^{-1})\|\boldsymbol{v}\|$. Hence, we can specify a large enough constant C such that I_1 dominates I_2 and I_3. This completes the proof of part (a).

Next, we turn to the proof of part (b). Lemma 5 gives the necessary and sufficient condition for an estimator to be a solution to the adaptive group LASSO objective function as defined by Equation (S.1). Now, to prove that all truly zero coefficients are set to zero almost surely, it suffices to show that

$$P\left(\forall g(i) \in \mathcal{A}_T \cap \mathcal{A}^c, \|\frac{1}{T}\boldsymbol{Z}'_{S,i}(\boldsymbol{Y} - \boldsymbol{Z}_{S,\mathcal{A}^*}\hat{\hat{\boldsymbol{\theta}}}_{S,\mathcal{A}^*})\| \leq \frac{1}{2}\lambda_S\|\hat{\boldsymbol{\theta}}^P_{S,i}\|^{-\gamma}\right) \to 1, \tag{S.8}$$

or equivalently

$$P\left(\exists g(i) \in \mathcal{A}_T \cap \mathcal{A}^c, \|\frac{1}{T}\boldsymbol{Z}'_{S,i}(\boldsymbol{Y} - \boldsymbol{Z}_{S,\mathcal{A}^*}\hat{\hat{\boldsymbol{\theta}}}_{S,\mathcal{A}^*})\| > \frac{1}{2}\lambda_S\|\hat{\boldsymbol{\theta}}^P_{S,i}\|^{-\gamma}\right) \to 0. \tag{S.9}$$

It holds that

$$\begin{aligned} &P\left(\exists g(i) \in \mathcal{A}_T \cap \mathcal{A}^c, \|\frac{1}{T}\boldsymbol{Z}'_{S,i}(\boldsymbol{Y} - \boldsymbol{Z}_{S,\mathcal{A}^*}\hat{\hat{\boldsymbol{\theta}}}_{S,\mathcal{A}^*})\| > \frac{1}{2}\lambda_S\|\hat{\boldsymbol{\theta}}^P_{S,i}\|^{-\gamma}\right) \\ \leq\ &P\left(\exists g(i) \in \mathcal{A}_T \cap \mathcal{A}^c, \|\frac{1}{T}\boldsymbol{Z}'_{S,i}\boldsymbol{U}\| > \frac{1}{2}\lambda_S\|\hat{\boldsymbol{\theta}}^P_{S,i}\|^{-\gamma} - \|\frac{1}{T}\boldsymbol{Z}'_{S,i}\boldsymbol{Z}_{S,\mathcal{A}^*}\left(\hat{\hat{\boldsymbol{\theta}}}_{S,\mathcal{A}^*} - \boldsymbol{\theta}^0_{S,\mathcal{A}^*}\right)\|\right). \end{aligned} \tag{S.10}$$

Further, we have

$$
\begin{aligned}
& \|\frac{1}{T}\boldsymbol{Z}'_{S,i}\boldsymbol{Z}_{S,\mathcal{A}^*}\left(\hat{\hat{\boldsymbol{\theta}}}_{S,\mathcal{A}^*}-\boldsymbol{\theta}^0_{S,\mathcal{A}^*}\right)\| \\
\leq\ & \frac{1}{\sqrt{T}}\left[\left(\hat{\hat{\boldsymbol{\theta}}}_{S,\mathcal{A}^*}-\boldsymbol{\theta}^0_{S,\mathcal{A}^*}\right)'\boldsymbol{Z}'_{S,\mathcal{A}^*}\left(\frac{1}{T}\boldsymbol{Z}_{S,i}\boldsymbol{Z}'_{S,i}\right)\boldsymbol{Z}_{S,\mathcal{A}^*}\left(\hat{\hat{\boldsymbol{\theta}}}_{S,\mathcal{A}^*}-\boldsymbol{\theta}^0_{S,\mathcal{A}^*}\right)\right]^{1/2} \\
\leq\ & c^*\|\hat{\hat{\boldsymbol{\theta}}}_{S,\mathcal{A}^*}-\boldsymbol{\theta}^0_{S,\mathcal{A}^*}\|,
\end{aligned}
\tag{S.11}
$$

and the first part of Theorem 5 implies that $\|\hat{\hat{\boldsymbol{\theta}}}_{S,\mathcal{A}^*}-\boldsymbol{\theta}^0_{S,\mathcal{A}^*}\| = O_p(T^{-\frac{1}{2}})$. Hence, we need to show that

$$
P\left(\exists g(i)\in\mathcal{A}_T\cap\mathcal{A}^c, \|\frac{1}{T}\boldsymbol{Z}'_{S,i}\boldsymbol{U}\| > \frac{1}{2}\lambda_S\|\hat{\boldsymbol{\theta}}^P_{S,i}\|^{-\gamma} - c^*\|\hat{\hat{\boldsymbol{\theta}}}_{S,\mathcal{A}^*}-\boldsymbol{\theta}^0_{S,\mathcal{A}^*}\|\right)\to 0. \tag{S.12}
$$

Using post-LASSO OLS estimates as our weights implies $\max_{g(i)\in\mathcal{A}^c}\|\hat{\boldsymbol{\theta}}^P_{S,i}\| = O_p(T^{-1/2})$ and we have

$$
\begin{aligned}
& P\left(\exists g(i)\in\mathcal{A}_T\cap\mathcal{A}^c, \|\frac{1}{T}\boldsymbol{Z}'_{S,i}\boldsymbol{U}\| > \frac{1}{2}\lambda_S\|\hat{\boldsymbol{\theta}}^P_{S,i}\|^{-\gamma} - c^*\|\hat{\hat{\boldsymbol{\theta}}}_{S,\mathcal{A}^*}-\boldsymbol{\theta}^0_{S,\mathcal{A}^*}\|\right) \\
\leq\ & P\left(\exists g(i)\in\mathcal{A}_T\cap\mathcal{A}^c, \|\frac{1}{T}\boldsymbol{Z}'_{S,i}\boldsymbol{U}\| > \frac{\lambda_S}{2\left(\max_{i\in\mathcal{A}^c}\|\hat{\boldsymbol{\theta}}^P_{S,i}\|\right)^{\gamma}} - c^*\|\hat{\hat{\boldsymbol{\theta}}}_{S,\mathcal{A}^*}-\boldsymbol{\theta}^0_{S,\mathcal{A}^*}\|\right) \\
\leq\ & P\left(\exists g(i)\in\mathcal{A}_T\cap\mathcal{A}^c, \|\frac{1}{T}\boldsymbol{Z}'_{S,i}\boldsymbol{U}\| > \frac{\lambda_S}{2}\left(\frac{C}{\sqrt{T}}\right)^{-\gamma} - Cc^*\frac{1}{\sqrt{T}}\right) + o_p(1) \\
\leq\ & P\left(\exists g(i)\in\mathcal{A}_T\cap\mathcal{A}^c, \|\frac{1}{T}\boldsymbol{Z}'_{S,i}\boldsymbol{U}\| > \frac{\lambda_S}{4}\left(\frac{C}{\sqrt{T}}\right)^{-\gamma}\right) + o_p(1) \\
\leq\ & \sum_{g(i)\in\mathcal{A}_T\cap\mathcal{A}^c} P\left(\|\frac{1}{T}\boldsymbol{Z}'_{S,i}\boldsymbol{U}\| > \frac{\lambda_S}{4}\left(\frac{C}{\sqrt{T}}\right)^{-\gamma}\right) \\
\leq\ & \sum_{g(i)\in\mathcal{A}_T\cap\mathcal{A}^c} \frac{E\|\frac{1}{T}\boldsymbol{Z}'_{S,i}\boldsymbol{U}\|^2}{\frac{1}{16}\lambda_S^2\left(\frac{C}{\sqrt{T}}\right)^{-2\gamma}},
\end{aligned}
\tag{S.13}
$$

for some $C>0$. The third inequality follows from the condition $\lambda_S^2 T^{\gamma}\to\infty$. Since Assumption 1 implies that $E\|\frac{1}{T}\boldsymbol{Z}'_{S,i}\boldsymbol{U}\|^2 = O_p(T^{-1})$, we have

$$
\frac{16E\|\frac{1}{T}\boldsymbol{Z}'_{S,i}\boldsymbol{U}\|^2}{C^{-2\gamma}\lambda_S^2 T^{\gamma}}\to 0, \tag{S.14}
$$

for all $g(i)\in\mathcal{A}_T\cap\mathcal{A}^c$. Note that $|\mathcal{A}_T| < M$ for all T and that all remaining indices i not included in $\mathcal{A}_T$ correspond to coefficients which have already been set to zero in the first step.

For the proof of model selection consistency, we still need to show that no truly non-zero coefficient changes are set to zero. It holds that

$$\min_{g(i)\in\mathcal{A}} \|\hat{\hat{\boldsymbol{\theta}}}_{S,i}\| \geq \min_{g(i)\in\mathcal{A}} \|\boldsymbol{\theta}^0_{S,i}\| - \max_{g(i)\in\mathcal{A}} \|\hat{\hat{\boldsymbol{\theta}}}_{S,i} - \boldsymbol{\theta}^0_{S,i}\|. \tag{S.15}$$

Since $\|\hat{\hat{\boldsymbol{\theta}}}_{S,i} - \boldsymbol{\theta}^0_{S,i}\| \xrightarrow{p} 0$ by part (a) and by considering Assumption 3-b, we have

$$P\left(\min_{g(i)\in\mathcal{A}} \|\hat{\hat{\boldsymbol{\theta}}}_{S,i}\| \geq \nu\right) \to 1. \tag{S.16}$$

This completes the proof of Theorem 5.

□

References

Behrendt, S., Schweikert, K., 2021. A Note on Adaptive Group Lasso for Structural Break Time Series. Econometrics and Statistics 17, 156–172. doi:10.1016/j.ecosta.2020.04.001.

Bickel, P.J., Ritov, Y., Tsybakov, A.B., 2009. Simultaneous analysis of lasso and dantzig selector. Annals of Statistics 37, 1705–1732. doi:10.1214/08-AOS620.

Davis, R.A., Lee, T.C.M., Rodriguez-Yam, G.A., 2006. Structural Break Estimation for Nonstationary Time Series Models. Journal of the American Statistical Association 101, 223–239. doi:10.1198/016214505000000745.

Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96, 1348–1360. doi:10.1198/016214501753382273.

Hansen, M.H., Yu, B., 2001. Model selection and the principle of minimum description length. Journal of the American Statistical Association 96, 746–774. doi:10.1198/016214501753168398.

Kock, A.B., 2016. Consistent and Conservative Model Selection With the Adaptive Lasso in Stationary and Nonstationary Autoregressions. Econometric Theory 32, 243–259. doi:10.1017/S0266466615000304.

Park, T., 1993. Equivalence of maximum likelihood estimation and iterative two-stage estimation for seemingly unrelated regression models. Communications in Statistics - Theory and Methods 22, 2285–2296. doi:10.1080/03610929308831148.

Qian, J., Su, L., 2016. Shrinkage Estimation of Regression Models With Multiple Structural Changes. Econometric Theory 32, 1376–1433. doi:10.1017/S0266466615000237.

Rissanen, J., 1989. Stochastic Complexity in Statistical Inquiry. 1 ed., World Scientific Publishing Company, River Edge, NJ, United States.

Schweikert, K., 2022. Oracle Efficient Estimation of Structural Breaks in Cointegrating Regressions. Journal of Time Series Analysis 43, 83–104. doi:10.1111/jtsa.12593.

Wang, H., Leng, C., 2008. A note on adaptive group lasso. Computational Statistics and Data Analysis 52, 5277–5286. doi:10.1016/j.csda.2008.05.006.