Supplementary Materials to "Multiscale Autoregression on Adaptively Detected Timescales"

Rafal Baranowski Yining Chen
Piotr Fryzlewicz

Department of Statistics, London School of Economics and Political Science

A Additional numerical experiments

Here we report the experiments with series simulated from non-stationary AR models with unit roots. The scenarios we consider are similar to (M1) – (M4) listed in the main manuscript, with their details outlined below.

(M1') Same as (M1) but with $\alpha_1 = 0.4$, $\alpha_2 = 0.6$ (i.e. $\boldsymbol{\beta} = (0.6, 0.2, 0.2)^T$).

(M2') Same as (M2) but with $\alpha_1 = 2$, $\alpha_2 = -1$ (i.e. $\boldsymbol{\beta} = (0.8, 0.8, -0.2, -0.2, -0.2)^T$).

(M3') Same as (M3) but with $\alpha_1 = 0.5$, $\alpha_2 = -1$, $\alpha_3 = 1.4$ (i.e. $\boldsymbol{\beta} = (0.5, -0.1, -0.1, -0.1, -0.1, 0.1, \ldots, 0.1)^T$).

(M4') Same as (M4) but with $\alpha_1 = 1$, $\alpha_2 = -4.8$, $\alpha_3 = 10.2$, $\alpha_4 = -6.4$ (i.e. $\boldsymbol{\beta} = (1, 0, \ldots, 0, 0.8, -0.8)^T$, so $\varepsilon_t = (1 - 0.8B^7)(1 - B)X_t$).

The corresponding results are summarised in Table 4.

We see that even in the setting of non-stationary observations, AMAR still performs much better than its competitors in most settings, even though all methods seem to perform worse as compared to the stationary settings, and here the reported results are associated with larger standard errors.

In addition, we note that the fused LASSO approach performs much worse than its competitors in terms of MSPE, especially in (M2') and (M4'). This is because the fused LASSO approach tends to over-estimate the number of scales, resulting in less accurate $\hat{\boldsymbol{\beta}}$, which greatly affects MSPE in the next 100 steps under the unit-root setting, where the prediction errors could accumulate exponentially.

Model (M1')										
	$E\lvert\hat{q}-q\rvert$		$E(D_H)$		$E\lVert\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rVert$			MSPE − 1		
Method	AMAR	Fused	AMAR	Fused	AMAR	Fused	AIC	AMAR	Fused	AIC
$T=400$	**0.269**	1.652	**1.447**	18.655	0.034	0.3661	**0.0272**	0.1095	3.4853	**3.2267**
$T=800$	**0.119**	2.124	**0.941**	22.277	**0.0177**	0.3785	0.0194	**0.0248**	3.3573	0.7336
$T=1500$	**0.148**	2.112	**1.312**	22.670	**0.0110**	0.3877	0.0161	**0.0069**	3.5081	0.2091
$T=3000$	**0.076**	1.719	**0.705**	22.808	**0.0047**	0.3938	0.0147	**0.0103**	3.632	0.0731
Model (M2')										
	$E\lvert\hat{q}-q\rvert$		$E(D_H)$		$E\lVert\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rVert$			MSPE − 1		
Method	AMAR	Fused	AMAR	Fused	AMAR	Fused	AIC	AMAR	Fused	AIC
$T=400$	**0.305**	2.840	**2.090**	15.815	**0.0630**	1.3423	0.7972	**212.039**	149893.509	2791.427
$T=800$	**0.201**	3.606	**1.825**	16.436	**0.0456**	1.3424	0.8090	**424.532**	154788.182	982.968
$T=1500$	**0.346**	3.635	**2.845**	16.551	**0.0436**	1.3426	0.8113	**466.524**	163393.438	523.160
$T=3000$	**0.263**	3.520	**2.289**	16.342	**0.0425**	1.3423	0.8135	340.344	151713.973	**109.213**
Model (M3')										
	$E\lvert\hat{q}-q\rvert$		$E(D_H)$		$E\lVert\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rVert$			MSPE − 1		
Method	AMAR	Fused	AMAR	Fused	AMAR	Fused	AIC	AMAR	Fused	AIC
$T=400$	**0.602**	2.554	**1.391**	18.637	**0.0220**	0.3028	0.0608	**0.0312**	0.7465	17.5274
$T=800$	**0.209**	1.800	**0.598**	21.809	**0.0162**	0.3095	0.0339	**0.0281**	0.7149	2.8428
$T=1500$	**0.181**	1.871	**0.483**	21.899	**0.0118**	0.3119	0.0217	**0.0252**	0.7264	0.8557
$T=3000$	**0.122**	1.696	**0.261**	22.216	**0.0023**	0.3190	0.0145	**0.0016**	0.7399	0.3303
Model (M4')										
	$E\lvert\hat{q}-q\rvert$		$E(D_H)$		$E\lVert\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rVert$			MSPE − 1		
Method	AMAR	Fused	AMAR	Fused	AMAR	Fused	AIC	AMAR	Fused	AIC
$T=400$	**0.407**	2.579	**1.667**	18.931	**0.1287**	2.2217	0.8769	**0.1919**	219.844	28.1879
$T=800$	**0.231**	2.527	**1.245**	22.706	**0.2628**	2.2347	0.8750	**0.4009**	226.569	6.658
$T=1500$	**0.407**	2.543	**2.136**	22.955	**0.1763**	2.2378	0.8462	**0.2551**	224.782	2.7918
$T=3000$	**0.284**	2.568	**1.543**	23.337	**0.1077**	2.2389	0.8419	**0.1584**	215.987	1.3470

Table 4: Performance of different methods under (M1') – (M4'). Results from the best method are highlighted in bold. Recall that $\hat{q}$ is the number of the fitted timescales, D_H is the Hausdorff distance between the fitted timescale locations $\{\hat{\tau}_1, \ldots, \hat{\tau}_{\hat{q}}\}$ and the true ones $\{\tau_1, \ldots, \tau_q\}$, $\lVert\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\rVert$ is the Euclidean distance between the fitted parameter vector and the true one, and MPSE is the mean squared prediction errors of different fitted models.

B Proofs

B.1 Proof of Proposition 2.1

For AR(p) processes, it has a stationary and causal solution if and only if all the roots of $b(z) = 0$ lie outside $\mathbb{T}$. Given that $\alpha_1, \ldots, \alpha_q \geq 0$, it implies that $\beta_1, \ldots, \beta_p \geq 0$. Consequently, for the roots of $b(z) := 1 - \beta_1 z - \cdots - \beta_p z^p = 0$ to lie outside $\mathbb{T}$, one would necessarily require $b(1) > 0$, i.e. $\beta_1 + \cdots + \beta_p < 1$. This is because $b(z)$ is a monotonically decreasing function with respect to $z \in (0, \infty)$, so $b(1) \leq 0$ would have implied a root on or inside $\mathbb{T}$. On the other hand, if $\beta_1 + \cdots + \beta_p < 1$, then for any z lie on or inside $\mathbb{T}$, we have that $\|\beta_1 z + \cdots + \beta_p z^p\| \leq \beta_1 \|z\| + \cdots + \beta_p \|z^p\| \leq \beta_1 + \cdots + \beta_p < 1$, hence $b(z) \neq 0$ for all z lie on or inside $\mathbb{T}$, i.e. all roots of $b(z) = 0$ are outside $\mathbb{T}$. As such, we have shown that $\beta_1 + \cdots + \beta_p < 1$ is the necessary and sufficient condition. Finally, we note that this condition under the AMAR framework is equivalent to

$$\sum_{j=1}^{p} \Big(\sum_{k:\tau_k \geq j} \frac{\alpha_k}{\tau_k} \Big) = \frac{\alpha_1}{\tau_1} \tau_1 + \cdots + \frac{\alpha_q}{\tau_q} \tau_q < 1$$

in view of Equations (1) and (2) and (3), which simply boils down to $\alpha_1 + \cdots + \alpha_q < 1$.

□

B.2 Proof of Theorem 2.1

We write the AR(p) model as

$$\mathbf{Y}_t = \mathbf{B}\mathbf{Y}_{t-1} + \varepsilon_t \mathbf{u}, \quad t = 1, \ldots, T, \tag{17}$$

where $\mathbf{Y}_t = (X_t, X_{t-1}, ..., X_{t-p+1})'$, the matrix of the coefficients

$$\mathbf{B} = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_p \\ & \mathbf{I}_{p-1} & & \mathbf{0} \end{pmatrix} \tag{18}$$

and $\mathbf{u} = (1, 0, \ldots, 0)' \in \mathbb{R}^p$. We start with a few auxiliary results.

Lemma B.1 (Parseval's identity, Theorem 1.9 in Duoandikoetxea (2001)) *For any complex-valued sequence $\{f_k\}_{k\in\mathbb{Z}}$ such that $\sum_{k\in\mathbb{Z}} |f_k|^2 < \infty$, the following identity holds*

$$\sum_{k\in\mathbb{Z}} |f_k|^2 = \int_{\mathbb{T}} |f(z)|^2 dm(z), \tag{19}$$

where $f(z) = \sum_{k\in\mathbb{Z}} f_k z^k$, $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$, $dm(z) = \frac{d|z|}{2\pi}$.

Lemma B.2 (Cauchy's integral formula) *Let* $\mathbf{M} \in \mathbb{R}^{p\times p}$ *be a real- or complex- valued matrix. Then for any curve* Γ *enclosing all eigenvalues of* $\mathbf{M}$ *and any* $j \in \mathbb{N}$ *the following holds*

$$\mathbf{M}^j = \frac{1}{2\pi i}\int_\Gamma z^j (z\mathbf{I}_p - \mathbf{M})^{-1} dz = \frac{1}{2\pi i}\int_\Gamma z^{j-1}(\mathbf{I}_p - z^{-1}\mathbf{M})^{-1} dz. \tag{20}$$

Lemma B.3 *Let* $\mathbf{B}$ *given by* (18) *be the matrix of coefficients of a stationary AR*(p) *process and let* $\mathbf{v} = (v_1, \ldots, v_p)' \in \mathbb{R}^p$. *For all* $z \in \mathbb{C}$ *such that* $\sum_{i=0}^\infty |\langle \mathbf{v}, \mathbf{B}^i\mathbf{u}\rangle||z^i| < \infty$, *we have*

$$b(z)\sum_{i=0}^\infty \left\langle \mathbf{v}, \mathbf{B}^i\mathbf{u}\right\rangle z^i = b(z)\left\langle \mathbf{v}, (\mathbf{I}_p - z\mathbf{B})^{-1}\mathbf{u}\right\rangle = v(z), \tag{21}$$

where $v(z) = v_1 + v_2 z + \ldots + v_p z^{p-1}$, *and where* $b(z)$ *is the AR polynomial.*

Proof. As $\sum_{i=0}^\infty |\langle \mathbf{v}, \mathbf{B}^i\mathbf{u}\rangle||z^i| < 0$, we can change the order of summation in the left-hand side of (21)

$$(1 - \beta_1 z - \ldots - \beta_p z^p)\sum_{i=0}^\infty \left\langle \mathbf{v}, \mathbf{B}^i\mathbf{u}\right\rangle z^i = \left\langle \mathbf{v}, \left(\sum_{i=0}^\infty (1 - \beta_1 z - \ldots - \beta_p z^p) z^i \mathbf{B}^i\right)\mathbf{u}\right\rangle.$$

Define $\beta_0 = -1$, $\beta_k = 0$ for $k > p$. By direct algebraic manipulation,

$$\sum_{i=0}^\infty (1 - \beta_1 z - \ldots - \beta_p z^p) z^i \mathbf{B}^i = -\sum_{i=0}^\infty \left(\sum_{k=0}^i \beta_k \mathbf{B}^{i-k}\right) z^i := -\sum_{i=0}^\infty \mathbf{D}_i z^i.$$

The characteristic polynomial of $\mathbf{B}$ is given by $\phi(z) = \sum_{k=0}^p \beta_k z^{p-k}$. From the Cayley–Hamilton theorem, $\mathbf{B}$ is a root of ϕ, and, consequently for $i \geq p$,

$$\mathbf{D}_i = \mathbf{B}^{i-p}\sum_{k=0}^i \beta_k \mathbf{B}^{p-k} = \mathbf{B}^{i-p}\sum_{k=0}^p \beta_k \mathbf{B}^{p-k} = 0.$$

It remains to demonstrate that $\langle \mathbf{v}, \mathbf{D}_i\mathbf{u}\rangle = -v_{i+1}$ for $i = 0, \ldots, p-1$, which we show by induction. For $i = 0$, $\langle \mathbf{v}, \mathbf{D}_i\mathbf{u}\rangle = \beta_0 \langle \mathbf{v}, \mathbf{u}\rangle = -v_1$. When $i \geq 1$, matrices $\mathbf{D}_i$ satisfy $\mathbf{D}_i = \mathbf{B}\mathbf{D}_{i-1} + \beta_i \mathbf{I}_p$, therefore

$$\begin{aligned}\langle \mathbf{v}, \mathbf{D}_i \mathbf{u}\rangle &= \langle \mathbf{v}, \mathbf{B}\mathbf{D}_{i-1}\mathbf{u}\rangle + \beta_i \langle \mathbf{v}, \mathbf{u}\rangle = \langle \mathbf{B}'\mathbf{v}, \mathbf{D}_{i-1}\mathbf{u}\rangle + \beta_i \langle \mathbf{v}, \mathbf{u}\rangle \\ &= \langle v_1(\beta_1, \ldots, \beta_p)' + (0, v_2, \ldots, v_p)', \mathbf{D}_{i-1}\mathbf{u}\rangle + \beta_i \langle \mathbf{v}, \mathbf{u}\rangle = -v_1\beta_i - v_{i+1} + v_1\beta_i \\ &= -v_{i+1},\end{aligned}$$

which completes the proof. □

Lemma B.4 *Let $Z_1, Z_2, \ldots$ be a sequence of i.i.d. $\mathcal{N}(0,1)$ random variables. Then for any integers $l \neq 0$ and $k > 0$, the following exponential probability bound holds for any $x > 0$:*

$$\mathbb{P}\left(\left|\sum_{t=1}^{k} Z_t Z_{t+l}\right| > kx\right) \leq 2\exp\left(-\frac{1}{8}\frac{kx^2}{4+x}\right). \tag{22}$$

Proof. We will show that $\mathbb{P}\left(\sum_{t=1}^{k} Z_t Z_{t+l} > kx\right) \leq \exp\left(-\frac{1}{8}\frac{kx^2}{4+x}\right)$, which would then imply (22) by symmetry. By Markov's inequality, for any $x > 0$ and $\lambda > 0$, it holds that

$$\mathbb{P}\left(\sum_{t=1}^{k} Z_t Z_{t+l} > kx\right) \leq \exp(-kx\lambda)\,\mathrm{E}\exp\left(\lambda\sum_{t=1}^{k} Z_t Z_{t+l}\right).$$

By the convexity of $y \mapsto \exp(\lambda y)$ for any $\lambda > 0$, Theorem 1 in Vershynin (2011) implies

$$\mathrm{E}\exp\left(\lambda\sum_{t=1}^{k} Z_t Z_{t+l}\right) \leq \mathrm{E}\exp\left(4\lambda\sum_{t=1}^{k} Z_t \tilde{Z}_t\right),$$

where $\tilde{Z}_1, \ldots, \tilde{Z}_k$ are independent copies of $Z_1, \ldots, Z_k$. Using the independence and by direct computation (see also Craig (1936)), we get

$$\mathrm{E}\exp\left(4\lambda\sum_{t=1}^{k} Z_t \tilde{Z}_t\right) = \left(\mathrm{E}\exp\left(4\lambda Z_1 \tilde{Z}_1\right)\right)^k = \left(\mathrm{E}\exp\left(8\lambda^2 \tilde{Z}_1^2\right)\right)^k = \left(1 - 16\lambda^2\right)^{-\frac{1}{2}k}$$

provided that $0 < \lambda < \frac{1}{4}$, therefore $\mathbb{P}\left(\sum_{t=1}^{k} Z_t Z_{t+l} > kx\right) \leq \exp\left(-kx\lambda - \frac{k}{2}\log\left(1 - 16\lambda^2\right)\right)$. Taking $\lambda = \frac{-2+\sqrt{4+x^2}}{4x}$ minimises the right-hand side of this inequality. With this value of λ and using

$\log(x) \leq x - 1$, we have

$$\mathbb{P}\left(\sum_{t=1}^{k} Z_t Z_{t+l} > kx\right) \leq \exp\left(\frac{k}{4}\left(2 - \sqrt{x^2+4} + 2\log\left(\frac{1}{4}\left(\sqrt{x^2+4}+2\right)\right)\right)\right)$$
$$\leq \exp\left(\frac{k}{4}\left(2 - \sqrt{x^2+4} + \frac{1}{2}\left(\sqrt{x^2+4}+2\right) - 2\right)\right)$$
$$= \exp\left(\frac{k}{8}\left(2 - \sqrt{x^2+4}\right)\right) = \exp\left(-\frac{1}{8}\frac{kx^2}{2+\sqrt{x^2+4}}\right)$$
$$\leq \exp\left(-\frac{1}{8}\frac{kx^2}{4+x}\right),$$

which completes the proof. □

Lemma B.5 (Lemma 1 in Laurent and Massart (2000)) *Let $Z_1, Z_2, \ldots$ be a sequence of i.i.d. $\mathcal{N}(0,1)$ random variables. For any integer $k > 0$ and $x > 0$, the following exponential probability bounds hold*

$$\mathbb{P}\left(\sum_{t=1}^{k} Z_t^2 \geq k + 2\sqrt{kx} + 2x\right) \leq \exp\left(-x\right), \tag{23}$$
$$\mathbb{P}\left(\sum_{t=1}^{k} Z_t^2 \leq k - 2\sqrt{kx}\right) \leq \exp\left(-x\right). \tag{24}$$

Proof of Theorem 2.1. For $\mathbf{C}_T = \sum_{t=1}^{T-1} \mathbf{Y}_t \mathbf{Y}_t'$ and $\mathbf{A}_T = \sum_{t=1}^{T-1} \varepsilon_{t+1} \mathbf{Y}_t$, we have $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathbf{C}_T^{-1}\mathbf{A}_T$. Here the distribution of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ is invariant to the value of σ. As such, in the following, we assume $\sigma = 1$ for notational convenience. Consequently,

$$\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\| \leq \lambda_{max}(\mathbf{C}_T^{-1}) \left\|\mathbf{A}_T\right\| = \lambda_{min}^{-1}(\mathbf{C}_T) \left\|\mathbf{A}_T\right\|, \tag{25}$$

where $\lambda_{min}(\mathbf{M})$ and $\lambda_{max}(\mathbf{M})$ denote, respectively, the smallest and the largest eigenvalues of a symmetric matrix $\mathbf{M}$. To provide an upper bound on $\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|$ given in Theorem 2.1, we will bound $\lambda_{min}(\mathbf{C}_T)$ from below and $\|\mathbf{A}_T\|$ from above, working on a set whose probability is large.

In the calculations below, we will repeatedly use the following representation of $\mathbf{Y}_t$, which follows from applying (17) recursively:

$$\mathbf{Y}_t = \mathbf{B}^t\mathbf{Y}_0 + \sum_{j=1}^{t} \varepsilon_{t-j+1}\mathbf{B}^{j-1}\mathbf{u}, \quad t = 1, \ldots, T. \tag{26}$$

In addition, to improve the presentational aspect of the proof, here we shall take $\mathbf{Y}_0 = \mathbf{0}$. All the results would go through (with minor modifications to handle the extra terms) if one instead assumes that $\mathbf{Y}_0$ is a realization from a stationary solution.

In the arguments below, we will show result more specific than (9), i.e.

$$\|\mathbf{A}_T\| \leq \left(32\underline{b}^{-2}\sqrt{1+\|\boldsymbol{\beta}\|^2}\right) p\log(T)\sqrt{(1+\log(T+p))T}, \tag{27}$$

$$\lambda_{min}(\mathbf{C}_T) \geq \bar{b}^{-2}\left(T - p(1+32\log(T)\sqrt{T})\right), \tag{28}$$

on the event

$$\mathcal{E}_T = \mathcal{E}_T^{(1)} \cap \mathcal{E}_T^{(2)} \cap \mathcal{E}_T^{(3)}, \tag{29}$$

where

$$\mathcal{E}_T^{(1)} = \bigcap_{1\leq i<j\leq p} \left\{ \left| \sum_{t=1}^{T-\max(i,j)} \varepsilon_t\varepsilon_{t+|i-j|} \right| < 32\log(T)\sqrt{T-\max(i,j)} \right\},$$

$$\mathcal{E}_T^{(2)} = \bigcap_{j=1}^{T} \left\{ \left| \sum_{t=1}^{T-j} \varepsilon_t\varepsilon_{t+j} \right| < 32\log(T)\sqrt{T-j} \right\},$$

$$\mathcal{E}_T^{(3)} = \left\{ \sum_{t=1}^{T-p} \varepsilon_t^2 > T - p - 2\sqrt{\log(T)(T-p)} \right\}.$$

Finally, we will demonstrate that $\mathcal{E}_T$ satisfies

$$\mathbb{P}\left(\mathcal{E}_T\right) \geq 1 - \frac{5}{T}. \tag{30}$$

Thus, (25), (27), (28) and (30) combined together imply the statement of Theorem 2.1. The remaining part of the proof is split into three parts, in which we show (27), (28) and (30) in turn.

Upper bound for $\|\mathbf{A}_T\|$. The Euclidean norm satisfies $\|\mathbf{A}_T\| = \sup_{\mathbf{v}:\in\mathbb{R}^p, \|\mathbf{v}\|=1} |\langle \mathbf{v}, \mathbf{A}_T\rangle|$, therefore we consider inner products $\langle \mathbf{v}, \mathbf{A}_T \rangle$ where $\mathbf{v} \in \mathbb{R}^p$ is any unit vector. By (26),

$$\begin{aligned}\langle \mathbf{v}, \mathbf{A}_T \rangle &= \sum_{t=1}^{T-1} \langle \mathbf{v}, \mathbf{Y}_t \rangle \varepsilon_{t+1} = \sum_{t=1}^{T-1} \sum_{j=1}^{t} \left\langle \mathbf{v}, \mathbf{B}^{j-1}\mathbf{u} \right\rangle \varepsilon_{t-j+1}\varepsilon_{t+1} \\ &= \sum_{j=1}^{T-1} \left\langle \mathbf{v}, \mathbf{B}^{j-1}\mathbf{u} \right\rangle a_j,\end{aligned}$$

where $a_j = \sum_{t=j}^{T-1} \varepsilon_{t-j+1}\varepsilon_{t+1} = \sum_{t=1}^{T-j} \varepsilon_t \varepsilon_{t+j}$.
Lemma B.2 and Lemma B.3 applied to the right-hand side of the above equation yield

$$\begin{aligned}\sum_{j=1}^{T-1} \left\langle \mathbf{v}, \mathbf{B}^{j-1}\mathbf{u} \right\rangle a_j &= \frac{1}{2\pi i} \int_{\mathbb{T}} \left(\sum_{j=1}^{T-1} z^{j-1} a_j \right) \left\langle \mathbf{v}, (z\mathbf{I}_p - \mathbf{B})^{-1}\mathbf{u} \right\rangle dz \\ &= \frac{1}{2\pi i} \int_{\mathbb{T}} \left(\sum_{j=1}^{T-1} z^{j-1} a_j \right) \left(\sum_{j=1}^{p} z^{p-j} v_j \right) q(z) dz \\ &= \frac{1}{2\pi i} \int_{\mathbb{T}} \left(\sum_{j=0}^{T+p-1} z^j c_j \right) q(z) dz,\end{aligned}$$

where $q(z) = (z^p b(z^{-1}))^{-1}$ and $c_j = \sum_{i=0}^{j} a_{i+1} v_{p-j+i}$. Integrating by parts, we get

$$\frac{1}{2\pi i} \int_{\mathbb{T}} \left(\sum_{j=0}^{T+p-1} z^j c_j \right) q(z) dz = -\frac{1}{2\pi i} \int_{\mathbb{T}} \left(\sum_{j=0}^{T+p-1} z^{j+1} \frac{c_j}{j+1} \right) q'(z) dz,$$

where $q'(\cdot)$ is the derivative of $q(\cdot)$. Combining the calculations above and using the fact that $\mathbb{T} = \{z \in \mathbb{C} : \ |z| = 1\}$, Cauchy's inequality and Lemma B.1, we obtain

$$\left| \sum_{j=1}^{T-1} \left\langle \mathbf{v}, \mathbf{B}^{j-1}\mathbf{u} \right\rangle a_j \right| \leq \sqrt{\sum_{j=0}^{T+p-1} \left(\frac{c_j}{j+1} \right)^2} \sqrt{\int_{\mathbb{T}} |q'(z)|^2 dm(z)}, \tag{31}$$

where we recall that $dm(z) = \frac{d|z|}{2\pi}$. To further bound the first term on the right-hand side of (31), we recall that on the event $\mathcal{E}_T$ coefficients $|a_j| \leq 32\log(T)\sqrt{T}$, hence

$$\begin{aligned}
\sqrt{\sum_{j=0}^{T+p-1}\left(\frac{c_j}{j+1}\right)^2} &= \sqrt{\sum_{j=0}^{T+p-1}\frac{1}{(j+1)^2}\left(\sum_{i=0}^{j} a_{i+1}v_{p-j+i}\right)^2} \\
&\leq \max_{j=0,\dots,T+p-1}|a_j|\sqrt{\sum_{j=0}^{T+p-1}\frac{1}{(j+1)^2}\left(\sum_{i=0}^{j}|v_{p-j+i}|\right)^2} \\
&\leq 32\log(T)\sqrt{T}\sqrt{\sum_{j=0}^{T+p-1}\frac{j+1}{(j+1)^2}} \\
&\leq 32\log(T)\sqrt{(1+\log(T+p))T}.
\end{aligned}$$

For the second term in (31), we calculate the derivative

$$q'(z) = -\frac{pz^{p-1} - \sum_{j=1}^{p}(p-j)\beta_j z^{p-j-1}}{(z^p b(z^{-p}))^2}$$

and use Lemma B.1 to bound

$$\begin{aligned}
\sqrt{\int_{\mathbb{T}}|q'(z)|^2 dm(z)} &= \sqrt{\int_{\mathbb{T}}\left|\frac{pz^{p-1} - \sum_{j=1}^{p}(p-j)\beta_j z^{p-j}}{(z^p b(z^{-p}))^2}\right|^2 dm(z)} \\
&\leq \frac{\sqrt{\int_{\mathbb{T}}\left|pz^{p-1} - \sum_{j=1}^{p}(p-j)\beta_j z^{p-j}\right|^2 dm(z)}}{\min_{|z|=1}|(z^p b(z^{-p}))|^2} \\
&= \underline{b}^{-2}\sqrt{\left(p^2 + \sum_{j=1}^{p}(p-j)^2\beta_j^2\right)} \leq \underline{b}^{-2}p\sqrt{1+\|\boldsymbol{\beta}\|^2}.
\end{aligned}$$

Combining the bounds on the two terms, we obtain

$$\sum_{j=1}^{T-1}\left\langle \mathbf{v}, \mathbf{B}^{j-1}\mathbf{u}\right\rangle a_j \leq \left(32\underline{b}^{-2}\sqrt{1+\|\boldsymbol{\beta}\|^2}\right) p\log(T)\sqrt{(1+\log(T+p))T}.$$

Taking supremum over $\mathbf{v} \in \mathbb{R}^p$ such that $\|\mathbf{v}\| = 1$ proves (27).

Lower bound for $\lambda_{min}(\mathbf{C}_T)$. Let $\mathbf{v} = (v_1, \dots, v_p)'$ be a unit vector in $\mathbb{R}^p$. We begin the proof by establishing the following inequality

$$\langle \mathbf{v}, \mathbf{C}_T \mathbf{v} \rangle \geq \bar{b}^{-2} \sum_{i,j=1}^{p} v_i v_j \sum_{t=1}^{T-1} \varepsilon_{t-j+1} \varepsilon_{t-i+1}, \tag{32}$$

where $\varepsilon_t = 0$ for $t \leq 0$ and $\bar{b} = \max_{z \in \mathbb{T}} |b(z)|$. By Lemma B.1 and (26), we rewrite the quadratic form on the left-hand side of (32) to

$$\begin{aligned} \langle \mathbf{v}, \mathbf{C}_T \mathbf{v} \rangle &= \sum_{t=1}^{T-1} \langle \mathbf{v}, \mathbf{Y}_t \rangle^2 && (33) \\ &= \int_{\mathbb{T}} \left| \sum_{t=1}^{T-1} \left\langle \mathbf{v}, \sum_{j=1}^{t} \varepsilon_j \mathbf{B}^{t-j} \mathbf{u} \right\rangle z^t \right|^2 dm(z) && (34) \\ &= \int_{\mathbb{T}} \left| \sum_{t=1}^{T-1} \sum_{j=1}^{T-1} \varepsilon_j \omega_{t-j} z^t \right|^2 dm(z) && (35) \end{aligned}$$

where $\omega_j = \langle \mathbf{v}, \mathbf{B}^j \mathbf{u} \rangle$ for $j \geq 0$, $\omega_j = 0$ for $j < 0$. Changing the order of summation and by a simple substitution we get

$$\sum_{t=1}^{T-1} \sum_{j=1}^{T-1} \varepsilon_j \omega_{t-j} z^t = \sum_{j=1}^{T-1} \varepsilon_j z^j \sum_{t=1}^{T-1} \omega_{t-j} z^{t-j} = \sum_{j=1}^{T-1} \varepsilon_j z^j \sum_{t=0}^{T-j-1} \omega_t z^t. \tag{36}$$

Using the definition of ω_j, the fact that all eigenvalues of $\mathbf{B}$ have modulus strictly lower than one and Lemma B.3, (36) simplifies to

$$\begin{aligned} \sum_{j=1}^{T-1} \varepsilon_j z^j \sum_{t=0}^{T-j-1} \omega_t z^t &= \sum_{j=1}^{T-1} \varepsilon_j z^j \left\langle \mathbf{v}, (\mathbf{I}_p - (\mathbf{B}z)^{T-j})(\mathbf{I}_p - \mathbf{B}z)^{-1} \mathbf{u} \right\rangle \\ &= \sum_{j=1}^{T-1} \varepsilon_j \left(z^j \left\langle \mathbf{v}, (\mathbf{I}_p - \mathbf{B}z)^{-1} \mathbf{u} \right\rangle - z^T \left\langle \mathbf{B}^{T-j} \mathbf{v}, (\mathbf{I}_p - \mathbf{B}z)^{-1} \mathbf{u} \right\rangle \right) \\ &= b(z)^{-1} \sum_{j=1}^{T-1} \varepsilon_j \left(z^j v(z) - z^T w_j(z) \right), \end{aligned}$$

where $v(z) = \sum_{k=1}^{p} v_k z_{k-1}$ and $w_j(z) = \sum_{k=1}^{p} (\mathbf{B}^{T-j} v)_k z^{k-1}$ for $j = 0, \ldots, T-1$. The equation above, (33) and (36) combined together imply the following inequality

$$\begin{aligned}\langle \mathbf{v}, \mathbf{C}_T \mathbf{v}\rangle &= \int_{\mathbb{T}} \left| b(z)^{-1} \sum_{j=1}^{T-1} \varepsilon_j \left(z^j v(z) - z^T w_j(z)\right)\right|^2 dm(z) \\ &\geq \overline{b}^{-2} \int_{\mathbb{T}} \left| \sum_{j=1}^{T-1} \varepsilon_j \left(z^j v(z) - z^T w_j(z)\right)\right|^2 dm(z).\end{aligned}$$

Observe that $\sum_{j=1}^{T-1} \varepsilon_j \left(z^j v(z) - z^T w_j(z)\right) = \sum_{j=1}^{T-1} \varepsilon_j \left(z^j v(z) - z^T w_j(z)\right) = \sum_{t=1}^{T+p-1} c_t z^t$ is a trigonometric polynomial, therefore by Lemma B.1 and simple algebra

$$\begin{aligned}\int_{\mathbb{T}} \left| \sum_{j=1}^{T-1} \varepsilon_j \left(z^j v(z) - z^T w_j(z)\right)\right| dm(z) &= \sum_{t=1}^{T+p-1} |c_t|^2 \geq \sum_{t=1}^{T-1} |c_t|^2 = \sum_{t=1}^{T-1} \left(\sum_{j=1}^{p} v_j \varepsilon_{t-j+1}\right)^2 = \\ &= \sum_{i,j=1}^{p} v_j v_i \sum_{t=1}^{T-1} \varepsilon_{t-j+1}\varepsilon_{t-i+1},\end{aligned}$$

which proves (32).

We are now in a position to bound $\langle \mathbf{v}, \mathbf{C}_T \mathbf{v}\rangle$ from below. Rearranging terms in (32) yields

$$\begin{aligned}\langle \mathbf{v}, \mathbf{C}_T \mathbf{v}\rangle &\geq \overline{b}^{-2} \left(\sum_{i=1}^{p} v_i^2 \sum_{t=1}^{n-i} \varepsilon_t^2 + \sum_{1\leq i<j\leq p} v_i v_j \sum_{t=1}^{T-\max(i,j)} \varepsilon_t \varepsilon_{t+|j-i|}\right) \\ &\geq \overline{b}^{-2} \left(\sum_{t=1}^{T-p} \varepsilon_t^2 \sum_{i=1}^{p} v_i^2 - \max_{1\leq i<j\leq p} \left| \sum_{t=1}^{T-\max(i,j)} \varepsilon_t \varepsilon_{t+|j-i|}\right| \left(\left(\sum_{i=1}^{p} |v_i| \right)^2 - \sum_{i=1}^{p} v_i^2 \right)\right) \\ &\geq \overline{b}^{-2} \left(\sum_{t=1}^{T-p} \varepsilon_t^2 - (p-1) \max_{1\leq i<j\leq p} \left| \sum_{t=1}^{T-\max(i,j)} \varepsilon_t \varepsilon_{t+|j-i|}\right| \right).\end{aligned}$$

Recalling the definition of $\mathcal{E}_T$, we conclude that on this event

$$\begin{aligned}\langle \mathbf{v}, \mathbf{C}_T \mathbf{v}\rangle &\geq \overline{b}^{-2} \left(T - p - 2\sqrt{\log(T)(T-p)} - (p-1)32\log(T)\sqrt{T}\right) \\ &\geq \overline{b}^{-2} \left(T - p(1 + 32\log(T)\sqrt{T})\right).\end{aligned}$$

Taking infimum over $\mathbf{v} \in \mathbb{R}^p$ such that $\|\mathbf{v}\| = 1$ in the inequality above proves (28).

Lower bound for $\mathbb{P}(\mathcal{E}_T)$. Recalling (29) and using a simple Bonferroni bound, we get

$$
\begin{aligned}
\mathbb{P}(\mathcal{E}_T^c) \le\ & p^2 \max_{1\le i<j\le p} \mathbb{P}\left(\left| \sum_{t=1}^{T-\max(i,j)} \varepsilon_t \varepsilon_{t+|i-j|} \right| \ge 32\log(T)\sqrt{T-\max(i,j)} \right) \\
& + T \max_{1\le j\le T} \mathbb{P}\left(\left| \sum_{t=1}^{T-j} \varepsilon_t \varepsilon_{t+j} \right| < 32\log(T)\sqrt{T-j} \right) \\
& + \mathbb{P}\left(\sum_{t=1}^{T-p} \varepsilon_t^2 > T-p-2\sqrt{\log(T)(T-p)} \right) \\
& := p^2 \max_{1\le i<j\le p} P_{i,j}^{(1)} + T \max_{1\le j\le T} P_j^{(2)} + P^{(3)}.
\end{aligned}
$$

Lemma B.4 implies that

$$
\begin{aligned}
P_{i,j}^{(1)} &\le 2\exp\left(-\frac{1}{8} \frac{(32\log(T))^2}{4+(\sqrt{T-\max(i,j)})^{-1}32\log(T)} \right) \le 2\exp\left(-2\log(T)\right) = \frac{2}{T^2}, \\
P_j^{(2)} &\le 2\exp\left(-\frac{1}{8} \frac{(32\log(T))^2}{4+(\sqrt{T-j})^{-1}32\log(T)} \right) \le 2\exp\left(-2\log(T)\right) = \frac{2}{T^2}.
\end{aligned}
$$

Moreover, by Lemma B.5, $P^{(3)} \le \exp\left(-\log(T)\right) = \frac{1}{T}$, hence, given that $p^2 < T$, we have $\mathbb{P}(\mathcal{E}_T^c) \le \frac{5}{T}$, which completes the proof. □

B.3 Proof of Theorem 2.2

In the proof below, we shall focus on the case where F_T^M consists of randomly drawn intervals (which is what Algorithm 2 does when p is large). For the case where all sub-intervals of $[1,p]$ are used, the same arguments would go through, because Algorithm 2 then produces a larger set F_T^M compared to the approach of random drawing.

We now split the proof into four steps.

Step 1. Consider the event $\left\{ \left\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\| \le \kappa_1 (\underline{b}/\overline{b})^2 \left\| \boldsymbol{\beta} \right\| \frac{p\log(T)\sqrt{\log(T+p)}}{\sqrt{T}-\kappa_2 p\log(T)} \right\}$ where κ_1, κ_2 are as in Theorem 2.1. Assumption (A3) implies that $\underline{b}/\overline{b}$ and $\|\boldsymbol{\beta}\|$ are bounded from above by constants. Furthermore, by Assumption (A2), $p \le c_1 T^{\theta}$, which implies that

$$\kappa_1 (\underline{b}/\bar{b})^2 \left\|\boldsymbol{\beta}\right\| \frac{p \log(T)\sqrt{\log(T+p)}}{\sqrt{T} - \kappa_2 p \log(T)} \le c_3 T^{\theta - 1/2} (\log(T))^{3/2} = c_3 \underline{\lambda}_T =: \lambda_T \tag{37}$$

for some constant $c_3 > 0$ and a sufficiently large T. Define now

$$A_T = \left\{ \left\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\| \le \lambda_T \right\} \tag{38}$$

By Theorem 2.1,

$$\mathbb{P}(A_T) \ge \mathbb{P}\left(\left\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\| \le \kappa_1 (\underline{b}/\bar{b})^2 \left\|\boldsymbol{\beta}\right\| \frac{p \log(T)\sqrt{\log(T+p)}}{\sqrt{T} - \kappa_2 p \log(T)} \right) \ge 1 - \kappa_3 T^{-1}, \tag{39}$$

for some constant $\kappa_3 > 0$.

Step 2. For $j = 1, \ldots, q$, define the intervals

$$\mathcal{I}_j^L = (\tau_j - \delta_T/3, \tau_j - \delta_T/6) \tag{40}$$

$$\mathcal{I}_j^R = (\tau_j + \delta_T/6, \tau_j + \delta_T/3) \tag{41}$$

Recall that F_T^M is the set of M randomly drawn intervals with endpoints in $\{1, \ldots, p\}$. Denote by $[s_1, e_1], \ldots, [s_M, e_M]$ the elements of F_T^M and let

$$D_T^M = \left\{ \forall j = 1, \ldots, q, \exists k \in \{1, \ldots, M\}, \text{ s.t. } s_k \times e_k \in \mathcal{I}_j^L \times \mathcal{I}_j^R \right\}. \tag{42}$$

We have that

$$\begin{aligned}
\mathbb{P}\left((D_T^M)^c\right) &\le \sum_{j=1}^{q} \Pi_{m=1}^{M} \left(1 - \mathbb{P}\left(s_m \times e_m \in \mathcal{I}_j^L \times \mathcal{I}_j^R\right)\right) \\
&\le q \left(1 - \frac{\delta_T^2}{6^2 p^2}\right)^M \le \frac{p}{\delta_T} \left(1 - \frac{\delta_T^2}{36 p^2}\right)^M.
\end{aligned}$$

Therefore, $\mathbb{P}\left(A_T \cap D_T^M\right) \ge 1 - \kappa_3 T^{-1} - p\delta_T^{-1}(1 - \delta_T^2 p^{-2}/36)^M \to 1$. In the remainder of the proof, assume that A_T and D_T^M all hold.

Note that Assumption (A4) implies that there exists $\underline{c} > 0$ such that $\delta_T^{1/2} \underline{\alpha}_T > \underline{c} \lambda_T$ for all sufficiently large T. We are now in the position to specify the constants explicitly as

$$C_1 = 2\sqrt{C_3} + c_3, \quad C_2 = \frac{1}{\sqrt{6}} - \frac{1}{\underline{c}}, \quad C_3 = (4\sqrt{2}+6)c_3^2,$$

where c_3 is in Equation (37).

Step 3. We focus on a generic interval $[s, e]$ such that

$$\exists j \in \{1, \dots, q\},\ \exists k \in \{1, \dots, M\},\ \text{s.t. } [s_k, e_k] \subset [s, e] \text{ and } s_k \times e_k \in \mathcal{I}_j^L \times \mathcal{I}_j^R. \tag{43}$$

Fix such an interval $[s, e]$ and let $j \in \{1, \dots, q\}$ and $k \in \{1, \dots, M\}$ be such that (43) is satisfied. Let $b_k^* = \operatorname{argmax}_{s_k \le b \le e_k} \mathcal{C}_{s_k,e_k}^b \left(\hat{\boldsymbol{\beta}}\right)$. By construction, $[s_k, e_k]$ satisfies $\tau_j - s_k + 1 \ge \delta_T/6$ and $e_k - \tau_j > \delta_T/6$. Let

$$\begin{aligned}
\mathcal{M}_{s,e} &= \left\{m : [s_m, e_m] \in F_T^M, [s_m, e_m] \subset [s, e]\right\}, \\
\mathcal{O}_{s,e} &= \{m \in \mathcal{M}_{s,e} : \max_{s_m \le b < e_m} \mathcal{C}_{s_m,e_m}^b \left(\hat{\boldsymbol{\beta}}\right) > \zeta_T\}.
\end{aligned}$$

Our first aim is to show that $\mathcal{O}_{s,e}$ is non-empty. This follows from Lemma 2 in Baranowski et al. (2019), the Cauchy–Schwarz inequality, and the calculation below, as

$$\begin{aligned}
\mathcal{C}_{s_k,e_k}^{b_k^*} \left(\hat{\boldsymbol{\beta}}\right) &\ge \mathcal{C}_{s_k,e_k}^{\tau_j} \left(\hat{\boldsymbol{\beta}}\right) \\
&\ge \mathcal{C}_{s_k,e_k}^{b_k^*} \left(\boldsymbol{\beta}\right) - \lambda_T \ge \left(\frac{\delta_T}{6}\right)^{1/2} |\alpha_j \tau_j^{-1}| - \lambda_T \ge \left(\frac{\delta_T}{6}\right)^{1/2} \underline{\alpha}_T - \lambda_T \\
&= \left(\frac{1}{\sqrt{6}} - \frac{\lambda_T}{\delta_T^{1/2}\underline{\alpha}_T}\right) \delta_T^{1/2} \underline{\alpha}_T \ge \left(\frac{1}{\sqrt{6}} - \frac{1}{\underline{c}}\right) \delta_T^{1/2} \underline{\alpha}_T = C_2 \delta_T^{1/2} \underline{\alpha}_T > \zeta_T.
\end{aligned}$$

Let $m^* = \operatorname{argmin}_{m \in \mathcal{O}_{s,e}} (e_m - s_m + 1)$ and $b^* = \operatorname{argmax}_{s_{m^*} \le b < e_{m^*}} \mathcal{C}_{s_{m^*},e_{m^*}}^b \left(\hat{\boldsymbol{\beta}}\right)$. Observe that $[s_{m^*}, e_{m^*})$ must contain at least one change in $\hat{\boldsymbol{\beta}}$. Indeed, if this were not the case, we would have $\mathcal{C}_{s_{m^*},e_{m^*}}^b \left(\boldsymbol{\beta}\right) = 0$ and

$$\mathcal{C}_{s_{m^*},e_{m^*}}^{b^*} \left(\hat{\boldsymbol{\beta}}\right) = |\mathcal{C}_{s_{m^*},e_{m^*}}^{b^*} \left(\hat{\boldsymbol{\beta}}\right) - \mathcal{C}_{s_{m^*},e_{m^*}}^{b^*} \left(\boldsymbol{\beta}\right)| \le \lambda_T < \frac{C_1}{c_3} \lambda_T = C_1 \underline{\lambda}_T \le \zeta_T,$$

which contradicted $\mathcal{C}_{s_{m^*},e_{m^*}}^{b^*} \left(\hat{\boldsymbol{\beta}}\right) > \zeta_T$. On the other hand, $[s_{m^*}, e_{m^*})$ cannot contain more than one change-points, because $e_{m^*} - s_{m^*} + 1 \le e_k - s_k + 1 \le \delta_T$.

Without loss of generality, assume $\tau_j \in [s_{m^*}, e_{m^*})$. Let $\eta_L = \tau_j - s_{m^*} + 1$, $\eta_R = e_{m^*} - \tau_j$ and $\eta_T = (C_1/c_3 - 1)^2 \alpha_j^2 \tau_j^{-2} \lambda_T^2$. We claim that $\min(\eta_L, \eta_R) > \eta_T$, because otherwise $\min(\eta_L, \eta_R) \leq \eta_T$ and Lemma 2 in Baranowski et al. (2019) would have implied

$$\begin{aligned} \mathcal{C}^{b^*}_{s_{m^*},e_{m^*}}\left(\hat{\boldsymbol{\beta}}\right) &\leq \mathcal{C}^{b^*}_{s_{m^*},e_{m^*}}(\boldsymbol{\beta}) + \lambda_T \leq \mathcal{C}^{\tau_j}_{s_{m^*},e_{m^*}}(\boldsymbol{\beta}) + \lambda_T \leq \eta_T^{1/2}|\alpha_j \tau_j^{-1}| + \lambda_T \\ &= (C_1/c_3 - 1 + 1)\lambda_T = C_1 \underline{\lambda}_T < \zeta_T, \end{aligned}$$

which contradicted $\mathcal{C}^{b^*}_{s_{m^*},e_{m^*}}\left(\hat{\boldsymbol{\beta}}\right) > \zeta_T$.

We are now in the position to prove $|b^* - \tau_j| \leq C_3 \underline{\lambda}_T \underline{\alpha}_T^{-2}$. Our aim is to find ϵ_T such that for any $b \in \{s_{m^*}, s_{m^*}+1, \ldots, e_{m^*}-1\}$ with $|b - \tau_j| > \epsilon_T$, we always have

$$\left\{\mathcal{C}^{\tau_j}_{s_{m^*},e_{m^*}}\left(\hat{\boldsymbol{\beta}}\right)\right\}^2 - \left\{\mathcal{C}^{b}_{s_{m^*},e_{m^*}}\left(\hat{\boldsymbol{\beta}}\right)\right\}^2 > 0. \tag{44}$$

This would then imply that $|b^* - \tau_j| \leq \epsilon_T$. By expansion and rearranging the terms, we see that (44) is equivalent to

$$\begin{aligned} &\langle \boldsymbol{\beta}, \boldsymbol{\psi}^{\tau_j}_{s_{m^*},e_{m^*}}\rangle^2 - \langle \boldsymbol{\beta}, \boldsymbol{\psi}^{b}_{s_{m^*},e_{m^*}}\rangle^2 > \langle \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \boldsymbol{\psi}^{b}_{s_{m^*},e_{m^*}}\rangle^2 - \langle \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \boldsymbol{\psi}^{\tau_j}_{s_{m^*},e_{m^*}}\rangle^2 \\ &\quad + 2\Big\langle \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \boldsymbol{\psi}^{b}_{s_{m^*},e_{m^*}}\langle \boldsymbol{\beta}, \boldsymbol{\psi}^{b}_{s_{m^*},e_{m^*}}\rangle - \boldsymbol{\psi}^{\tau_j}_{s_{m^*},e_{m^*}}\langle \boldsymbol{\beta}, \boldsymbol{\psi}^{\tau_j}_{s_{m^*},e_{m^*}}\rangle\Big\rangle. \end{aligned} \tag{45}$$

Here $\boldsymbol{\psi}^b_{s,e}$ (with $1 \leq s < b < e \leq p$) is a p-dimensional vector, with its s-th to b-th component being $\sqrt{\frac{e-b}{(e-s+1)()b-s+1}}$, its $b+1$-th to e-th component being $\sqrt{\frac{b-s+1}{(e-s+1)(e-b)}}$, and the remaining elements being 0. In the following, we assume that $b \geq \tau_j$. The case that $b < \tau_j$ can be handled in a similar fashion. By Lemma 4 in Baranowski et al. (2019), we have

$$\begin{aligned} \langle \boldsymbol{\beta}, \boldsymbol{\psi}^{\tau_j}_{s_{m^*},e_{m^*}}\rangle^2 - \langle \boldsymbol{\beta}, \boldsymbol{\psi}^{b}_{s_{m^*},e_{m^*}}\rangle^2 &= (\mathcal{C}^{\tau_j}_{s^*,e^*}(\boldsymbol{\beta}))^2 - (\mathcal{C}^{b}_{s_{m^*},e_{m^*}}(\boldsymbol{\beta}))^2 \\ &= \frac{|b-\tau_j|\eta_L}{|b-\tau_j| + \eta_L}(\alpha_j \tau_j^{-1})^2 =: \kappa. \end{aligned}$$

In addition, since we assume event A_T,

$$\begin{aligned} &\langle \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \boldsymbol{\psi}^{b}_{s_{m^*},e_{m^*}}\rangle^2 - \langle \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \boldsymbol{\psi}^{\tau_j}_{s_{m^*},e_{m^*}}\rangle^2 \leq \lambda_T^2, \\ &2\Big\langle \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \boldsymbol{\psi}^{b}_{s_{m^*},e_{m^*}}\langle \boldsymbol{\beta}, \boldsymbol{\psi}^{b}_{s_{m^*},e_{m^*}}\rangle - \boldsymbol{\psi}^{\tau_j}_{s_{m^*},e_{m^*}}\langle \boldsymbol{\beta}, \boldsymbol{\psi}^{\tau_j}_{s_{m^*},e_{m^*}}\rangle\Big\rangle \\ &\qquad \leq 2\|\boldsymbol{\psi}^{b}_{s_{m^*},e_{m^*}}\langle \boldsymbol{\beta}, \boldsymbol{\psi}^{b}_{s_{m^*},e_{m^*}}\rangle - \boldsymbol{\psi}^{\tau_j}_{s_{m^*},e_{m^*}}\langle \boldsymbol{\beta}, \boldsymbol{\psi}^{\tau_j}_{s_{m^*},e_{m^*}}\rangle\|_2 \lambda_T = 2\kappa^{1/2}\lambda_T, \end{aligned}$$

where the final equality is also implied by Lemma 4 in Baranowski et al. (2019). Consequently, (45) can be deducted from the stronger inequality $\kappa - 2\lambda_T \kappa^{1/2} - \lambda_T^2 > 0$. This quadratic inequality is implied by $\kappa > (\sqrt{2}+1)^2\lambda_T^2$, and could be restricted further to

$$\frac{2|b-\tau_j|\eta_L}{|b-\tau_j|+\eta_L} \geq \min(|b-\tau_j|, \eta_L) > (4\sqrt{2}+6)(\alpha_j\tau_j^{-1})^{-2}\lambda_T^2 = C_3(\alpha_j\tau_j^{-1})^{-2}\underline{\lambda}_T^2. \tag{46}$$

But since

$$\eta_L \geq \eta_T = (C_1/c_3-1)^2(\alpha_j\tau_j^{-1})^{-2}\lambda_T^2 = (2\sqrt{C_3}/c_3)^2(\alpha_j\tau_j^{-1})^{-2}\lambda_T^2 > C_3(\alpha_j\tau_j^{-1})^{-2}\underline{\lambda}_T^2,$$

we see that (46) is implied by $|b-\tau_j| > C_3(\alpha_j\tau_j^{-1})^{-2}\underline{\lambda}_T^2$. To sum up, $|b^*-\tau_j| > C_3(\alpha_j\tau_j^{-1})^{-2}\underline{\lambda}_T^2$ would result in (44), a contradiction. So we have proved that $|b^*-\tau_j| \leq C_3(\alpha_j\tau_j^{-1})^{-2}\underline{\lambda}_T^2$.

Step 4. With the arguments above valid on the event $A_T \cap B_T \cap D_T^M$, we can now proceed with the proof of the theorem. At the start of Algorithm 1, we have $s=1$ and $e=p$ and, provided that $q \geq 1$, condition (43) is satisfied. Therefore the algorithm detects a change-point b^* in that interval such that $|b^*-\tau_j| \leq C_3(\alpha_j\tau_j^{-1})^{-2}\underline{\lambda}_T^2$. By construction, we also have that $|b^*-\tau_j| < 2/3\delta_T$. This in turn implies that for all $l=1,\ldots,q$ such that $\tau_l \in [s,e]$ and $l \neq j$ we have either $\mathcal{I}_l^L, \mathcal{I}_l^R \subset [s,b^*]$ or $\mathcal{I}_l^L, \mathcal{I}_l^R \subset [b^*+1,e]$. Therefore (43) is satisfied within each segment containing at least one change-point. Note that before all q change points are detected, each change point will not be detected twice. To see this, we suppose that τ_j has already been detected by b, then for all intervals $[s_k,e_k] \subset [\tau_j - C_3(\alpha_j\tau_j^{-1})^{-2}\underline{\lambda}_T^2+1, \tau_j - C_3(\alpha_j\tau_j^{-1})^{-2}\underline{\lambda}_T^2+2/3\delta_T+1] \cup [\tau_j + C_3(\alpha_j\tau_j^{-1})^{-2}\underline{\lambda}_T^2 - 2/3\delta_T, \tau_j + C_3(\alpha_j\tau_j^{-1})^{-2}\underline{\lambda}_T^2]$, Lemma 2 in Baranowski et al. (2019), together with the definition of A_T, guarantee that

$$\begin{aligned}\max_{s_k \leq b < e} \mathcal{C}_{s_k,e_k}^b\left(\hat{\boldsymbol{\beta}}\right) &\leq \max_{s \leq b < e} \mathcal{C}_{s_k,e_k}^b(\boldsymbol{\beta}) + \lambda_T \\ &\leq \sqrt{C_3(\alpha_j\tau_j^{-1})^{-2}\underline{\lambda}_T^2}\alpha_j\tau_j^{-1} + \sqrt{C_3(\alpha_{j+1}\tau_{j+1}^{-1})^{-2}\underline{\lambda}_T^2}\alpha_{j+1}\tau_{j+1}^{-1} + \lambda_T \\ &< (2\sqrt{C_3}/c_3+1)\lambda_T = C_1\underline{\lambda}_T < \zeta_T.\end{aligned}$$

Once all the change-points have been detected, we then only need to consider $[s_k,e_k]$ such that

$$[s_k,e_k] \subset [\tau_j - C_3(\alpha_j\tau_j^{-1})^{-2}\underline{\lambda}_T^2+1, \tau_{j+1} + C_3(\alpha_{j+1}\tau_{j+1}^{-1})^{-2}\underline{\lambda}_T^2]$$

for $j=1,\ldots,q$. For such intervals, we have, by Lemmas 2 and 3 of Baranowski et al. (2019)

$$\begin{aligned}\max_{s_k \leq b < e_k} \mathcal{C}_{s_k,e_k}^b\left(\hat{\boldsymbol{\beta}}\right) &\leq \max_{s \leq b < e} \mathcal{C}_{s_k,e_k}^b(\boldsymbol{\beta}) + \lambda_T \\ &\leq \sqrt{C_3(\alpha_j\tau_j^{-1})^{-2}\underline{\lambda}_T^2}\alpha_j\tau_j^{-1} + \sqrt{C_3(\alpha_{j+1}\tau_{j+1}^{-1})^{-2}\underline{\lambda}_T^2}\alpha_{j+1}\tau_{j+1}^{-1} + \lambda_T \leq C_1\underline{\lambda}_T < \zeta_T.\end{aligned}$$

Hence no further scales is detected and the algorithm terminates. □

References

R. Baranowski, Y. Chen, and P. Fryzlewicz. Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society Series B*, 81:649–672, 2019.

C.C. Craig. On the frequency function of xy. *The Annals of Mathematical Statistics*, 7(1):1–15, 1936.

J. Duoandikoetxea. *Fourier Analysis*, volume 29 of *Graduate Studies in Mathematics*. American Mathematical Society, 2001.

B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28:1302–1338, 2000.

R. Vershynin. A simple decoupling inequality in probability theory. Technical report, University of Michigan, 2011. URL `https://www.math.uci.edu/~rvershyn/papers/decoupling-simple.pdf`.