SCIENCE CHINA
Mathematics

• ARTICLES •

https://doi.org/10.1007/s11425-023-xxxx-x

A robust and efficient tail-adaptive approach for change-point detection in semiparametric models

Wangyong Chen[1], Yewen Chen[2], Bingyi Jing[3,4], Wang Zhou[5] & Yao Hu[1,6,*]

[1]*Department of Statistics, Guizhou University, Guiyang* 550025*, China;*
[2]*Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, GA* 30602*, USA;*
[3]*Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen* 518055*, China;*
[4]*School of Artificial Intelligence, The Chinese University of Hong Kong, Shenzhen, Shenzhen 518172, China;*
[5]*Department of Statistics and Data Science, National University of Singapore, Singapore* 117546*, Singapore;*
[6]*Guizhou Key Laboratory of Artificial Intelligence and Brain-inspired Computing, Guizhou University, Guiyang* 550025*, China*

Email: gs.wychen22@gzu.edu.cn, yewen.chen@uga.edu, jingby@sustech.edu.cn, wangzhou@nus.edu.sg, yhu1@gzu.edu.cn

Received September 23, 2025; accepted March 24, 2026; published online 2026

Abstract Structural breaks frequently arise in economic policy evaluation, biomedical monitoring, and other dynamic systems, posing substantial challenges for partially linear models (PLMs) that contain nonparametric nuisance components. This paper develops a unified tail-adaptive cumulative sum (CUSUM) framework for structural break testing and estimation in PLMs. The proposed methodology integrates cross-fitted semiparametric composite quantile regression with a weighted CUSUM process, allowing robust inference under heterogeneous and heavy-tailed error distributions. A key theoretical contribution is the establishment of a uniform oracle equivalence result, showing that the feasible plug-in CUSUM process converges uniformly to its oracle linear-model counterpart over the time index. This equivalence provides the theoretical foundation for valid tail-adaptive testing and multiplier bootstrap inference in high-dimensional settings. We further establish non-asymptotic localization bounds, showing that the proposed estimator attains the optimal rate up to logarithmic factors. The procedure is implemented within a seeded binary segmentation scheme to accommodate multiple change-points. Simulation studies and an empirical application to energy policy data demonstrate the effectiveness and robustness of the proposed method.

Keywords partially linear models, structural break, tail-adaptive, change-point detection, cumulative sum

MSC(2020) 62M10, 62F35, 62F40, 62G08

Citation: Chen W Y, Chen Y W, Jing B Y, et al. A robust and efficient tail-adaptive approach for change-point detection in semiparametric models. Sci China Math, 2026, 69, https://doi.org/10.1007/s11425-023-xxxx-x

1 Introduction

Partially linear models (PLMs), also known as semiparametric models, play a central role in modern statistics and econometrics by combining the flexibility of nonparametric regression with the interpretability of parametric components. More specifically, a general PLM with multiple nonparametric functions can be represented as follows:

$$Y_i = \mathbf{X}_i^\top \boldsymbol{\beta} + \mathbf{Z}_i^\top \boldsymbol{\alpha}(U_i) + \gamma(U_i) + \varepsilon_i, \tag{1.1}$$

* Corresponding author

where $Y_i \in \mathbb{R}$ for the i-th sampling point is the response of interest. The covariates $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})^\top \in \mathbb{R}^d$ represent the parametric component, with regression coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^\top$. The covariates $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iq})^\top \in \mathbb{R}^q$, together with the index variable $U_i \in \mathbb{R}$, capture more complex interaction effects on Y_i through the smooth nonparametric functions $\boldsymbol{\alpha}(U_i) = (\alpha_1(U_i), \ldots, \alpha_q(U_i))^\top$. In addition, $\gamma(U_i)$ is another smooth nonparametric function representing varying baseline effects, and ε_i denotes a zero-mean error term. In this work, our interest lies in the scenario where the index i corresponds to time.

This unique structure in (1.1) makes them not only theoretically elegant but also highly practical for real-world applications, including economics [34], biomedical research [24], and others involving complex covariate relationships [15]. However, classical PLMs (1.1) face major challenges in statistical inference due to increasingly complex data-generating mechanisms. In particular, when real-world datasets exhibit structural heterogeneity, commonly referred to as structural breaks, classical PLMs often become inapplicable. There remains a pressing need to develop more advanced PLMs and rigorously validate them, both theoretically and empirically.

To facilitate the application of PLMs in settings with structural breaks, change-point analysis—a statistical framework for detecting structural changes [32]—has attracted considerable interest [17–19, 29, 41]. Typically, for model (1.1), the core scientific question is whether there is a structural break in the parametric coefficient sequence $\{\boldsymbol{\beta}_i\}$, commonly referred to as the change-point existence or testing problem. We further formulate this problem as a hypothesis test:

$$\mathrm{H}_0: \quad \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_n, \quad \text{versus} \quad \mathrm{H}_1: \ \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_{l_0-1} \ \neq \ \boldsymbol{\beta}_{l_0} = \boldsymbol{\beta}_{l_0+1} = \cdots = \boldsymbol{\beta}_n, \tag{1.2}$$

where l_0 denotes the location of the (unknown) change-point. In many practical applications in economics, finance, and the social sciences, nonparametric functions tend to capture long-term trends or smooth relationships that remain relatively stable over time. In contrast, changes in the parametric component typically reflect potential structural shifts in the data-generating process. This work focuses on the latter and tests for significant changes in the underlying system relationships, as formulated in (1.2).

From an application perspective, (1.2) involves two important scenarios for describing structural breaks of PLMs. The first is a posteriori (retrospective, offline) change-point detection, which analyzes a fixed historical sample to test whether structural breaks have occurred, followed by estimating their number and locations [2, 4, 5, 16]. By contrast, the second is prospective (sequential, online) monitoring, which treats observations as arriving over time and designs stopping rules to signal a change as quickly as possible while controlling false alarms [9, 12, 30]. Although these two scenarios of structural breaks have been widely studied using fully parametric [1, 10, 20, 31, 33] or nonparametric [3, 27, 37, 38] models, PLMs that accommodate both remain underdeveloped.

Under the simple setting of PLMs with $\boldsymbol{\alpha}(\cdot) \equiv 0$, the structural breaks in the linear coefficients $\boldsymbol{\beta}$ have been studied [36] using the cumulative sum (CUSUM) test [6]. CUSUM is widely used due to its computational simplicity and ease of implementation. However, previous CUSUM-based methods often rely on strong and difficult-to-validate assumptions, such as Gaussian or sub-Gaussian error distributions and low-dimensional covariates [14, 25, 37, 40]. In addition, these methods allow only a small subset of coefficients to change while others remain constant, limiting the flexibility of PLMs for modeling complex data. Furthermore, their covariate selection procedures heavily depend on prior knowledge and lack data-driven strategies for identifying covariates associated with the response. This introduces a risk of model misspecification in PLMs and can result in large variance in coefficient estimation, ultimately compromising the accuracy of detecting structural breaks inherent in the data.

Recent advances in high-dimensional change-point detection have introduced tail-adaptive CUSUM-type procedures that remain valid under unknown and potentially heavy-tailed error distributions [26,28]. These methods employ aggregation schemes that adapt to the tail behavior of the errors and thereby avoid restrictive sub-Gaussian assumptions. Meanwhile, cross-fitting and debiased machine learning techniques have become standard tools for semiparametric inference in the presence of high-dimensional nuisance components [8]. Such approaches mitigate the impact of nuisance estimation errors and facilitate valid inference for low-dimensional target parameters.

Despite these advances, extending high-dimensional adaptive change-point detection to partially linear models remains nontrivial. In PLMs with nonparametric nuisance components, residuals are not directly observable, and inaccurate plug-in estimates of these residuals can quickly propagate through the time-indexed partial sum process underlying CUSUM statistics. This propagation may affect the uniform weak convergence and directly prevents the application of existing adaptive change-point detection frameworks. Developing a new valid methodology for structural break inference thus becomes necessary.

In addition to testing structural breaks, efficient estimation of the change-point location is equally important for understanding structural changes. However, theoretical guarantees for precise localization in partially linear models remain limited, particularly in high-dimensional settings and when multiple change-points are present. This is because cross-iteration errors between complex nonparametric nuisance components and high-dimensional linear components can further distort the time-indexed partial sum processes underlying CUSUM statistics if these components are not carefully developed and implemented. This work addresses both testing and estimation within a unified framework.

Building on previous work, we propose a tail-adaptive CUSUM-type change-point detection procedure for partially linear models. To overcome the aforementioned difficulty, we introduce a cross-fitted semiparametric composite quantile regression (CS–CQR) framework that reduces the partially linear model to a plug-in linear representation suitable for structural break analysis. Both theoretical results and empirical evidence from simulations and an energy data application demonstrate the effectiveness of the proposed method. The main contributions of this paper are summarized as follows.

(1) **Uniform oracle equivalence for semiparametric change-point inference.** We develop a cross-fitted semiparametric composite quantile regression framework and establish a uniform oracle equivalence result showing that the feasible plug-in CUSUM process uniformly approximates its oracle linear-model counterpart over $t \in [q_0, 1-q_0]$ at rate $o_p(1)$. This provides a rigorous bridge between semiparametric nuisance estimation and high-dimensional change-point analysis.

(2) **High-dimensional inference and bootstrap validity.** The aggregated CUSUM statistic involves the supremum of a high-dimensional and sparsity-adaptive stochastic process over time, whose limiting null distribution is analytically intractable. This difficulty is further compounded in the semiparametric setting, where the statistic is constructed from plug-in residuals affected by nuisance estimation. We develop a tuning-free multiplier bootstrap procedure tailored to the proposed plug-in CUSUM framework. Combined with the established uniform oracle equivalence result, the bootstrap consistently approximates the nonstandard limiting distribution under high-dimensional scaling, even when the dimension d diverges with the sample size.

(3) **Localization theory and empirical illustration.** We derive non-asymptotic localization bounds achieving the optimal rate $\log(dn)/n$ up to logarithmic factors for single change-point detection. The procedure is further implemented within a seeded binary segmentation framework and illustrated through simulations and an empirical energy policy application.

The remainder of this paper is organized as follows. In Section 2, we present our tail-adaptive procedure for locating a single change-point. Section 3 develops its theoretical guarantees, including size control, power analysis, and the asymptotic accuracy of the estimated change-point. In Section 4, we report extensive simulation experiments, and in Section 5 we illustrate the method on the energy dataset. Finally, Section 6 contains the conclusion and proposes prospects for further research.

Notations: Throughout this paper, for a vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^\top \in \mathbb{R}^d$, we define its ℓ_p-norm as $\|\boldsymbol{\mu}\|_p = \left(\sum_{j=1}^d |\mu_j|^p\right)^{1/p}$ for $1 \leqslant p < \infty$, and $\|\boldsymbol{\mu}\|_\infty = \max_{1\leqslant j\leqslant d} |\mu_j|$. In addition, we write $\|\boldsymbol{\mu}\|_0 := \sum_{j=1}^d \mathbb{I}\{\mu_j \neq 0\}$. For two real-valued sequences $\{x_n\}$ and $\{y_n\}$, we write $x_n = O(y_n)$ if there exists $C > 0$ such that $|x_n| \leqslant C|y_n|$ for all large n, and $x_n = o(y_n)$ if $x_n/y_n \to 0$ as $n \to \infty$. For a random sequence $\{\zeta_n\}$, we write $\zeta_n \xrightarrow{\mathbb{P}} \zeta$ if ζ_n converges to ζ in probability, and let $\zeta_n = o_p(1)$ if $\zeta_n \xrightarrow{\mathbb{P}} 0$, and $\zeta_n = O_p(1)$ if $\{\zeta_n\}$ is stochastically bounded. For any finite set $\mathcal{B}$, we denote its cardinality by $|\mathcal{B}|$. For $x \in \mathbb{R}^+$, let $\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leqslant x\}$ be the floor function. The indicator function is denoted by $\mathbb{I}\{\cdot\}$. For any $a, b \in \mathbb{R}$, we write $a \vee b := \max\{a, b\}$.

2 Methodology

In this section, we further elaborate on the statistical challenges and propose a novel procedure to tackle the hypothesis testing problem in (1.2).

2.1 Change-point model

We restrict attention to the case of at most one change-point in model (1.1). We assume that $\{(\mathbf{X}_i, \mathbf{Z}_i, U_i, Y_i)\}_{i=1}^n$ are independent, but (time) ordered observations. Then, to motivate our testing strategy, we reparameterize the linear component as follows:

$$Y_i = \mathbf{X}_i^\top \big(\boldsymbol{\beta}^{(1)} + \boldsymbol{\delta}_{l_0}\, \mathbb{I}\{i \geqslant l_0\}\big) + \mathbf{Z}_i^\top \boldsymbol{\alpha}(U_i) + \gamma(U_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where $\boldsymbol{\delta}_{l_0} = \boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}$ captures the shift in the parametric coefficients at the (unknown) change-point l_0, which we represent as $l_0 = \lfloor n\, t_0 \rfloor$, for some $t_0 \in (0, 1)$ denoting its relative location. To avoid boundary effects, we further assume $t_0 \in [q_0, 1-q_0]$ for some fixed $q_0 \in (0, 0.5)$, which is a common assumption in the change-point literature (see [11,21,26]). Finally, to mitigate the curse of dimensionality in nonparametric estimation, we restrict U to be univariate throughout.

A natural question for model (2.1) is whether a change-point exists. Equivalently, we consider the following hypothesis test:

$$\mathrm{H}_0 :\ \boldsymbol{\delta}_{l_0} = \mathbf{0} \text{ for all } l_0 \quad \text{versus} \quad \mathrm{H}_1 :\ \exists\, l_0 \in \{2, \ldots, n-1\} \text{ such that } \boldsymbol{\delta}_{l_0} \neq \mathbf{0}. \tag{2.2}$$

Note that test (2.2) is equivalent to problem (1.2) in Section 1: accepting H_0 implies there is no structural break in the parametric coefficients of either model (2.1) or model (1.1).

To construct our test statistic, we first reformulate model (2.1) as a linear regression model. In particular, since the nonparametric components $\boldsymbol{\alpha}(u)$ and $\gamma(u)$ are unknown (but assumed time-invariant), we estimate them from the data and denote the resulting estimators by $\widetilde{\boldsymbol{\alpha}}(u)$ and $\widetilde{\gamma}(u)$. For each observation i, we let $\eta(U_i) := \mathbf{Z}_i^\top \alpha(U_i) + \gamma(U_i)$. Then we define the adjusted response as follows:

$$\widetilde{Y}_i := Y_i - \widetilde{\eta}(U_i) = Y_i - \mathbf{Z}_i^\top \widetilde{\boldsymbol{\alpha}}(U_i) - \widetilde{\gamma}(U_i). \tag{2.3}$$

Substituting it into (2.1), we obtain the residualized change-point regression model given by

$$\widetilde{Y}_i = \mathbf{X}_i^\top \boldsymbol{\beta}^{(1)} + \mathbf{X}_i^\top \boldsymbol{\delta}_{l_0}\, \mathbb{I}\{i \geqslant l_0\} - \vartheta_i + \varepsilon_i, \tag{2.4}$$

where $\vartheta_i := \widetilde{\eta}(U_i) - \eta(U_i)$ captures the plug-in bias from estimating the nonparametric components.

With this adjustment, the conditional expectation of $\widetilde{Y}_i$ given $\mathbf{X}_i$ becomes

$$\mathbb{E}\big[\widetilde{Y}_i \mid \mathbf{X}_i\big] = \mathbf{X}_i^\top \boldsymbol{\beta}^{(1)} + \mathbf{X}_i^\top \boldsymbol{\delta}_{l_0}\, \mathbb{I}\{i \geqslant l_0\} - \mathbb{E}\big[\vartheta_i \mid \mathbf{X}_i\big]. \tag{2.5}$$

Next, let $0 < \tau_1 < \cdots < \tau_L < 1$ denote a set of candidate quantile levels. For each τ_l, define the conditional error quantile $r_l^*(\mathbf{X}_i) := \inf\{\, t : \mathbb{P}(\varepsilon_i \leqslant t + \vartheta_i \mid \mathbf{X}_i) \geqslant \tau_l\}$, so that the conditional τ_l-quantile of $\widetilde{Y}_i$ given $\mathbf{X}_i$ is

$$M_{\tau_l}(\widetilde{Y}_i \mid \mathbf{X}_i) = r_l^*(\mathbf{X}_i) \ + \ \mathbf{X}_i^\top \boldsymbol{\beta}^{(1)} + \ \mathbf{X}_i^\top \boldsymbol{\delta}_{l_0}\, \mathbb{I}\{i \geqslant l_0\}, \quad l = 1, \ldots, L, \tag{2.6}$$

where $M_{\tau_l}(\widetilde{Y}_i \mid \mathbf{X}_i) := \inf\big\{t : \mathbb{P}(\widetilde{Y}_i \leqslant t \mid \mathbf{X}_i) \geqslant \tau_l\big\}$.

In fact, from (2.3) the adjusted response subtracts the smooth nonparametric terms, so any discontinuity in the regression of $\widetilde{Y}_i$ on $\mathbf{X}_i$ at $i = l_0$ must arise from a jump in the parametric coefficient $\boldsymbol{\beta}$. Conversely, a structural break in the original regression of Y_i on $\mathbf{X}_i$ at l_0 is exactly mirrored in the transformed regression. Hence, inference based on $\widetilde{Y}_i$ is equivalent to that in the original model in terms of the change-point structure. Therefore, testing H_0 in (1.2) can be carried out equivalently based on either (2.5) or (2.6).

Note that in the oracle case, where the plug-in bias vanishes identically, i.e., $\vartheta_i = \eta(U_i) - \widetilde{\eta}(U_i) \equiv 0$, model (2.4) reduces to the change-point regression setting studied by [26]. We adopt a weighted composite

loss framework to construct a test statistic that balances robustness and efficiency. Specifically, for any fixed weight $\lambda \in [0,1]$, define the weighted composite loss function as

$$\ell_\lambda(\boldsymbol{x}, y; \widetilde{\boldsymbol{\tau}}, \boldsymbol{r}, \boldsymbol{\beta}) := \frac{1-\lambda}{L} \sum_{l=1}^{L} \rho_{\tau_l}(y - r_l - \boldsymbol{x}^\top \boldsymbol{\beta}) + \frac{\lambda}{2}(y - \boldsymbol{x}^\top \boldsymbol{\beta})^2, \tag{2.7}$$

where $\rho_\tau(u) := u(\tau - \mathbb{I}\{u \leqslant 0\})$ is the check loss at quantile level τ, $\widetilde{\boldsymbol{\tau}} = (\tau_1, \ldots, \tau_L)^\top$ denotes the user-specified quantile levels, and $\boldsymbol{r} = (r_1, \ldots, r_L)^\top \in \mathbb{R}^L$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^\top \in \mathbb{R}^d$. The tuning parameter λ governs the trade-off between robustness and efficiency. When $\lambda = 0$, the loss reduces to the composite quantile loss, which is more robust to outliers. When $\lambda = 1$, it becomes the squared loss, which is optimal under Gaussian errors.

The subgradient of ℓ_λ with respect to $\boldsymbol{\beta}$ yields the following composite score function:

$$\boldsymbol{S}_\lambda^{\mathrm{lin}}(\boldsymbol{x}, y; \widetilde{\boldsymbol{\tau}}, \boldsymbol{\theta}) := \frac{1-\lambda}{L} \sum_{l=1}^{L} \boldsymbol{x} \left(\mathbb{I}\{y - r_l - \boldsymbol{x}^\top \boldsymbol{\beta} \leqslant 0\} - \tau_l \right) - \lambda\, \boldsymbol{x}(y - \boldsymbol{x}^\top \boldsymbol{\beta}), \tag{2.8}$$

where $\boldsymbol{\theta} := (\boldsymbol{r}^\top, \boldsymbol{\beta}^\top)^\top \in \mathbb{R}^{L+d}$ denotes the vector of parameters to be estimated. Following [26], we then define the individual CUSUM process for each $\lambda \in [0,1]$ and $t \in (0,1)$ as

$$\boldsymbol{G}_\lambda^{\mathrm{lin}}(t; \widetilde{\boldsymbol{\tau}}, \boldsymbol{\theta}) = \frac{1}{\sqrt{n}\, \sigma(\lambda, \widetilde{\boldsymbol{\tau}})} \left[\boldsymbol{S}_\lambda(t; \widetilde{\boldsymbol{\tau}}, \boldsymbol{\theta}) - \frac{\lfloor nt \rfloor}{n} \boldsymbol{S}_\lambda(1; \widetilde{\boldsymbol{\tau}}, \boldsymbol{\theta}) \right], \tag{2.9}$$

where the partial sum process is defined by

$$\boldsymbol{S}_\lambda(t; \widetilde{\boldsymbol{\tau}}, \boldsymbol{\theta}) = \sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{S}_\lambda^{\mathrm{lin}}(\boldsymbol{X}_i, \widetilde{Y}_i; \widetilde{\boldsymbol{\tau}}, \boldsymbol{\theta}), \tag{2.10}$$

and the variance is given by $\sigma^2(\lambda, \widetilde{\boldsymbol{\tau}}) = \mathrm{Var}\big[(1-\lambda) e_i(\widetilde{\boldsymbol{\tau}}) - \lambda\, \varepsilon_i\big]$ with $e_i(\widetilde{\boldsymbol{\tau}}) = \frac{1}{L} \sum_{l=1}^{L} \big(\mathbb{I}\{\varepsilon_i \leqslant r_l^*\} - \tau_l \big)$.

To detect structural breaks in model (2.4), [26] proposed a family of oracle test statistics, each indexed by a pair (λ, s_0). The corresponding statistic is

$$\Gamma_\lambda = \max_{t \in [q_0, 1-q_0]} \left\| \boldsymbol{G}_\lambda^{\mathrm{lin}}\big(t; \widetilde{\boldsymbol{\tau}}, \boldsymbol{\theta}\big) \right\|_{(s_0,2)}, \quad \lambda \in [0,1], \tag{2.11}$$

where the $(s_0, 2)$-norm is a special case of the $(s_0, \widetilde{p})$-norm introduced by [28], defined for any $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^\top \in \mathbb{R}^d$ as $\|\boldsymbol{\mu}\|_{(s_0,\widetilde{p})} := \Big(\sum_{j=1}^{s_0} |\mu|_{(j)}^{\widetilde{p}} \Big)^{1/\widetilde{p}}$. Here, $|\mu|_{(1)} \geqslant \cdots \geqslant |\mu|_{(d)}$ are the order statistics of $|\mu_1|, \ldots, |\mu_d|$, and $s_0 \in \{1, \ldots, d\}$, $\widetilde{p} \in [1, \infty)$.

It is important to note that $\{\Gamma_\lambda, \lambda \in [0,1]\}$ is constructed under the oracle setting, where the nonparametric components $\boldsymbol{\alpha}(\cdot)$ and $\gamma(\cdot)$ are assumed known. In this case, the adjusted response $\widetilde{Y}_i$ is observable, and the plug-in bias ϑ_i vanishes identically. However, in the partially linear model (1.1), both $\boldsymbol{\alpha}(\cdot)$ and $\gamma(\cdot)$ are unknown and must be estimated from the data. Consequently, the plug-in bias is non-negligible and needs to be properly addressed to ensure valid inference.

2.2 Test statistic

To adapt the oracle test statistic (2.11) to the partially linear setting, we first estimate the nonparametric components $\boldsymbol{\alpha}(\cdot)$ and $\gamma(\cdot)$ in model (1.1). For this purpose, we develop a novel cross-fitted semiparametric composite quantile regression (CS-CQR) procedure, which allows us to construct adjusted responses that asymptotically approximate the oracle form. Our procedure is motivated by the semiparametric CQR of [22] and by the double/debiased machine-learning (DML) framework of [8]. More precisely, let $\{(\mathbf{X}_i, \mathbf{Z}_i, U_i, Y_i)\}_{i=1}^n$ be an i.i.d. sample generated by (1.1). The CS-CQR procedure consists of the following two main steps (see Algorithm 1 for more details):

Step 1 (Local-linear estimation on the auxiliary sample). We randomly partition the index set $\{1,\dots,n\}$ into K folds $\mathcal{I}_1,\dots,\mathcal{I}_K$ and, for each k, define the following auxiliary index set:

$$\mathcal{I}_k^c \;=\; \{1,\dots,n\}\setminus \mathcal{I}_k \;=\; \bigcup_{j=1,j\neq k} \mathcal{I}_j. \tag{2.12}$$

For each k and each target point $u_0 \in \{U_i : i \in \mathcal{I}_k\}$, obtain the local-linear CQR estimates $(\widetilde{\boldsymbol{\alpha}}^{(-k)}(u_0), \widetilde{\gamma}^{(-k)}(u_0))$ with a fixed offset $\boldsymbol{\beta}^\dagger$ by solving

$$\min_{\substack{\mathbf{a}_0,\mathbf{a}_1\in\mathbb{R}^q\\ \gamma_0,\gamma_1\in\mathbb{R}}} \sum_{\ell=1}^{L}\sum_{i\in\mathcal{I}_k^c} \rho_{\tau_\ell}\Big(Y_i - \mathbf{X}_i^\top\boldsymbol{\beta}^\dagger - \mathbf{Z}_i^\top\big(\mathbf{a}_0+\mathbf{a}_1(U_i-u_0)\big) - \big(\gamma_0+\gamma_1(U_i-u_0)\big)\Big)\,\mathcal{K}_h(U_i-u_0), \tag{2.13}$$

and then set $\widetilde{\boldsymbol{\alpha}}^{(-k)}(u_0) := \mathbf{a}_0$ and $\widetilde{\gamma}^{(-k)}(u_0) := \gamma_0$. Here, $\rho_\tau(u) = u\big(\tau - \mathbb{I}\{u<0\}\big)$ is the check loss, and $\mathcal{K}_h(u) = \mathcal{K}(u/h)/h$ is a kernel with bandwidth h.

Remark 2.1. This sample-splitting design prevents the same observations from being used both to estimate $\big(\boldsymbol{\alpha}(\cdot),\gamma(\cdot)\big)$ and to evaluate the test statistic, thereby mitigating overfitting and reducing small-sample bias. We compute $\boldsymbol{\beta}^\dagger$ once, using Algorithm A.2 in Appendix D, and treat it as fixed throughout CS-CQR; it is not re-estimated within folds or across bandwidths. Under mild regularity conditions, we show that $\boldsymbol{\beta}^\dagger$ is $\sqrt{n}$-consistent (see Appendix D.2 for details).

Step 2 (Cross-fitting onto the held-out fold). For each held-out observation $i \in \mathcal{I}_k$, we set

$$\widetilde{\boldsymbol{\alpha}}(U_i) = \widetilde{\boldsymbol{\alpha}}^{(-k)}(U_i), \widetilde{\gamma}(U_i) = \widetilde{\gamma}^{(-k)}(U_i). \tag{2.14}$$

Repeating Steps **1-2** for $k = 1,\dots,K$ yields the cross-fitted estimators $\widetilde{\boldsymbol{\alpha}}(\cdot)$ and $\widetilde{\gamma}(\cdot)$ for all $i = 1,\dots,n$.

Algorithm 1 Cross-fitted semiparametric CQR estimator

Input: Data $\{(\mathbf{X}_i,\mathbf{Z}_i,U_i,Y_i)\}_{i=1}^n$, quantile levels $\widetilde{\boldsymbol{\tau}} = (\tau_1,\dots,\tau_L)^\top$, kernel $\mathcal{K}_h(u) = \mathcal{K}(u/h)/h$, bandwidth grid $\mathcal{H}$, number of folds K.

1: Randomly partition $\{1,\dots,n\}$ into K folds $\mathcal{I}_1,\dots,\mathcal{I}_K$. For $k = 1,\dots,K$, set $\mathcal{I}_k^c = \bigcup_{j\neq k}\mathcal{I}_j$.
2: Compute the fixed offset $\boldsymbol{\beta}^\dagger$ via Algorithm A.2 (Appendix D)
3: Around any target u_0, use the local-linear approximations $\mathbf{Z}_j^\top\boldsymbol{\alpha}(U_j) \approx \mathbf{Z}_j^\top\{\boldsymbol{\alpha}_0+\boldsymbol{\alpha}_1(U_j-u_0)\}$, and $\gamma(U_j) \approx \gamma_0 + \gamma_1(U_j-u_0)$, with weights $\mathcal{K}_h(U_j-u_0)$ for each $j = 1,\dots,q$.
4: **for** $h \in \mathcal{H}$ **do**
5: **for** $k = 1,\dots,K$ **do**
6: For each $i \in \mathcal{I}_k$ (set $u_0 := U_i$), compute the local-linear composite quantile problem

$$(\widehat{\boldsymbol{\alpha}}_0,\widehat{\boldsymbol{\alpha}}_1,\widehat{\gamma}_0,\widehat{\gamma}_1) \leftarrow \operatorname*{argmin}_{\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_1,\gamma_0,\gamma_1} \sum_{j\in\mathcal{I}_k^c}\sum_{\ell=1}^{L} \rho_{\tau_\ell}\Big(Y_j - \mathbf{X}_j^\top\boldsymbol{\beta}^\dagger - \mathbf{Z}_j^\top\{\boldsymbol{\alpha}_0+\boldsymbol{\alpha}_1(U_j-u_0)\} - \{\gamma_0+\gamma_1(U_j-u_0)\}\Big)\mathcal{K}_h(U_j-u_0),$$

and set $\widehat{g}_h^{(-k)}(U_i,\mathbf{Z}_i) = \mathbf{Z}_i^\top\widehat{\boldsymbol{\alpha}}_0 + \widehat{\gamma}_0$.
7: **end for**
8: Compute the cross-validated loss $\mathrm{CV}(h) = \sum_{k=1}^K\sum_{i\in\mathcal{I}_k}\sum_{\ell=1}^L \rho_{\tau_\ell}\big(Y_i - \mathbf{X}_i^\top\boldsymbol{\beta}^\dagger - \widehat{g}_h^{(-k)}(U_i,\mathbf{Z}_i)\big)$.
9: **end for**
10: Choose $\widehat{h} \leftarrow \arg\min_{h\in\mathcal{H}} \mathrm{CV}(h)$.
11: **for** $k = 1,\dots,K$ **do**
12: For each $i \in \mathcal{I}_k$ (set $u_0 := U_i$), re-solve the local-linear problem on $\mathcal{I}_k^c$ with $h = \widehat{h}$ and offset $\boldsymbol{\beta}^\dagger$ to obtain $(\widehat{\boldsymbol{\alpha}}_0,\widehat{\gamma}_0)$, and set $\widetilde{\boldsymbol{\alpha}}(U_i) = \widehat{\boldsymbol{\alpha}}_0, \widetilde{\gamma}(U_i) = \widehat{\gamma}_0$.
13: **end for**

Output: The final estimators $\widetilde{\boldsymbol{\alpha}}(\cdot)$ and $\widetilde{\gamma}(\cdot)$.

Using the cross-fitted estimates $\widetilde{\boldsymbol{\alpha}}(U_i)$ and $\widetilde{\gamma}(U_i)$, we construct the pseudo-data as

$$\widetilde{Y}_i = Y_i - \mathbf{Z}_i^\top\widetilde{\boldsymbol{\alpha}}(U_i) - \widetilde{\gamma}(U_i), \quad \widetilde{\mathbf{X}}_i = \mathbf{X}_i, \quad i = 1,\dots,n. \tag{2.15}$$

Based on the pseudo-data $\{(\widetilde{\mathbf{X}}_i,\widetilde{Y}_i)\}_{i=1}^n$, we define the plug-in composite score as

$$\boldsymbol{S}_\lambda^{\mathrm{plug}}(\widetilde{\mathbf{X}}_i,\widetilde{Y}_i;\widetilde{\boldsymbol{\tau}},\boldsymbol{\theta}) = \frac{1-\lambda}{L}\sum_{\ell=1}^{L}\widetilde{\mathbf{X}}_i\Big(\mathbb{I}\{\widetilde{Y}_i - r_\ell - \widetilde{\mathbf{X}}_i^\top\boldsymbol{\beta} \leqslant 0\} - \tau_\ell\Big) - \lambda\,\widetilde{\mathbf{X}}_i\big(\widetilde{Y}_i - \widetilde{\mathbf{X}}_i^\top\boldsymbol{\beta}\big). \tag{2.16}$$

The corresponding plug-in partial sum process is defined by

$$\widetilde{\boldsymbol{S}}_\lambda(t;\widetilde{\boldsymbol{\tau}},\boldsymbol{\theta}) = \sum_{i=1}^{\lfloor nt\rfloor} \boldsymbol{S}_\lambda^{\mathrm{plug}}(\widetilde{\mathbf{X}}_i,\widetilde{Y}_i;\widetilde{\boldsymbol{\tau}},\boldsymbol{\theta}), \tag{2.17}$$

and the plug-in CUSUM process is given by

$$\widetilde{\boldsymbol{G}}_\lambda(t;\widetilde{\boldsymbol{\tau}},\boldsymbol{\theta}) = \frac{1}{\sqrt{n}\,\sigma(\lambda,\widetilde{\boldsymbol{\tau}})}\Big[\widetilde{\boldsymbol{S}}_\lambda(t;\widetilde{\boldsymbol{\tau}},\boldsymbol{\theta}) - \tfrac{\lfloor nt\rfloor}{n}\,\widetilde{\boldsymbol{S}}_\lambda(1;\widetilde{\boldsymbol{\tau}},\boldsymbol{\theta})\Big]. \tag{2.18}$$

To complete the construction of our test statistic, we replace the unknown scale $\sigma(\lambda,\widetilde{\boldsymbol{\tau}})$ in (2.18) with a consistent estimator $\widehat{\sigma}(\lambda,\widetilde{\boldsymbol{\tau}})$ in the spirit of [26]. Since the true parameter vector $\boldsymbol{\theta}$ is unknown under H_0, we estimate it by solving, for each $\lambda\in[0,1]$, the following penalized composite quantile optimization problem:

$$\widehat{\boldsymbol{\theta}}_\lambda = \underset{\boldsymbol{\theta}\in\mathbb{R}^{L+d}}{\operatorname{argmin}}\left[\frac{1-\lambda}{nL}\sum_{i=1}^{n}\sum_{l=1}^{L}\rho_{\tau_l}\Big(\widetilde{Y}_i - \widetilde{\mathbf{X}}_i^\top\boldsymbol{\beta} - r_l\Big) + \frac{\lambda}{2n}\,\|\boldsymbol{\Xi}\|_2^2 + \kappa_\lambda\|\boldsymbol{\beta}\|_1\right], \tag{2.19}$$

where $\boldsymbol{\Xi} := \Big(\widetilde{Y}_1 - \widetilde{\mathbf{X}}_1^\top\boldsymbol{\beta},\ldots,\widetilde{Y}_n - \widetilde{\mathbf{X}}_n^\top\boldsymbol{\beta}\Big)^\top$, $\widehat{\boldsymbol{\theta}}_\lambda := (\widehat{r}_{\lambda,1},\ldots,\widehat{r}_{\lambda,L},\,\widehat{\beta}_{\lambda,1},\ldots,\widehat{\beta}_{\lambda,d})^\top$, and $\kappa_\lambda\geqslant 0$ controls the sparsity of $\widehat{\boldsymbol{\beta}}_\lambda$.

After estimating $\widehat{\boldsymbol{\theta}}_\lambda$, we substitute it into $\widetilde{\boldsymbol{S}}_\lambda(\cdot;\widetilde{\boldsymbol{\tau}},\boldsymbol{\theta})$. Consequently, we obtain the oracle plug-in CUSUM process $\Big\{\widetilde{\boldsymbol{G}}_\lambda\big(t;\widetilde{\boldsymbol{\tau}},\widehat{\boldsymbol{\theta}}_\lambda\big),\,t\in[q_0,1-q_0]\Big\}$, defined by

$$\widetilde{\boldsymbol{G}}_\lambda\big(t;\widetilde{\boldsymbol{\tau}},\widehat{\boldsymbol{\theta}}_\lambda\big) = \frac{1}{\sqrt{n}\widehat{\sigma}(\lambda,\widetilde{\boldsymbol{\tau}})}\left(\widetilde{\boldsymbol{S}}_\lambda\big(t;\widetilde{\boldsymbol{\tau}},\widehat{\boldsymbol{\theta}}_\lambda\big) - \frac{\lfloor nt\rfloor}{n}\widetilde{\boldsymbol{S}}_\lambda\big(1;\widetilde{\boldsymbol{\tau}},\widehat{\boldsymbol{\theta}}_\lambda\big)\right). \tag{2.20}$$

Note that we refer to $\widetilde{\boldsymbol{G}}_\lambda\big(t;\widetilde{\boldsymbol{\tau}},\widehat{\boldsymbol{\theta}}_\lambda\big)$ as an oracle because there are no unknown parameters in the test statistic. Hence, for any $\lambda\in[0,1]$ and user-specified s_0, we define the oracle individual test statistic as

$$\widetilde{\Gamma}_\lambda = \max_{t\in[q_0,\,1-q_0]}\Big\|\widetilde{\boldsymbol{G}}_\lambda\big(t;\widetilde{\boldsymbol{\tau}},\widehat{\boldsymbol{\theta}}_\lambda\big)\Big\|_{(s_0,2)}. \tag{2.21}$$

Specifically, for each given $\lambda\in[0,1]$, once H_0 is rejected, the change-point estimator is defined by

$$\widehat{t}_\lambda = \underset{q_0\leqslant t\leqslant 1-q_0}{\operatorname{argmax}}\Big\|\widetilde{\boldsymbol{G}}_\lambda\big(t;\widetilde{\boldsymbol{\tau}},\widehat{\boldsymbol{\theta}}_\lambda\big)\Big\|_{(s_0,2)}. \tag{2.22}$$

Although the statistic $\widetilde{\Gamma}_\lambda$ in (2.21) is well-defined, its limiting null distribution is analytically intractable, especially when the parameter dimension d increases with the sample size n. To address this, we employ a multiplier bootstrap procedure to approximate the null distribution of $\widetilde{\Gamma}_\lambda$ under H_0, which enables accurate determination of the corresponding critical values and p-values.

2.3 Bootstrap procedure for test statistics

In this subsection, we adopt a multiplier bootstrap procedure (see [21, 26, 28]) to approximate the null distribution of $\widetilde{\Gamma}_\lambda$. Let B denote the number of bootstrap replications. For each $b=1,\ldots,B$, generate i.i.d. standard normal variates $\varepsilon_1^b,\ldots,\varepsilon_n^b\sim N(0,1)$ and set $e_i^b(\widetilde{\boldsymbol{\tau}}) = \frac{1}{L}\sum_{l=1}^{L}\Big[\mathbb{I}\{\varepsilon_i^b\leqslant\Phi^{-1}(\tau_l)\}-\tau_l\Big]$, where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution.

We now define the bth bootstrap-based CUSUM process as

$$\widetilde{\boldsymbol{G}}_\lambda^b(t;\widetilde{\boldsymbol{\tau}}) = \frac{1}{\sqrt{n}\,\nu(\lambda,\widetilde{\boldsymbol{\tau}})}\Big[\widetilde{\boldsymbol{S}}_\lambda^b(t;\widetilde{\boldsymbol{\tau}}) - \tfrac{\lfloor nt\rfloor}{n}\,\widetilde{\boldsymbol{S}}_\lambda^b(1;\widetilde{\boldsymbol{\tau}})\Big], \tag{2.23}$$

where $\widetilde{\boldsymbol{S}}_\lambda^b(t;\widetilde{\boldsymbol{\tau}}) = \sum_{i=1}^{\lfloor nt\rfloor}\mathbf{X}_i\Big[(1-\lambda)\,e_i^b(\widetilde{\boldsymbol{\tau}}) - \lambda\,\varepsilon_i^b\Big]$, and the variance term is defined by $\nu^2(\lambda,\widetilde{\boldsymbol{\tau}}) := (1-\lambda)^2\operatorname{Var}\big[e_i^b(\widetilde{\boldsymbol{\tau}})\big] + \lambda^2 - 2\lambda(1-\lambda)\operatorname{Cov}\big(e_i^b(\widetilde{\boldsymbol{\tau}}),\varepsilon_i^b\big)$. Since $\varepsilon_i^b\sim N(0,1)$ and $e_i^b(\widetilde{\boldsymbol{\tau}})$ are known, $\nu^2(\lambda,\widetilde{\boldsymbol{\tau}})$ can be computed directly within the bootstrap procedure.

Building on the bootstrap samples $\{\widetilde{\boldsymbol{G}}_\lambda^b(t;\widetilde{\boldsymbol{\tau}})\}_{b=1}^B$ and a user-specified tuning constant s_0, we define the bth bootstrap version of the test statistic $\widetilde{\Gamma}_\lambda$ as

$$\widetilde{\Gamma}_\lambda^b = \max_{q_0\leqslant t\leqslant 1-q_0} \left\|\widetilde{\boldsymbol{G}}_\lambda^b(t;\widetilde{\boldsymbol{\tau}})\right\|_{(s_0,2)}, \quad \lambda\in[0,1]. \tag{2.24}$$

For a given significance level $\alpha\in(0,0.5)$, let $C_\alpha := \inf\left\{t\in\mathbb{R}:\mathbb{P}\left(\widetilde{\Gamma}_\lambda\leqslant t\right)\geqslant 1-\alpha\right\}$ and P_λ denote the critical value and theoretical p-value of the test statistic $\widetilde{\Gamma}_\lambda$, respectively. Based on the bootstrap samples $\{\widetilde{\Gamma}_\lambda^b\}_{b=1}^B$, we estimate them by

$$\widehat{C}_\alpha = \inf\Big\{t:\frac{1}{B}\sum_{b=1}^B\mathbb{I}\{\widetilde{\Gamma}_\lambda^b\leqslant t\}\geqslant 1-\alpha\Big\} \text{ and } \widehat{P}_\lambda = \frac{1}{B+1}\sum_{b=1}^B\mathbb{I}\{\widetilde{\Gamma}_\lambda^b>\widetilde{\Gamma}_\lambda\}, \text{ with } \lambda\in[0,1]. \tag{2.25}$$

We then define the individual test decision by

$$\Theta_{\lambda,\alpha} := \mathbb{I}\{\widehat{P}_\lambda\leqslant\alpha\}, \quad \lambda\in[0,1]. \tag{2.26}$$

Note that $\widehat{P}_\lambda\leqslant\alpha$ if and only if $\widetilde{\Gamma}_\lambda\geqslant\widehat{C}_\alpha$. Hence, we reject the null hypothesis H_0 at the significance level α precisely when $\Theta_{\lambda,\alpha}=1$.

However, existing research (see [26]) has shown that (i) for a fixed error distribution, the power of the test statistic $\widetilde{\Gamma}_\lambda$ varies with the choice of λ, and (ii) smaller p-values provide stronger evidence against the null hypothesis H_0. To guarantee optimal testing performance, we therefore introduce a tail-adaptive statistic that selects the tuning parameter λ in a data-driven manner. Concretely, for the family of individual test statistics $\{\widetilde{\Gamma}_\lambda:\lambda\in[0,1]\}$, we define the tail-adaptive test statistic as

$$\Gamma_{\mathrm{ad}} = \min_{\lambda\in\mathcal{D}}\widehat{P}_\lambda, \tag{2.27}$$

where $\mathcal{D}\subseteq[0,1]$ is a prespecified candidate set of λ.

For the tail-adaptive test statistic Γ_{ad}, let $F(t)=\mathbb{P}\big(\Gamma_{\mathrm{ad}}\leqslant t\big)$ and $P_{\mathrm{ad}}=1-F\big(\Gamma_{\mathrm{ad}}\big)$ denote its (unknown) distribution function and theoretical p-value, respectively. Since F cannot be evaluated directly, we approximate P_{ad} using the low-cost bootstrap method proposed by [42]. The core idea of the low-cost bootstrap method is to efficiently utilize the collection of bootstrap samples $\left\{\widetilde{\Gamma}_\lambda^1,\ldots,\widetilde{\Gamma}_\lambda^B\right\}$. More specifically, for $b=1,\ldots,B$, we set the b-th low-cost bootstrap sample for the theoretical p-values of $\widetilde{\Gamma}_\lambda$ as

$$\widehat{P}_\lambda^b = \frac{1}{B}\sum_{b'\neq b}\mathbb{I}\left(\widetilde{\Gamma}_\lambda^{b'}>\widetilde{\Gamma}_\lambda^b\right). \tag{2.28}$$

Then, we define the b-th bootstrap sample for the tail-adaptive test statistic Γ_{ad} as

$$\Gamma_{\mathrm{ad}}^b = \min_{\lambda\in\mathcal{D}}\widehat{P}_\lambda^b. \tag{2.29}$$

Furthermore, on the basis of the bootstrap samples $\left\{\Gamma_{\mathrm{ad}}^1,\ldots,\Gamma_{\mathrm{ad}}^B\right\}$, we can approximate P_{ad} by

$$\widehat{P}_{\mathrm{ad}} = \frac{1}{B+1}\sum_{b=1}^B\mathbb{I}\left(\Gamma_{\mathrm{ad}}^b\leqslant\Gamma_{\mathrm{ad}}\right). \tag{2.30}$$

Then, given the significance level α, we define the tail-adaptive test as

$$\Theta_{\mathrm{ad},\alpha} := \mathbb{I}\{\widehat{P}_{\mathrm{ad}}\leqslant\alpha\}. \tag{2.31}$$

We reject H_0 at the significance level α when $\Theta_{\mathrm{ad},\alpha}=1$. In that case, the selected weight parameter is

$$\widehat{\lambda} = \operatorname*{argmin}_{\lambda\in\mathcal{D}}\widehat{P}_\lambda, \tag{2.32}$$

and the corresponding change-point estimator is

$$\widehat{t}_{\widehat{\lambda}} = \operatorname*{argmax}_{q_0\leqslant t\leqslant 1-q_0}\left\|\widetilde{\boldsymbol{G}}_{\widehat{\lambda}}\big(t;\widetilde{\boldsymbol{\tau}},\widehat{\boldsymbol{\theta}}_{\widehat{\lambda}}\big)\right\|_{(s_0,2)}. \tag{2.33}$$

3 Theoretical properties

In this section, we establish the theoretical properties of the proposed method.

Let $\{\widetilde{\boldsymbol{\alpha}}(\cdot), \widetilde{\gamma}(\cdot)\}$ be the cross-fitted estimators produced by Algorithm 1, and define

$$\eta(U_i) := \mathbf{Z}_i^\top \boldsymbol{\alpha}(U_i) + \gamma(U_i), \qquad \widetilde{\eta}(U_i) := \mathbf{Z}_i^\top \widetilde{\boldsymbol{\alpha}}(U_i) + \widetilde{\gamma}(U_i).$$

Set the plug-in bias $\vartheta_i := \widetilde{\eta}(U_i) - \eta(U_i)$ and the pseudo-responses $\widetilde{Y}_i := Y_i - \widetilde{\eta}(U_i)$.

Note that, since $\widetilde{Y}_i = \mathbf{X}_i^\top \boldsymbol{\beta} + \varepsilon_i - \vartheta_i$, it follows that $\widetilde{Y}_i - \mathbf{X}_i^\top \boldsymbol{\beta} = \varepsilon_i - \vartheta_i$ and $\mathbb{I}\{\widetilde{Y}_i - r_\ell - \mathbf{X}_i^\top \boldsymbol{\beta} \leqslant 0\} = \mathbb{I}\{\varepsilon_i \leqslant r_\ell + \vartheta_i\}$. For theoretical comparison with the (linear model) CUSUM process of [26], we decompose the plug-in composite score into an oracle (linear model) score part and a plug-in remainder. Recalling $\widetilde{\mathbf{X}}_i = \mathbf{X}_i$ and $\varepsilon_i := Y_i - \eta(U_i) - \mathbf{X}_i^\top \boldsymbol{\beta}$, write

$$\boldsymbol{S}_\lambda^{\mathrm{plug}}(\widetilde{\mathbf{X}}_i, \widetilde{Y}_i; \widetilde{\boldsymbol{\tau}}, \boldsymbol{\theta}) = \boldsymbol{S}_\lambda^{\mathrm{lin}}(\mathbf{X}_i, \varepsilon_i; \widetilde{\boldsymbol{\tau}}, \boldsymbol{\theta}) \,+\, \Delta \boldsymbol{S}_\lambda(\mathbf{X}_i, \vartheta_i), \tag{3.1}$$

where $\boldsymbol{\theta} := (\boldsymbol{r}^\top, \boldsymbol{\beta}^\top)^\top$. The first term on the right-hand side is the oracle (linear model) composite score,

$$\boldsymbol{S}_\lambda^{\mathrm{lin}}(\mathbf{X}_i, \varepsilon_i; \widetilde{\boldsymbol{\tau}}, \boldsymbol{\theta}) = \frac{1-\lambda}{L} \sum_{\ell=1}^{L} \mathbf{X}_i \big(\mathbb{I}\{\varepsilon_i \leqslant r_\ell\} - \tau_\ell\big) - \lambda\, \mathbf{X}_i\, \varepsilon_i, \tag{3.2}$$

and the second term is the plug-in remainder,

$$\Delta \boldsymbol{S}_\lambda(\mathbf{X}_i, \vartheta_i) = \lambda\, \mathbf{X}_i\, \vartheta_i \,+\, \frac{1-\lambda}{L} \sum_{\ell=1}^{L} \mathbf{X}_i \Big[\mathbb{I}\{\varepsilon_i \leqslant r_\ell + \vartheta_i\} - \mathbb{I}\{\varepsilon_i \leqslant r_\ell\}\Big]. \tag{3.3}$$

To facilitate theoretical analysis, we impose the following assumptions.

Assumption A (Sampling and design). The observations $\{(\mathbf{X}_i, \mathbf{Z}_i, U_i)\}_{i=1}^n$ are i.i.d., where $\mathbf{X}_i \in \mathbb{R}^d$ denotes the parametric regressors and $\mathbf{Z}_i \in \mathbb{R}^q$ the varying–coefficient regressors. Let $\mathbf{X} = [\mathbf{X}_1^\top; \ldots; \mathbf{X}_n^\top] \in \mathbb{R}^{n\times d}$ denote the design matrix for the parametric component, and write $\boldsymbol{\Sigma} = \mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top]$. Assume that there exist constants $\kappa_1, \kappa_2 > 0$ such that $\lambda_{\min}(\boldsymbol{\Sigma}) \geqslant \kappa_1$ and $\lambda_{\max}(\boldsymbol{\Sigma}) \leqslant \kappa_2 < \infty$. Moreover, there exists $M \geqslant 1$ such that $\|\mathbf{X}_i\|_\infty \leqslant M$ almost surely. The covariate U_i has compact support $\mathcal{U} \subset \mathbb{R}$.

Assumption B (Errors and density). Define $\varepsilon_i := Y_i - \eta(U_i) - \mathbf{X}_i^\top \boldsymbol{\beta}$.

(**B**.1) Assume $\{\varepsilon_i\}_{i=1}^n$ are i.i.d., independent of $(\mathbf{X}_i, \mathbf{Z}_i, U_i)$, with $\mathbb{E}[\varepsilon_i] = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma_\varepsilon^2 \in (0, \infty)$.

(**B**.2) Let F_ε and f_ε denote the cumulative distribution function and density of ε, respectively, with f_ε continuously differentiable on $\mathbb{R}$. Assume that there exist constants $c_f, C_f, C_f' > 0$ such that

$$\sup_{t\in\mathbb{R}} |f_\varepsilon(t)| \leqslant C_f, \qquad \sup_{t\in\mathbb{R}} |f_\varepsilon'(t)| \leqslant C_f', \qquad \inf_{1\leqslant\ell\leqslant L} f_\varepsilon(r_\ell^*) \geqslant c_f.$$

and, F_ε is globally Lipschitz with constant C_f, i.e., $|F_\varepsilon(u) - F_\varepsilon(v)| \leqslant C_f\, |u - v|$ for all $u, v \in \mathbb{R}$.

Assumption C (Cross-fitting consistency). Let $\widetilde{\eta}$ be the out-of-fold cross-fitted estimator of $\eta(\cdot)$, constructed by using training folds independent of each evaluation observation $(Y_i, \mathbf{X}_i, \mathbf{Z}_i, U_i)$. Define

$$r_n := \|\widetilde{\eta} - \eta\|_{L_2} = \Big\{\mathbb{E}\big[(\widetilde{\eta}(U) - \eta(U))^2\big]\Big\}^{1/2}.$$

Assume that $r_n = o_p(1)$ and $\sqrt{n}\, r_n^2 \to 0$.

Remark 3.1. Assumption **C** is a standard nuisance-rate condition in modern semiparametric inference with cross-fitting; see, for example, [8] and the related double/debiased machine learning literature. Such conditions are imposed to ensure that the first-stage nuisance estimation converges sufficiently fast for valid second-stage inference. A concrete verification under one-dimensional local-linear smoothing is provided in Appendix C.

Assumption D (Moment constraints). Let $\mathcal{V} := \{\boldsymbol{\nu} \in \mathbb{R}^d : \|\boldsymbol{\nu}\|_2 = 1,\ \|\boldsymbol{\nu}\|_0 \leqslant s_0\}$.

(D.1) Assume that there exist constants $0 < c_{s_0} \leqslant C_{s_0}^{\mathrm{up}} < \infty$ such that, for all $\boldsymbol{\nu} \in \mathcal{V}$,

$$c_{s_0} \ \leqslant\ \mathbb{E}\big[(\boldsymbol{\nu}^\top \mathbf{X}_i)^2\big] \ \leqslant\ C_{s_0}^{\mathrm{up}}.$$

(D.2) Assume that there exists $C_{s_0} > 0$ such that, for all $\boldsymbol{\nu} \in \mathcal{V}$,

$$\mathbb{E}\big[(\boldsymbol{\nu}^\top \mathbf{X}_i\, \varepsilon_i)^2\big] \ \geqslant\ C_{s_0}, \qquad \mathbb{E}\big[(\boldsymbol{\nu}^\top \mathbf{X}_i\, e_i(\widetilde{\boldsymbol{\tau}}))^2\big] \ \geqslant\ C_{s_0},$$

and, in addition, $\inf_{j \leqslant d} \mathbb{E}[X_{ij}^2] \geqslant C_{s_0}$.

(D.3) Assume that there exists a positive constant $\widetilde{C} < \infty$ such that $\mathbb{E}|\varepsilon_i|^{2+k} \leqslant \widetilde{C}^{\,k}$ for $k = 1, 2$.

(D.4) Assume that $\mathbb{E}\|\mathbf{X}_i\|_2^2 < \infty$ and $\mathbb{E}\|\mathbf{Z}_i\|_2^{2+\delta} < \infty$ for some $\delta > 0$.

Assumption E (Parameter space). Assume the following holds:

(**E**.1) There exist constants $0 < K_1 < \frac{1}{7}$ and $0 < K_2 < \frac{1}{6}$ such that $s_0^3 \log(dn) = O\big(n^{K_1}\big)$, and $s_0^4 \log(dn) = O\big(n^{K_2}\big)$.

(**E**.2) The overall sparsity $s = \|\boldsymbol{\beta}^{(1)}\|_0 + \|\boldsymbol{\beta}^{(2)}\|_0$ satisfies $\frac{s_0^2\, s^3\ \log^3(dn)}{n} \longrightarrow 0$ as $n, d \to \infty$.

(**E**.3) There is a constant $C_{\boldsymbol{\beta}} > 0$ such that $\|\boldsymbol{\beta}^{(1)}\|_\infty \vee \|\boldsymbol{\beta}^{(2)}\|_\infty \leqslant C_\beta$, and a constant $C_\Delta > 0$ with $\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_1 \leqslant C_\Delta$.

(**E**.4) The tuning parameters $\{\kappa_\lambda\}$ in (2.21) are chosen such that $\kappa_\lambda = C_\lambda \sqrt{\frac{\log(dn)}{n}}$ for some $C_\lambda > 0$.

Assumption F (Signal-shift). Let $\boldsymbol{\Delta} = \{j : \beta_j^{(1)} \neq \beta_j^{(2)}\}$ denote the set of coordinates with a change, and define the shift vector $\boldsymbol{\delta} \in \mathbb{R}^d$ by $\delta_j = \beta_j^{(2)} - \beta_j^{(1)}$. Define $\delta_{\min} = \min_{j \in \boldsymbol{\Delta}} |\delta_j|$ and $\delta_{\max} = \max_{j \in \boldsymbol{\Delta}} |\delta_j|$. There exist positive constants $0 < c \leqslant C < \infty$ such that

$$0 \ <\ c \ \leqslant\ \liminf_{d \to \infty} \frac{\delta_{\min}}{\delta_{\max}} \ \leqslant\ \limsup_{d \to \infty} \frac{\delta_{\max}}{\delta_{\min}} \ \leqslant\ C \ <\ \infty.$$

We briefly discuss the roles of the above assumptions below.

Assumption **A** specifies regularity conditions on the design and sampling scheme. In particular, the non-degeneracy of the covariance matrix $\boldsymbol{\Sigma}$ ensures that the parametric component is identifiable and that the associated high-dimensional estimation and inference procedures are well behaved. The boundedness condition on $\mathbf{X}_i$ and the compact support of U_i are standard in semiparametric models and facilitate uniform control of empirical processes. Assumption **B** concerns the distributional properties of the error term. Assumption **B.1** imposes independence and finite variance, which are essential for establishing the asymptotic behavior of the CUSUM-type processes under both H_0 and H_1. Assumption **B.2** further requires smoothness and boundedness of the error density, which is mainly used to control the bias introduced by quantile-based score functions and to justify the Gaussian and bootstrap approximations for the individual and adaptive tests. Assumption **C** characterizes the accuracy of the cross-fitted estimator $\widetilde{\eta}$ for the nonparametric component. The condition $r_n = o_p(1)$ guarantees consistency, while the stronger requirement $\sqrt{n}\, r_n^2 \to 0$ ensures that the plug-in effect arising from estimating $\eta(\cdot)$ is asymptotically negligible relative to the stochastic fluctuations of the CUSUM process. Assumption **D** imposes moment and non-degeneracy conditions on sparse linear combinations of the covariates. These conditions ensure that the $(s_0, 2)$-norm–based test statistics are well defined and that their variances are bounded away from zero. In particular, Assumptions **D.1**–**D.3** provide the basic moment conditions required for high-dimensional Gaussian and bootstrap approximations, while Assumption **D.4** controls the overall second and higher-order moments of the covariates. Assumption **E** specifies the growth rates of the dimensional and sparsity parameters. Assumptions **E.1** and **E.2** regulate the relationship between (n, d, s, s_0) and allow the ambient dimension d to grow much faster than the sample size n, provided that the effective sparsity remains moderate. Assumptions **E.3** and **E.4** impose mild boundedness conditions on the regression coefficients and the regularization parameters, which are standard in high-dimensional change-point analysis and are mainly used to control the estimation error of the Lasso-type estimators. Finally,

Assumption **F** characterizes the signal structure under the alternative hypothesis. By requiring the relative magnitudes of the nonzero signal shifts to be bounded away from zero and infinity, this assumption avoids pathological cases where the change is either too weak or dominated by a single coordinate, and it guarantees stable detection power for the proposed $(s_0, 2)$-norm–based tests.

Definition 3.2. For $t \in (0,1)$ and a fixed $\lambda \in [0,1]$, define the plug-in partial sum and CUSUM processes by

$$\widetilde{\boldsymbol{S}}_\lambda(t) := \sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{S}_\lambda^{\mathrm{plug}}(\widetilde{Y}_i, \mathbf{X}_i; \widetilde{\boldsymbol{\tau}}, \boldsymbol{\theta}), \widetilde{\boldsymbol{G}}_\lambda(t) := \frac{1}{\sqrt{n}\,\sigma(\lambda, \widetilde{\boldsymbol{\tau}})}\Big[\widetilde{\boldsymbol{S}}_\lambda(t) - \tfrac{\lfloor nt \rfloor}{n}\,\widetilde{\boldsymbol{S}}_\lambda(1)\Big],$$

where $\boldsymbol{S}_\lambda^{\mathrm{plug}}$ is given in (3.1)-(3.3) and $\sigma^2(\lambda, \widetilde{\boldsymbol{\tau}}) := \mathrm{Var}\{(1-\lambda)e_i(\widetilde{\boldsymbol{\tau}}) - \lambda\,\varepsilon_i\}$ with $e_i(\widetilde{\boldsymbol{\tau}}) = \frac{1}{L}\sum_{\ell=1}^{L}(\mathbb{I}\{\varepsilon_i \leqslant r_\ell^*\} - \tau_\ell)$ and r_ℓ^* the τ_ℓ-quantile of ε. Similarly, define the oracle (linear model) counterparts as

$$\boldsymbol{S}_\lambda^{\mathrm{lin}}(t) := \sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{S}_\lambda^{\mathrm{lin}}(\mathbf{X}_i, \varepsilon_i; \widetilde{\boldsymbol{\tau}}, \boldsymbol{\theta}), \boldsymbol{G}_\lambda^{\mathrm{lin}}(t) := \frac{1}{\sqrt{n}\,\sigma(\lambda, \widetilde{\boldsymbol{\tau}})}\Big[\boldsymbol{S}_\lambda^{\mathrm{lin}}(t) - \tfrac{\lfloor nt \rfloor}{n}\,\boldsymbol{S}_\lambda^{\mathrm{lin}}(1)\Big],$$

where $\boldsymbol{S}_\lambda^{\mathrm{lin}}$ is defined in (3.2).

Before analyzing the asymptotic size and power of the individual test $\Theta_{\lambda,\alpha}$ and the tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$, we first show that the standardized feasible plug-in CUSUM process $\widetilde{\boldsymbol{G}}_\lambda(t)$ is uniformly close to the oracle (linear model) CUSUM process of [26]. Specifically, Corollary 3.5 establishes that $\sup_{t\in[q_0,1-q_0]} \|\widetilde{\boldsymbol{G}}_\lambda(t) - \boldsymbol{G}_\lambda^{\mathrm{lin}}(t)\|_2 = o_p(1)$ and, consequently, $\widetilde{\Gamma}_\lambda - \Gamma_\lambda^{\mathrm{lin}} = o_p(1)$.

Theorem 3.3. *Let $\Delta\boldsymbol{S}_\lambda(\mathbf{X}_i, \vartheta_i)$ be as in* (3.3) *and $a_{i,n}(t) := \mathbb{I}\{i \leqslant \lfloor nt \rfloor\} - \lfloor nt \rfloor/n$. Fix $q_0 \in (0, 1/2)$ and $\lambda \in [0,1]$. Under Assumptions **A**–**D**, we have*

$$\sup_{t\in[q_0,1-q_0]} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_{i,n}(t)\,\Delta\boldsymbol{S}_\lambda(\mathbf{X}_i, \vartheta_i) \right\|_2 = o_p(1) \ \text{as } n \to \infty. \tag{3.4}$$

Remark 3.4. A rate-refined bound is

$$\sup_{t\in[q_0,1-q_0]} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_{i,n}(t)\,\Delta\boldsymbol{S}_\lambda(\mathbf{X}_i, \vartheta_i) \right\|_2 = O_p\big(r_n^{1/2} \ \vee \ \sqrt{n}\, r_n^2\big),$$

uniformly in $t \in [q_0, 1-q_0]$ for each fixed $\lambda \in [0,1]$. Under Assumption **C** (namely $r_n = o_p(1)$ and $\sqrt{n}\, r_n^2 \to 0$), this bound is $o_p(1)$. In particular, for one-dimensional local-linear smoothing with bandwidth $h \asymp n^{-1/5}$, we have $r_n = O_p(n^{-2/5})$ and therefore the left-hand side is $O_p(n^{-1/5}) = o_p(1)$. This condition is mild and is satisfied by standard nonparametric estimators; see Appendix C.

Corollary 3.5. *For any given $q_0 \in (0, 1/2)$ and $\lambda \in [0,1]$, and under the conditions of Theorem 3.3, as $n \to \infty$, we have*

$$\sup_{t\in[q_0,1-q_0]} \big\|\widetilde{\boldsymbol{G}}_\lambda(t) - \boldsymbol{G}_\lambda^{\mathrm{lin}}(t)\big\|_2 = o_p(1), \ \textit{and}\ \widetilde{\Gamma}_\lambda - \Gamma_\lambda^{\mathrm{lin}} = o_p(1),$$

where $\widetilde{\Gamma}_\lambda := \max_{t\in[q_0,1-q_0]} \|\widetilde{\boldsymbol{G}}_\lambda(t)\|_{(s_0,2)}$ and $\Gamma_\lambda^{\mathrm{lin}} := \max_{t\in[q_0,1-q_0]} \|\boldsymbol{G}_\lambda^{\mathrm{lin}}(t)\|_{(s_0,2)}$.

Proof. By the decomposition $\boldsymbol{S}_\lambda^{\mathrm{plug}} = \boldsymbol{S}_\lambda^{\mathrm{lin}} + \Delta\boldsymbol{S}_\lambda$ and the definition of the CUSUM processes,

$$\widetilde{\boldsymbol{G}}_\lambda(t) - \boldsymbol{G}_\lambda^{\mathrm{lin}}(t) = \frac{1}{\sqrt{n}\,\sigma(\lambda, \widetilde{\boldsymbol{\tau}})} \sum_{i=1}^{n} a_{i,n}(t)\,\Delta\boldsymbol{S}_\lambda(\mathbf{X}_i, \vartheta_i),$$

where $a_{i,n}(t) = \mathbb{I}\{i \leqslant \lfloor nt \rfloor\} - \lfloor nt \rfloor/n$ and $\sigma^2(\lambda, \widetilde{\boldsymbol{\tau}}) = \mathrm{Var}\big((1-\lambda)e_i(\widetilde{\boldsymbol{\tau}}) - \lambda\,\varepsilon_i\big)$ is the variance factor used to standardize the processes. Assumption **B** implies that $\sigma^2(\lambda, \widetilde{\boldsymbol{\tau}})$ is finite and bounded away from zero (uniformly over fixed λ), and hence $1/\sigma(\lambda, \widetilde{\boldsymbol{\tau}}) = O(1)$. Theorem 3.3 then yields

$$\sup_{t\in[q_0,1-q_0]} \big\|\widetilde{\boldsymbol{G}}_\lambda(t) - \boldsymbol{G}_\lambda^{\mathrm{lin}}(t)\big\|_2 = o_p(1).$$

For any $u, v \in \mathbb{R}^d$, we have $\big|\|u\|_{(s_0,2)} - \|v\|_{(s_0,2)}\big| \leqslant \|u - v\|_{(s_0,2)} \leqslant \|u - v\|_2$. Thus, for the maximal statistics, we have

$$\big|\widetilde{\Gamma}_\lambda - \Gamma_\lambda^{\mathrm{lin}}\big| \leqslant \sup_{t\in[q_0,1-q_0]} \big\|\widetilde{\boldsymbol{G}}_\lambda(t) - \boldsymbol{G}_\lambda^{\mathrm{lin}}(t)\big\|_2 = o_p(1),$$

which completes the proof. □

Corollary 3.5 shows that our feasible plug-in CUSUM process $\widetilde{\boldsymbol{G}}_\lambda(t)$ and its associated statistic $\widetilde{\Gamma}_\lambda$ approximate the oracle counterparts $\boldsymbol{G}_\lambda^{\mathrm{lin}}(t)$ and $\Gamma_\lambda^{\mathrm{lin}}$ of [26] to within $o_p(1)$ uniformly over $t \in [q_0, 1-q_0]$.

Proposition 3.6. *Assume that Assumptions* **A**–**E** *hold, and any given* $\lambda \in [0,1]$ *and significance level* $\alpha \in (0,1)$. *Then, under* H_0 *and as* $n, d, B \to \infty$, *we have*

$$\sup_{t>0}\big|\mathbb{P}(\widetilde{\Gamma}_\lambda \leqslant t) - \mathbb{P}(\widetilde{\Gamma}_\lambda^b \leqslant t)\big| = o_p(1)$$

and

$$\mathbb{P}(\Theta_{\lambda,\alpha} = 1) \longrightarrow \alpha.$$

Proposition 3.6 demonstrates that, for each $\lambda \in [0,1]$, the limiting null distribution of $\widetilde{\Gamma}_\lambda$ can be uniformly approximated by that of its bootstrap counterpart $\widetilde{\Gamma}_\lambda^b$, and that, for any given significance level $\alpha \in (0,1)$, the bootstrap-based individual test $\Theta_{\lambda,\alpha}$ asymptotically controls the Type I error.

Having established the size properties under H_0, we now turn to the behavior of the individual test under H_1. In particular, Proposition 3.7 provides a non-asymptotic error bound for the argmax-based estimator $\widehat{t}_\lambda$ defined in (2.22) for each $\lambda \in [0,1]$, and Corollary 3.8 further implies that our individual test can reject H_0 with probability tending to one.

Proposition 3.7. *Let* $\boldsymbol{\delta} = \boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}$, *and assume* $\|\boldsymbol{\delta}\|_{(s_0,2)} \gg \sqrt{\log(dn)/n}$. *Define the signal-to-noise ratio*

$$\mathrm{SNR}(\lambda, \widetilde{\boldsymbol{\tau}}) = \frac{(1-\lambda)\left(\frac{1}{L}\sum_{l=1}^{L} f_\varepsilon(r_l^*)\right) + \lambda}{\sigma(\lambda, \widetilde{\boldsymbol{\tau}})}, \lambda \in [0,1].$$

In addition to Assumptions **A**, **D**, **E**.2–**E**.4, *and* **F**, *suppose*:

(a) *if* $\lambda = 1$, *then Assumptions* **B**.1, **D**.2, *and* **D**.3 *hold, as well as* $n^{1/4} = o(s)$ *holds;*

(b) *if* $\lambda = 0$, *then Assumptions* **B**.2 *and* **D**.2 *hold, as well as* $s_0^{1/2} s^2 \sqrt{\frac{\log(dn)}{n}}\, \|\boldsymbol{\delta}\|_{(s_0,2)} \to 0$ *holds;*

(c) *if* $\lambda \in (0,1)$, *then Assumptions* **B**, **D**.2, *and* **D**.3 *hold, as well as* $s_0^{1/2} s^2 \sqrt{\frac{\log(dn)}{n}}\, \|\boldsymbol{\delta}\|_{(s_0,2)} \to 0$ *and* $n^{1/4} = o(s)$ *hold.*

Then, under H_1 *and as* $n, d \to \infty$, *with probability tending to one, we have*

$$\big|\widehat{t}_\lambda - t_0\big| \;\leqslant\; C^*(s_0, \lambda, \widetilde{\boldsymbol{\tau}})\, \frac{\log(dn)}{n\,\mathrm{SNR}(\lambda, \widetilde{\boldsymbol{\tau}})^2\, \|\boldsymbol{\delta}\|_{(s_0,2)}^2}, \lambda \in [0,1]$$

where t_0 *is the true change-point and* $C^*(s_0, \lambda, \widetilde{\boldsymbol{\tau}})$ *is a positive constant.*

Proposition 3.7 shows that, under the signal condition $\|\boldsymbol{\delta}\|_{(s_0,2)} \gg \sqrt{\log(dn)/n}$ and given the signal-to-noise ratio $\mathrm{SNR}(\lambda, \widetilde{\boldsymbol{\tau}})$, the argmax-based estimator $\widehat{t}_\lambda$ is consistent for the true change-point t_0 and attains the optimal convergence rate up to a $\log(dn)$ factor.

Proposition 3.8. *Let* $\boldsymbol{\Omega} = (\Omega_1, \ldots, \Omega_d)^\top$ *with* $\Omega_j = \mathrm{SNR}(\lambda, \widetilde{\boldsymbol{\tau}})\big|t_0(1-t_0)\,[\boldsymbol{\Sigma}\,\boldsymbol{\delta}]_j\big|\,\mathbb{I}\{j \in \boldsymbol{\Delta}\}$, *where* $\boldsymbol{\Delta} = \{j : \delta_j \neq 0\}$. *Define*

$$\epsilon_n = O\Big(s_0^{1/2} s \sqrt{\tfrac{\log(dn)}{n}}\Big) \;\vee\; O\Big(s_0^{1/2} s^2 \sqrt{\tfrac{\log(dn)}{n}}\, \|\boldsymbol{\delta}\|_{(s_0,2)}\Big).$$

Assume, in addition to Assumptions **A**, **D**, *and* **E**.2–**E**.4, *suppose*:

(a) *if* $\lambda = 1$, *then Assumptions* **B**.1, **D**.2, *and* **D**.3 *hold;*

(b) *if* $\lambda = 0$, *then Assumptions* **B**.2 *and* **D**.2 *hold, as well as* $s_0^{1/2} s^2 \sqrt{\frac{\log(dn)}{n}}\, \|\boldsymbol{\delta}\|_{(s_0,2)} \to 0$ *holds;*

(c) *if* $\lambda \in (0,1)$, *then Assumptions* **B**, **D**.2, *and* **D**.3 *hold, as well as* $s_0^{1/2} s^2 \sqrt{\frac{\log(dn)}{n}}\, \|\boldsymbol{\delta}\|_{(s_0,2)} \to 0$ *and* $n^{1/4} = o(s)$ *hold.*

Under H_1*, if* $\boldsymbol{\Omega}$ *satisfies*

$$\sqrt{n}\,\|\boldsymbol{\Omega}\|_{(s_0,2)} \;\geqslant\; \frac{C(\lambda,\widetilde{\boldsymbol{\tau}})}{1-\epsilon_n}\, s_0^{1/2}\,\big(\sqrt{\log(dn)}+\sqrt{\log(1/\alpha)}\big),$$

for some constant $C(\lambda,\widetilde{\boldsymbol{\tau}})>0$*, then, for any given significance level* $\alpha\in(0,1)$*, as* $n,d,B\to\infty$*, we have*

$$\mathbb{P}(\Theta_{\lambda,\alpha}=1)\longrightarrow 1, \lambda\in[0,1].$$

Proposition 3.8 shows that, for any fixed significance level $\alpha\in(0,1)$ and each $\lambda\in[0,1]$, once the signal-to-noise ratio exceeds a suitable threshold, the bootstrap-based individual test $\Theta_{\lambda,\alpha}$ can detect the existence of a structural change in model (2.1) with probability tending to one.

Furthermore, under suitable regularity conditions, Proposition 3.9 establishes the asymptotic size and power properties of our tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$.

Proposition 3.9. *Assume that Assumptions* **A**–**D** *and* **E**.2–**E**.4 *hold and any given significance level* $\alpha\in(0,1)$.

(i) *Suppose further that, for any* $\varsigma\in(0,1)$*,* $h^{0.6}(\varsigma)s_0^3\log(dn)=o\big(n^{1/10}\big)$ *holds. Then under* H_0 *and as* $n,d,B\to\infty$*, we have*

$$\mathbb{P}(\Theta_{\mathrm{ad},\alpha}=1)\longrightarrow\alpha,\quad \widehat{P}_{\mathrm{ad}}-P_{\mathrm{ad}}=o_p(1),$$

where P_{ad} *denotes the theoretical p-value of* Γ_{ad}.

(ii) *Under* H_1*, define* ϵ_n *as in Proposition 3.8. If* $\boldsymbol{\Omega}$ *satisfies*

$$\sqrt{n}\,\|\boldsymbol{\Omega}\|_{(s_0,2)} \;\geqslant\; \frac{C(\lambda,\widetilde{\boldsymbol{\tau}})}{1-\epsilon_n}\, s_0^{1/2}\,\big(\sqrt{\log(dn)}+\sqrt{\log(|\mathcal{D}|/\alpha)}\big),$$

for some constant $C(\lambda,\widetilde{\boldsymbol{\tau}})>0$*, then, as* $n,d,B\to\infty$*, we have*

$$\mathbb{P}(\Theta_{\mathrm{ad},\alpha}=1)\longrightarrow 1.$$

Note that, by Proposition 3.7, our tail-adaptive change-point estimator $\widehat{t}_{\widehat{\lambda}}$ defined in (2.33) is also consistent under the alternative hypothesis H_1. Moreover, Propositions 3.6-3.9 follow directly from Theorems 3.2-3.9 of [26], whose proofs we refer to for details and therefore omit here for brevity.

4 Numerical studies

This section investigates the finite-sample performance of the proposed method under a general partially linear model with $\gamma(U)\neq 0$, allowing for both single and multiple change-point scenarios. Simulation results under a simplified model with a single change-point and $\gamma(U)=0$, which corresponds to a special case of the proposed framework, are reported in Appendix A.

We consider the following partially linear model:

$$Y_i=\mathbf{X}_i^\top\boldsymbol{\beta}^{(1)}\mathbb{I}\{i<l_0\}+\mathbf{X}_i^\top\boldsymbol{\beta}^{(2)}\mathbb{I}\{i\geqslant l_0\}+\alpha(U_i)Z_i+\gamma(U_i)+\varepsilon_i,\quad i=1,\ldots,n, \tag{4.1}$$

where $l_0=\lfloor n\,t_0\rfloor$. Since our primary interest lies in detecting structural breaks in the parametric component, we assume throughout that the nonparametric function remains time-invariant.

In our simulations, we set $\alpha(U_i)=\sin(2\pi U_i)$ and $\gamma(U_i)=\sin(4\pi U_i)$, where $U_i\sim\mathrm{Unif}(0,1)$. The binary covariate Z_i is generated independently of $\mathbf{X}_i$ and U_i, taking the value 0 with probability 0.2 and 1 with probability 0.8. Moreover, we generate the error term ε_i from various distributions—namely the standard normal distribution $N(0,1)$ and Student's t_ν distributions with degrees of freedom $\nu=3$ and $\nu=5$—to demonstrate the adaptability of our proposed method to different tail behavior. For the parametric component, we specify the regression coefficient vector $\boldsymbol{\beta}=(1,1,1,0,\ldots,0)^\top\in\mathbb{R}^d$, so that only the first three entries are nonzero (and equal to one). We then introduce a shift vector $\boldsymbol{\delta}=(\delta_1,\ldots,\delta_d)^\top\in\mathbb{R}^d$, where $\delta_j=C_\delta\sqrt{\log(d)/n}\,\mathbb{I}(1\leqslant j\leqslant 3), j=1,\ldots,d$, and $C_\delta>0$ controls the magnitude of the jump. Under the null hypothesis H_0, we have $\boldsymbol{\beta}^{(1)}=\boldsymbol{\beta}^{(2)}=\boldsymbol{\beta}$, whereas under the alternative hypothesis H_1, we have $\boldsymbol{\beta}^{(2)}=\boldsymbol{\beta}^{(1)}+\boldsymbol{\delta}$. For the design matrix $\mathbf{X}_i\in\mathbb{R}^d$, we assume $\mathbf{X}_i\overset{\text{i.i.d.}}{\sim}N\big(\mathbf{0},\boldsymbol{\Sigma}\big)$. We consider two cases for $\boldsymbol{\Sigma}$:

• **Scenario 1 (Toeplitz):** Following [26], we take $\mathbf{\Sigma} = \mathbf{\Sigma}^*$, where $\mathbf{\Sigma}^* = (\sigma_{ij}^*) \in \mathbb{R}^{d\times d}$ with $\sigma_{ij}^* = 0.8^{|\,i-j\,|}$ for $i, j = 1, \ldots, d$.

• **Scenario 2 (Compound Symmetric):** We take $\mathbf{\Sigma} = \mathbf{\Sigma}^{'}$, where $\mathbf{\Sigma}^{'} = (\sigma_{ij}^{'}) \in \mathbb{R}^{d\times d}$ with $\sigma_{ij}^{'} = 0.3\,\mathbb{I}(i \neq j) + \mathbb{I}(i = j)$ for $i, j = 1, \ldots, d$. Here $\mathbb{I}(\cdot)$ denotes the indicator function.

Throughout these experiments, we set the sample size to $n = 100$, the dimension to $d = 10$, and the number of bootstrap replications to $B = 100$. In Algorithm 1, we employ the Epanechnikov kernel $\mathcal{K}(u) = 0.75\,(1 - u^2)_+$, the quantile level $\tau = 0.5$, and $L = 1$. In addition, both the bandwidth h in Algorithm 1 and the tuning parameter κ_λ in (2.19) are selected by the cross-validation technique [35]. The optimization problems arising in our proposed method are solved by using the L-BFGS-B algorithm proposed by [7].

4.1 Size performance

In this subsection, we consider the empirical size of our proposed tests at the significance level $\alpha = 0.05$. Table 1 summarizes the size results of the individual test $\Theta_{\lambda,\alpha}$ with $\lambda \in \{0, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ and the tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$ under both Scenario 1 and Scenario 2, across three error distributions. To examine the impact of the number of active coefficients, we vary $s_0 \in \{1, 2, 3, 4, 5\}$. All results are based on 500 replications, each employing $B = 100$ bootstrap resamples as in Subsection 2.3.

As reported in Table 1, for all considered values of s_0, the proposed individual tests $\Theta_{\lambda,\alpha}$ with different choices of λ, as well as the tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$, exhibit empirical sizes that are well aligned with the nominal significance level $\alpha = 0.05$. This pattern holds uniformly across the two covariance structures (Scenario 1 and Scenario 2) and across all three error distributions, indicating that the bootstrap calibration procedure provides an accurate approximation to the null distributions.

More specifically, for the individual test $\Theta_{\lambda,\alpha}$, the empirical sizes fluctuate mildly around 0.05 as λ varies. Such fluctuations are observed under all error distributions, including the heavy-tailed t_3 distribution, the moderately heavy-tailed t_5 distribution, and the Gaussian distribution. Importantly, these variations do not display any systematic inflation or deflation of the rejection probability. Even in the presence of heavy tails or stronger dependence, the empirical sizes remain within a narrow range around the nominal level, suggesting satisfactory finite-sample size control.

In comparison, the tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$ exhibits particularly stable empirical size control. Across different values of λ, error distributions, and covariance structures, its empirical sizes consistently remain close to 0.05, with less variability than those of the individual tests. This indicates that aggregating information across multiple tuning parameters through the adaptive strategy does not compromise size accuracy, and instead leads to more uniform performance.

Moreover, the sparsity parameter s_0, which determines the number of active components in the $(s_0, 2)$-norm test statistic, has little impact on the empirical size. As s_0 increases from 1 to 5, both the individual and adaptive tests exhibit stable size performance in both scenarios. For a fixed λ and a fixed error distribution, the differences in empirical sizes between Scenario 1 and Scenario 2 are minor, indicating that the proposed tests are robust to changes in the underlying covariance structure.

Overall, the simulation results in Table 1 demonstrate that the proposed testing procedures achieve reliable control of the nominal significance level over a wide range of distributional assumptions, dependence structures, and sparsity levels. The tail-adaptive test further improves stability across tuning parameters, providing a robust and practically appealing alternative to individual tests in partially linear models.

4.2 Power performance

In this subsection, we further investigate the finite-sample power of the proposed tests. The experimental settings follow those in Section 4.1, except that the relative change-point location is fixed at $t_0 = 0.5$ and we consider two signal sizes $C_\delta \in \{1.5, 2.5, 3.5, 4.5, 5\}$. We evaluate the individual tests $\Theta_{\lambda,\alpha}$ with $\lambda \in \{0, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ and the tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$. Empirical powers are computed based on

Table 1 Empirical sizes of the individual test $\Theta_{\lambda,\alpha}$ with $\lambda \in \{0, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ and the tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$ under Toeplitz and Compound Symmetric covariance structures for $s_0 \in \{1, 2, 3, 4, 5\}$, at the significance level $\alpha = 0.05$. The results are based on 500 replications with $B = 100$ bootstrap draws per replication.

Empirical sizes for Scenario 1									
Dist.	s_0	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 0.3$	$\lambda = 0.5$	$\lambda = 0.7$	$\lambda = 0.9$	$\lambda = 1$	Adaptive
t_3	1	0.046	0.056	0.054	0.050	0.046	0.048	0.042	0.044
	2	0.044	0.044	0.046	0.048	0.060	0.044	0.042	0.052
	3	0.046	0.048	0.066	0.056	0.046	0.048	0.058	0.048
	4	0.042	0.044	0.058	0.038	0.046	0.044	0.050	0.046
	5	0.048	0.046	0.050	0.048	0.046	0.058	0.054	0.052
t_5	1	0.056	0.052	0.052	0.050	0.046	0.042	0.044	0.054
	2	0.050	0.046	0.054	0.056	0.060	0.054	0.052	0.046
	3	0.066	0.046	0.052	0.052	0.050	0.048	0.058	0.048
	4	0.052	0.040	0.056	0.046	0.048	0.044	0.048	0.040
	5	0.040	0.054	0.052	0.058	0.048	0.060	0.040	0.042
$N(0,1)$	1	0.044	0.052	0.042	0.044	0.054	0.046	0.048	0.046
	2	0.054	0.060	0.052	0.054	0.056	0.040	0.044	0.048
	3	0.048	0.038	0.044	0.042	0.046	0.050	0.040	0.036
	4	0.054	0.046	0.062	0.042	0.048	0.050	0.048	0.050
	5	0.056	0.054	0.058	0.052	0.052	0.050	0.040	0.048
Empirical sizes for Scenario 2									
Dist.	s_0	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 0.3$	$\lambda = 0.5$	$\lambda = 0.7$	$\lambda = 0.9$	$\lambda = 1$	Adaptive
t_3	1	0.044	0.054	0.050	0.060	0.046	0.048	0.052	0.056
	2	0.040	0.042	0.046	0.050	0.054	0.052	0.052	0.048
	3	0.052	0.056	0.044	0.036	0.046	0.048	0.046	0.040
	4	0.052	0.054	0.042	0.056	0.044	0.046	0.048	0.054
	5	0.048	0.052	0.056	0.052	0.048	0.046	0.062	0.044
t_5	1	0.056	0.054	0.040	0.046	0.052	0.054	0.052	0.050
	2	0.054	0.050	0.042	0.044	0.052	0.056	0.054	0.050
	3	0.048	0.064	0.054	0.056	0.040	0.042	0.048	0.052
	4	0.062	0.054	0.044	0.046	0.058	0.048	0.044	0.054
	5	0.050	0.048	0.056	0.064	0.054	0.056	0.040	0.046
$N(0,1)$	1	0.062	0.030	0.044	0.048	0.046	0.046	0.044	0.048
	2	0.054	0.060	0.056	0.058	0.064	0.046	0.048	0.050
	3	0.052	0.058	0.046	0.064	0.040	0.044	0.050	0.044
	4	0.044	0.054	0.062	0.058	0.056	0.038	0.048	0.044
	5	0.052	0.042	0.052	0.040	0.056	0.044	0.046	0.046

500 replications with $B = 100$ bootstrap draws per replication. Figures 1 and 2 summarize all empirical power results. More specifically, the main conclusions are as follows:

First, the empirical power increases monotonically with the signal strength C_δ across all error distributions, covariance structures, and sparsity levels. In particular, for both Toeplitz and Compound Symmetric covariance designs, the power at $C_\delta = 5$ is uniformly higher than that at $C_\delta = 1.5$. Under Gaussian errors, most procedures achieve power close to one when C_δ is sufficiently large, which is consistent with the theoretical detection guarantees established in Section 3.

Second, the presence of heavy-tailed errors leads to a noticeable loss of power compared with the Gaussian case, especially for small to moderate signal strengths. This effect is most pronounced under t_3 errors. Nevertheless, the tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$ consistently maintains competitive and often superior power relative to the individual tests in these settings, demonstrating robustness to heterogeneous tail behavior. This advantage becomes more evident as the signal strength increases, where the adaptive procedure rapidly approaches the upper envelope of the individual tests.

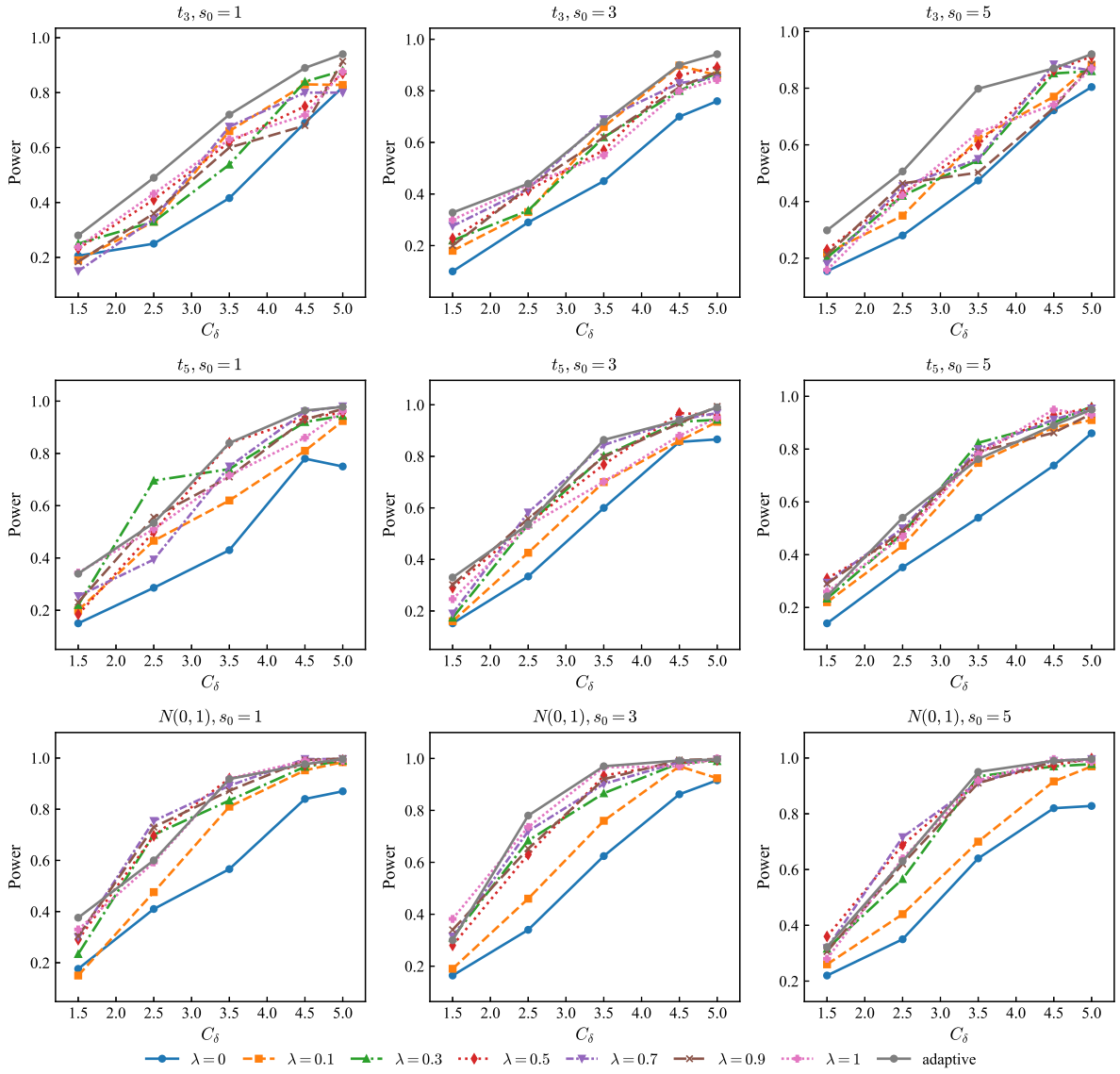

Figure 1 Empirical power of the individual tests $\Theta_{\lambda,\alpha}$ with $\lambda \in \{0, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ and the tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$ under the Toeplitz covariance structure. From left to right, columns correspond to $s_0 = 1, 3, 5$, and from top to bottom, rows correspond to error distributions t_3, t_5, and $N(0,1)$. The change-point location is $t_0 = 0.5$, and the nominal level is $\alpha = 0.05$. Results are based on 500 replications with $B = 100$ bootstrap draws per replication. (Color online)

Third, the sparsity level s_0 plays an important role in detection performance. When the signal is strong, changes involving a larger number of coordinates are generally easier to detect, leading to higher power for larger s_0. For weaker signals, the relationship between power and s_0 is less monotone, reflecting finite-sample variability. Despite this, the proposed $(s_0, 2)$-norm aggregation strategy remains effective: for moderate to large C_δ, power typically improves with increasing s_0 under both covariance structures.

Fourth, the overall ranking of the procedures is qualitatively similar under the Toeplitz and Compound Symmetric covariance designs. This indicates that the proposed tests are not overly sensitive to moderate changes in the dependence structure of the covariates. While a well-tuned individual test can occasionally match or slightly outperform the adaptive test under Gaussian errors, such dominance is not uniform across error distributions or sparsity levels.

Finally, no single fixed value of λ yields uniformly optimal power across all scenarios. The performance of the individual tests depends critically on the error distribution, the covariance structure, and the sparsity level, all of which are typically unknown in practice. By contrast, the tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$

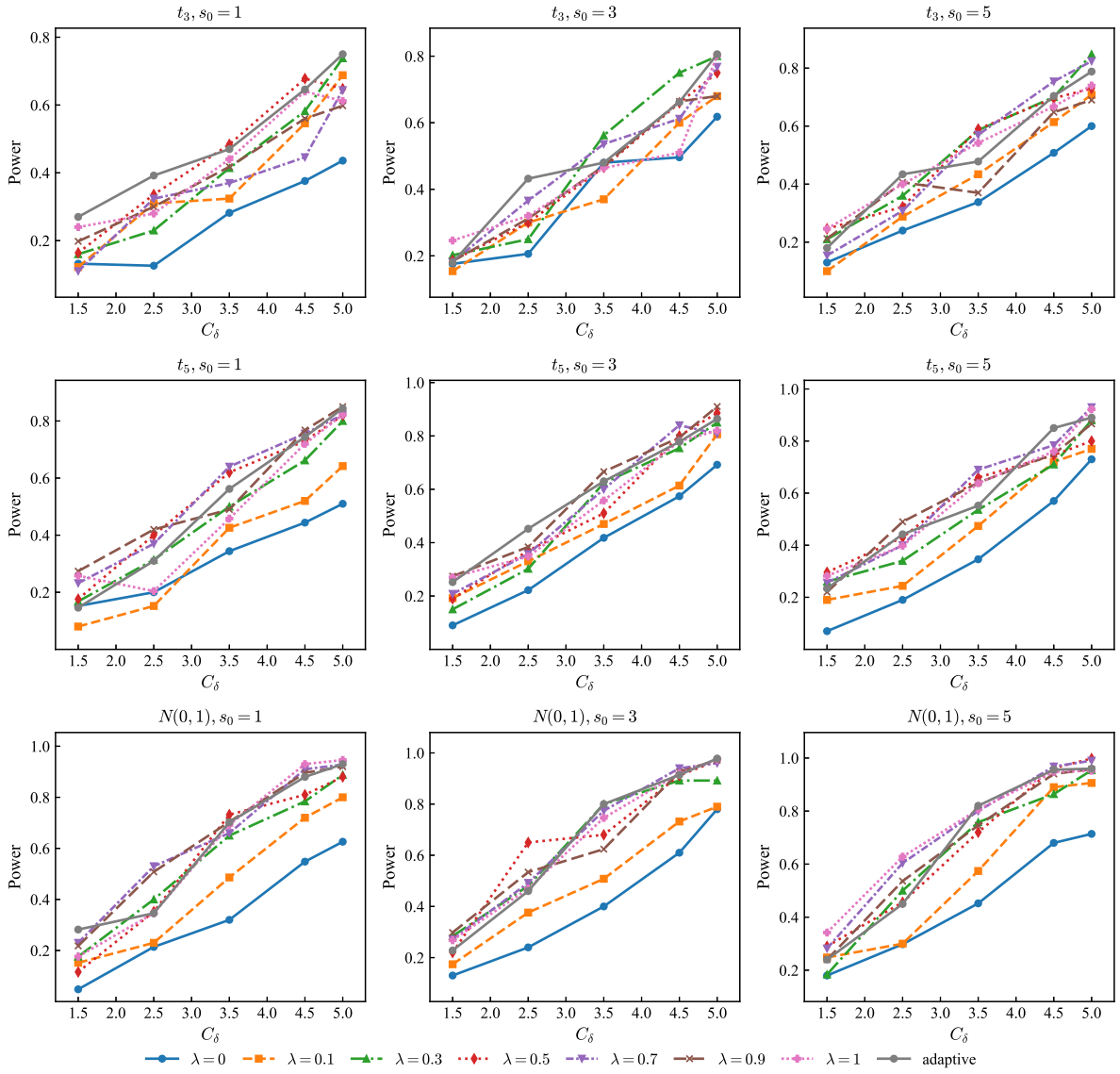

Figure 2 Empirical power of the individual tests $\Theta_{\lambda,\alpha}$ with $\lambda \in \{0, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ and the tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$ under the Compound Symmetric covariance structure. From left to right, columns correspond to $s_0 = 1, 3, 5$, and from top to bottom, rows correspond to error distributions t_3, t_5, and $N(0,1)$. The change-point location is $t_0 = 0.5$, and the nominal level is $\alpha = 0.05$. Results are based on 500 replications with $B = 100$ bootstrap draws per replication. (Color online)

effectively tracks the best-performing individual test across a wide range of designs, particularly under non-Gaussian errors, thereby avoiding the need for tuning parameter selection.

From a practical perspective, these results suggest that the tail-adaptive test provides a reliable default choice with strong power across heterogeneous settings. When there is credible prior knowledge indicating light-tailed errors and weak dependence, an individual test with a moderate value of λ (e.g., $\lambda \in [0.3, 0.7]$) may offer marginal gains. Otherwise, $\Theta_{\mathrm{ad},\alpha}$ delivers stable and robust performance without requiring additional tuning.

Overall, the numerical results demonstrate that the proposed testing procedures achieve high power over a broad range of signal strengths, error distributions, covariance structures, and sparsity levels, with the adaptive procedure in particular providing consistently reliable performance.

4.3 Multiple change-points detection

We next investigate the finite-sample performance of the proposed procedure in a multiple change-point scenario. Throughout this experiment, we set the sample size to $n = 200$ and the dimension to $d = 10$. Specifically, we consider a partially linear model with two structural breaks in the parametric component as follows:

$$Y_i = \begin{cases} \mathbf{X}_i^\top \boldsymbol{\beta}^{(1)} + \alpha(U_i)Z_i + \gamma(U_i) + \varepsilon_i,\ 1 \leqslant i \leqslant l_1, \\ \mathbf{X}_i^\top \boldsymbol{\beta}^{(2)} + \alpha(U_i)Z_i + \gamma(U_i) + \varepsilon_i,\ l_1 + 1 \leqslant i \leqslant l_2, \\ \mathbf{X}_i^\top \boldsymbol{\beta}^{(3)} + \alpha(U_i)Z_i + \gamma(U_i) + \varepsilon_i,\ l_2 + 1 \leqslant i \leqslant n, \end{cases} \tag{4.2}$$

where the true change-point locations are set to $l_1 = 66$ and $l_2 = 133$. The covariates $\mathbf{X}_i \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is specified as in Scenario 1 (Toeplitz) or Scenario 2 (Compound Symmetric). The regression coefficients satisfy $\boldsymbol{\beta}^{(2)} = \boldsymbol{\beta}^{(1)} + \boldsymbol{\delta}, \boldsymbol{\beta}^{(3)} = \boldsymbol{\beta}^{(1)}$, where $\boldsymbol{\beta}^{(1)} = (1, 1, 1, 0, \ldots, 0)^\top \in \mathbb{R}^d$, and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_d)^\top$ is defined as $\delta_j = C_\delta \sqrt{\log(d)/n}\, \mathbb{I}(1 \leqslant j \leqslant 3), j = 1, \ldots, d$, with $C_\delta > 0$ controlling the magnitude of the structural change. For our proposed methods, we embedded the individual and tail-adaptive procedures into the seeded binary segmentation (SBS) framework [23]. Algorithm A.1 was implemented with parameters $\zeta = 1/\sqrt{2}$, $q_0 = 0.01$, $s_0 = 1$, $\tau = 0.5$, $K = 2$, $\alpha = 0.05$, $B = 100$, $m = \lfloor 0.2\, n \rfloor$, and $\lambda \in \{0, 0.1, 0.5, 0.9, 1\}$. All results are based on 100 replications.

To evaluate the performance in identifying change-points, we adopted the scaled Hausdorff distance to quantify the accuracy of change-point estimation. Specifically, let $\mathcal{T}_0 = \{l_1, l_2\}$ denote the set of true change-point locations, and let $\widehat{\mathcal{T}} = \{\widehat{l}_1, \ldots, \widehat{l}_{\widehat{K}}\}$ denote the set of estimated change-points obtained from the SBS procedure. The scaled Hausdorff distance between $\widehat{\mathcal{T}}$ and $\mathcal{T}_0$ is defined as

$$\mathrm{HD}(\widehat{\mathcal{T}}, \mathcal{T}_0) = \frac{1}{n} \max \left\{ \max_{\widehat{l} \in \widehat{\mathcal{T}}} \min_{l \in \mathcal{T}_0} |\widehat{l} - l|,\ \max_{l \in \mathcal{T}_0} \min_{\widehat{l} \in \widehat{\mathcal{T}}} |l - \widehat{l}| \right\}.$$

Smaller values of HD indicate more accurate recovery of both the number and locations of change-points, with HD = 0 corresponding to exact detection.

Table 2 Performance of multiple change-point estimation under the Toeplitz covariance structure with $(n, d) = (200, 10)$. The mean scaled Hausdorff distance is reported with standard deviations in parentheses, based on 100 replications with $B = 100$ bootstrap for each replication.

	$N(0,1)$ $(C_\delta = 5\sqrt{2})$		$N(0,1)$ $(C_\delta = 6\sqrt{2})$		t_3 $(C_\delta = 5\sqrt{2})$		t_3 $(C_\delta = 6\sqrt{2})$	
Methods	HD	(Sd)	HD	(Sd)	HD	(Sd)	HD	(Sd)
$\lambda = 0$	0.080	(0.039)	0.132	(0.028)	0.130	(0.055)	0.127	(0.040)
$\lambda = 0.1$	0.118	(0.020)	0.107	(0.048)	0.110	(0.058)	0.118	(0.033)
$\lambda = 0.5$	0.143	(0.024)	0.146	(0.008)	0.078	(0.028)	0.133	(0.014)
$\lambda = 0.9$	0.152	(0.039)	0.140	(0.012)	0.193	(0.075)	0.119	(0.051)
$\lambda = 1$	0.135	(0.015)	0.124	(0.026)	0.158	(0.072)	0.116	(0.011)
Adaptive	0.137	(0.023)	0.143	(0.023)	0.113	(0.031)	0.090	(0.042)

Tables 2 and 3 summarize the performance of multiple change-point estimation under the Toeplitz and compound symmetric covariance structures, respectively, using the scaled Hausdorff distance as the evaluation metric.

Under the Toeplitz covariance structure (Table 2), the proposed procedure achieves relatively small Hausdorff distances across all settings, indicating accurate detection and localization of multiple change-points. Overall, the estimation accuracy improves as the signal strength increases from $C_\delta = 5\sqrt{2}$ to $6\sqrt{2}$, particularly under the Gaussian error distribution. Among the fixed choices of λ, moderate values tend to yield more stable performance, while the adaptive procedure performs competitively and often attains smaller Hausdorff distances, especially under heavy-tailed t_3 errors.

Table 3 Performance of multiple change-point estimation under the Compound Symmetric covariance structure with $(n, d) = (200, 10)$. The mean scaled Hausdorff distance is reported with standard deviations in parentheses, based on 100 replications with $B = 100$ bootstrap for each replication.

	$N(0,1)$ $(C_\delta = 5\sqrt{2})$		$N(0,1)$ $(C_\delta = 6\sqrt{2})$		t_3 $(C_\delta = 5\sqrt{2})$		t_3 $(C_\delta = 6\sqrt{2})$	
Methods	HD	(Sd)	HD	(Sd)	HD	(Sd)	HD	(Sd)
$\lambda = 0$	0.178	(0.070)	0.061	(0.047)	0.089	(0.033)	0.095	(0.040)
$\lambda = 0.1$	0.080	(0.026)	0.097	(0.068)	0.121	(0.092)	0.091	(0.032)
$\lambda = 0.5$	0.119	(0.013)	0.107	(0.047)	0.137	(0.031)	0.117	(0.043)
$\lambda = 0.9$	0.120	(0.048)	0.119	(0.025)	0.106	(0.052)	0.116	(0.026)
$\lambda = 1$	0.109	(0.046)	0.112	(0.037)	0.109	(0.071)	0.129	(0.024)
Adaptive	0.070	(0.058)	0.123	(0.025)	0.058	(0.019)	0.104	(0.045)

Similar patterns are observed under the compound symmetric covariance structure (Table 3). Despite the stronger cross-correlation among covariates, the proposed method continues to produce reasonably small Hausdorff distances, demonstrating robustness to different dependence structures. In particular, the adaptive procedure consistently achieves favorable performance across both Gaussian and heavy-tailed error distributions, suggesting its effectiveness in balancing sensitivity and localization accuracy without requiring manual tuning of λ.

Overall, these results indicate that embedding the proposed change-point test within the SBS framework enables reliable detection and accurate localization of multiple change-points across a wide range of covariance structures and error distributions.

Finally, we assess the computational complexity of the proposed multi–change-point detection procedure for partially linear models. The algorithm consists of two main components. The first is the cross-fitted semiparametric composite quantile regression (CS–CQR) for estimating the nonparametric functions $\{\boldsymbol{\alpha}(\cdot), \gamma(\cdot)\}$, which involves local linear smoothing and evaluation of the composite check loss at L quantile levels for each observation, leading to a computational complexity of order $O(nqL)$, where n is the sample size and q is the dimension of the varying-coefficient regressors. The second component concerns the construction of the individual and tail-adaptive testing statistics under the Seeded Binary Segmentation (SBS) framework. Based on the pseudo-responses obtained from the first component, the individual test requires solving one Lasso problem for each seed interval, resulting in a complexity of $O(S \cdot \text{Lasso}(n, d))$, while the tail-adaptive test further evaluates the statistic over a collection of candidate adaptive weights, yielding a total cost of $O(S \cdot |\mathcal{A}| \cdot \text{Lasso}(n, d))$, where S denotes the number of seed intervals, d is the dimension of the parametric regressors, and $|\mathcal{A}|$ is the number of candidate tail-adaptive weights. Consequently, the overall computational cost is dominated by the repeated solution of Lasso problems in the SBS step when S or $|\mathcal{A}|$ is large, whereas the CS–CQR step scales linearly with the sample size and the number of quantile levels and is comparatively less computationally demanding.

Figure 3 displays the average running time of the proposed method under different values of λ with sample sizes $n \in \{100, 200, 300\}$, where the dimension is fixed at $d = 10$, the covariance structure corresponds to Scenario 1, the errors are normally distributed, and $\delta_j = 0.5\,\mathbb{I}(1 \leqslant j \leqslant 3), j = 1, \ldots, 10$. We observe that the computational cost increases with the sample size for all choices of λ. In particular, the tail-adaptive testing method incurs the highest computational cost. This is expected, since the method is designed to adapt to the tail structure of the error distribution. To achieve such adaptivity, Lasso estimators are computed repeatedly with different weighting parameters $\lambda \in \mathcal{A} = \{0, 0.1, 0.5, 0.9, 1\}$ for each seed interval, thereby increasing the computational burden. Nevertheless, the overall running time remains moderate for the considered sample sizes, indicating that the proposed method is computationally feasible in practice.

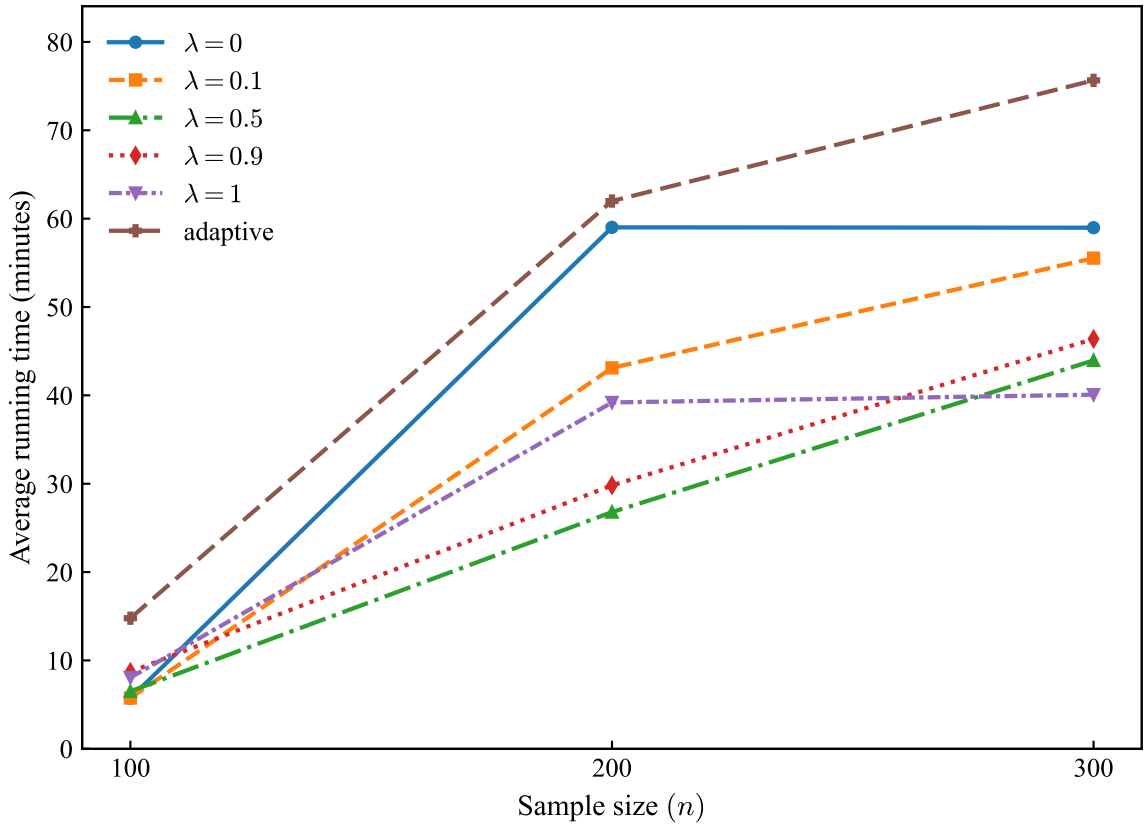

Figure 3 Average running time (in minutes) versus the sample size $n \in \{100, 200, 300\}$ for different values of λ. (Color online)

5 Application to energy data

We apply the proposed method to a publicly available monthly dataset for China obtained from Kaggle (see https://www.kaggle.com). The sample spans January 2015 to April 2025 and retains months with complete observations on five series: CO_2 intensity (g CO_2/kWh; hereafter CDI), electricity demand (TWh; ED), clean-energy generation (TWh; CG), fossil-fuel generation (TWh; FG), and fossil-fuel emissions (Mt CO_2; FE). The final sample size is $n = 124$.

In energy and environmental policy evaluations, it is essential to assess whether the relationship between CO_2 intensity and production-side variables remains stable over time [39]. Structural changes in this relationship are of particular interest to supervisory authorities, policymakers, energy regulators, corporate decision-makers, and researchers, as they not only help assess the effectiveness of implemented measures but also serve as indirect evidence regarding the timing and enforcement of specific policies.

We explore the relationships between CO_2 intensity (CDI) and the remaining time series (see Figure 4). The scatterplots reveal several features. To begin with, CDI is negatively related to electricity demand (ED) and clean-energy generation (CG), with the decline flattening at higher levels—indicative of a stable concave nonlinearity. Additionally, CDI is positively related to fossil-fuel generation (FG) and fossil-fuel emissions (FE) over most of the support; however, at comparable FG/FE levels, later observations lie systematically lower, implying that the relationship is not time-invariant. Notably, FG and FE are mechanically collinear. Collectively, these features motivate modeling ED and CG with smooth nonlinear terms, while treating production-side relations as near-linear but temporally heterogeneous.

Accordingly, we adopt a partially linear varying-coefficient framework. Let Y_t denote CO_2 intensity (CDI) and set $U_t = t/n$ be the normalized monthly index for $t = 1, \ldots, n$. To mitigate the collinearity between FG and FE, we introduce the emission factor $\mathrm{EF}_t = \mathrm{FE}_t/\mathrm{FG}_t$ (defined as fossil-fuel emissions

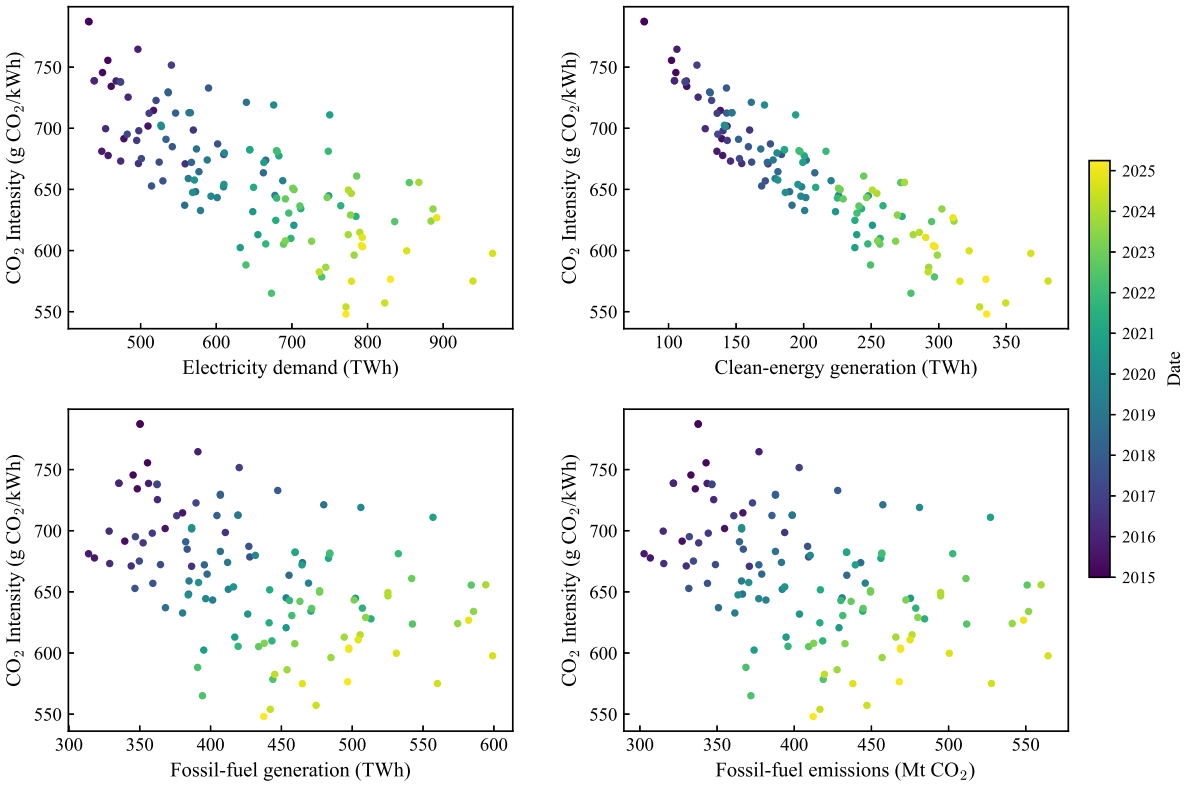

Figure 4 Scatter plots of the CO_2 intensity ($g\,CO_2$/kWh) against four covariates: electricity demand (TWh), clean-energy generation (TWh), fossil-fuel generation (TWh), and fossil-fuel emissions (Mt CO_2). (Color online)

per unit of fossil-fuel generation). Specifically, we place $\mathbf{X}_t = (\mathrm{FG}_t, \mathrm{EF}_t)^\top$ in the linear component (with coefficients allowed to vary over t; FE is not entered separately), and $\mathbf{Z}_t = (\mathrm{ED}_t, \mathrm{CG}_t)^\top$ in the nonparametric component (smooth but time-invariant). Our empirical question is whether the coefficients on $\mathbf{X}_t$ are stable over time or display structural breaks, while $\mathbf{Z}_t$ captures the time–invariant nonlinear effects on Y_t. To remove the nonparametric effects, we first estimate them using Algorithm 1 and construct pseudo-responses by residualizing the fitted components. Prior to change-point detection, each series is standardized so that its mean is zero and its variance is one. We then detect multiple change-points by applying our tail-adaptive test within the seeded binary segmentation framework [23], using the seeding parameter $\zeta = 1/\sqrt{2}$. For implementation, we set $q_0 = 0.01$, $s_0 = 1$, $\tau = 0.5$, $K = 2$, $\alpha = 0.05$, $B = 100$, and $\lambda \in \{0, 0.1, \ldots, 1\}$. There are four change-points detected: June 2017, November 2020, June 2022, and May 2023 (see Figure 5).

To further interpret the detected breaks, we relate them to contemporaneous developments in China's electricity system. Figure 5 displays the CO_2 intensity series, with red dashed lines indicating the estimated change-points. The first break, in June 2017, aligns with the launch of spot market pilots that shifted dispatch and pricing toward market signals, thereby altering the merit order, the generation mix, and per-unit emissions of incremental generation. The second break, in November 2020, follows the national carbon-neutrality pledge and coincides with flexibility-oriented reforms—wider time-of-use pricing, demand-response programs, storage rules, and flexibility retrofits of thermal units—which plausibly affected both the scale of fossil-fuel generation and the emission factor. In June 2022, the third break overlaps with an extreme heatwave and drought that curtailed hydropower, inducing substitution toward fossil-fuel generation and temporarily increasing per-unit emissions at the margin. The fourth break, in May 2023, coincides with rapid additions of wind and solar capacity alongside expanded green-power and spot trading and growing storage, reducing curtailment and shifting incremental supply toward lower-emission sources. Taken together, these episodes are consistent with discrete changes in the emission factor and the scale of fossil-fuel generation—conditional on electricity demand and clean-energy

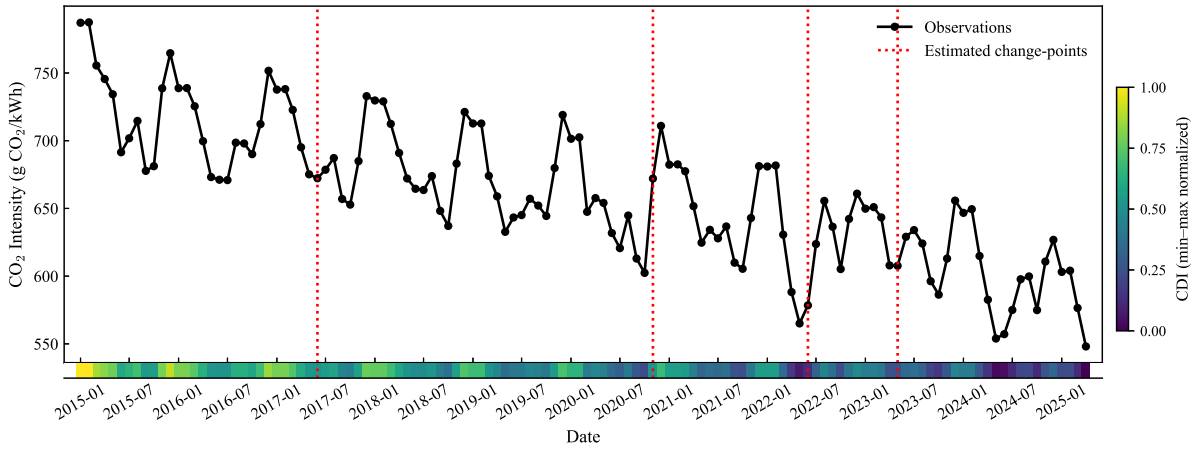

Figure 5 Plot of the CO_2 intensity with the estimated change-points (red dashed lines). (Color online)

generation—and hence with time variation in the relationship between CO_2 intensity and production-side quantities.

Overall, the previously detected change-points coincide with documented policy shifts or external shocks, indicating structural transitions in how CO_2 intensity responds to generation-side variables. While these alignments do not imply causality, they offer insight into the timing and persistence of regime changes. Such break patterns may serve as a valuable reference for future policy design, helping to anticipate the system's behavior under similar interventions or shocks and to adapt strategies accordingly.

6 Conclusion

This paper establishes a unified tail-adaptive CUSUM framework for structural break testing and estimation in partially linear models. The procedure integrates cross-fitted semiparametric nuisance estimation with high-dimensional change-point inference, enabling robust detection of structural changes in the parametric component under sparsity and nonlinear effects. The central theoretical contribution is the establishment of a uniform oracle equivalence result. We show that the feasible plug-in CUSUM process and its associated statistics converge uniformly, at rate $o_p(1)$, to their oracle linear-model counterparts over $t \in [q_0, 1-q_0]$. This provides a rigorous bridge between semiparametric nuisance estimation and high-dimensional change-point analysis, and forms the basis for the asymptotic validity of both the individual and tail-adaptive tests. To address the intractable null distribution of the aggregated high-dimensional CUSUM statistic, a tuning-free multiplier bootstrap procedure is developed and shown to be consistent under mild regularity conditions. In addition, we obtain non-asymptotic guarantees for change-point localization, showing that the proposed estimator achieves the optimal rate $\log(dn)/n$ up to logarithmic factors. The procedure is further implemented within a seeded binary segmentation framework to accommodate multiple change-points. Simulation studies demonstrate reliable detection and accurate localization performance. An empirical application to energy policy data illustrates the practical relevance of the proposed methodology. Future research may explore more scalable nuisance estimation strategies and extensions to dependent data structures, which would further broaden the applicability of the framework.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 12161016 and 11661018), the Guizhou Provincial Basic Research Program (Natural Science) (No. QianKeHeJiChu-ZK[2024] YiBan082), and the Guizhou Key Laboratory of Artificial Intelligence and Brain-inspired Computing. Jing's research was partially supported by National Natural Science Foundation of China (Grant No. 12371290) and the SUSTech-NUS Joint Research Program. Wang Zhou was partially supported by the SUSTech-NUS Joint Research Program. We extend our sincere gratitude to the Gui'an Supercomputing

Center, managed by Gui'an New Area Science and Technology Innovation Industries Development Co., Ltd., for the invaluable provision of computational resources and technical support, which were instrumental in carrying out the experiments for this study. The authors would also like to thank the editor and the anonymous reviewers for their constructive comments and valuable suggestions, which have significantly improved the quality and clarity of this manuscript.

References

1 Aue A, Horváth L, Hušková M, et al. Change-point monitoring in linear models. Econom J, 2006, 9: 373–403
2 Aue A, Horváth L. Structural breaks in time series. J Time Ser Anal, 2013, 34: 1–16
3 Aue A, Kirch C. The state of cumulative sum sequential changepoint testing 70 years after Page. Biometrika, 2024, 111: 367–391
4 Bai J, Perron P. Estimating and testing linear models with multiple structural changes. Econometrica, 1998, 66: 47–78
5 Baranowski R, Chen Y, Fryzlewicz P. Narrowest-over-threshold detection of multiple change-points and change-point-like features. J R Stat Soc Ser B Stat Methodol, 2019, 81: 649–672
6 Brown R L, Durbin J, Evans J M. Techniques for testing the constancy of regression relationships over time. J R Stat Soc Ser B Stat Methodol, 1975, 37: 149–163
7 Byrd R H, Lu P, Nocedal J, et al. A limited memory algorithm for bound constrained optimization. SIAM J Sci Comput, 1995, 16: 1190–1208
8 Chernozhukov V, Chetverikov D, Demirer M, et al. Double/debiased machine learning for treatment and structural parameters. Econom J, 2018, 21: C1–C68
9 Chu C S J, Stinchcombe M, White H. Monitoring structural change. Econometrica, 1996, 64: 1045–1065
10 Deng A, Perron P. A non-local perspective on the power properties of the CUSUM and CUSUM of squares tests for structural change. J Econom, 2008, 142: 212–240
11 Dette H, Gösmann J. Relevant change-points in high dimensional time series. Electron J Stat, 2018, 12: 2578–2636
12 Dette H, Gösmann J. A likelihood ratio approach to sequential change-point detection for a general class of parameters. J Am Stat Assoc, 2020, 115: 1361–1377
13 Fan J, Gijbels I. Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66 (1st ed.). New York: Routledge, 1996
14 Fan J, Huang T. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli, 2005, 11: 1031–1057
15 Ferreira C S, Zeller C B, Mimura A M, et al. Partially linear models and their applications to change-point detection of chemical process data. J Appl Stat, 2017, 44: 2125–2141
16 Fryzlewicz P. Wild binary segmentation for multiple change-point detection. Ann Stat, 2014, 42: 2243–2281
17 Hocking T D, Rigaill G, Fearnhead P, et al. Generalized functional pruning optimal partitioning (gfpop) for constrained changepoint detection in genomic data. J Stat Softw, 2022, 101: 1–31
18 Huang X, Chen L, Chen W, et al. Change-point detection in temporal networks based on graph snapshot similarity measures. Appl Math Comput, 2025, 489: 129165
19 Jia S, Shi L. Efficient change-points detection for genomic sequences via cumulative segmented regression. Bioinformatics, 2022, 38: 311–317
20 Jiang P, Kurozumi E. Power properties of the modified CUSUM tests. Commun Stat Theory Methods, 2019, 48: 2962–2981
21 Jirak J M. Uniform change-point tests in high dimension. Ann Stat, 2015, 43: 2451–2483
22 Kai B, Li R, Zou H. New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. Ann Stat, 2011, 39: 305–332
23 Kovács S, Bühlmann P, Li H, et al. Seeded binary segmentation: a general methodology for fast and optimal changepoint detection. Biometrika, 2023, 110: 249–256
24 Li N, Fei Y, Zhang X. Partial linear model averaging prediction for longitudinal data. J Syst Sci Complex, 2024, 37: 863–885
25 Liu D, Li R, Wang Z. Testing for structural breaks in panel varying coefficient models: with an application to OECD health expenditure. Empir Econ, 2011, 40: 95–118
26 Liu B, Qi Z, Zhang X, et al. Change-point detection for high-dimensional linear models: A general tail-adaptive approach. arXiv:2207.11532, 2022
27 Liu B, Zhang X, Liu Y. High dimensional change-point inference: recent developments and extensions. J Multivar Anal, 2022, 188: 104833
28 Liu B, Zhou C, Zhang X, et al. A unified data-adaptive framework for high dimensional change point detection. J R Stat Soc Ser B Stat Methodol, 2020, 82: 933–963

29 Lund R B, Beaulieu C, Killick R, et al. Good practices and common pitfalls in climate time series changepoint techniques: a review. J Clim, 2023, 36: 8041–8057
30 Moustakides G V. Optimal stopping times for detecting changes in distributions. Ann Stat, 1986, 14: 1379–1387
31 Otto S, Breitung J. Backward CUSUM for testing and monitoring structural change with an application to COVID-19 pandemic data. Econom Theory, 2023, 39: 659–692
32 Page E S. Continuous inspection schemes. Biometrika, 1954, 41: 100–115
33 Ploberger W, Krämer W. The CUSUM test with OLS residuals. Econometrica, 1992, 60: 271–285
34 Qu Z, Yoon J, Perron P. Inference on conditional quantile processes in partially linear models with applications to the impact of unemployment benefits. Rev Econ Stat, 2024, 106: 521–541
35 Stone C J. An asymptotically optimal window selection rule for kernel density estimates. Ann Stat, 1984, 12: 1285–1297
36 Su L, White H. Testing structural change in partially linear models. Econom Theory, 2010, 26: 1761–1806
37 Su L, Xiao Z. Testing structural change in time-series nonparametric regression models. Stat Its Interface, 2008, 1: 347–366
38 Vogt M. Testing for structural change in time-varying nonparametric regression models. Econom Theory, 2015, 31: 811–859
39 Wang Y, Shi R, Zhang C, et al. Structural changes and trends in China's renewable electricity production in the policy evolution process. Renew. Energy, 2022, 182: 879–886
40 Yang Q, Li Y N, Zhang Y. Change-point detection for nonparametric regression under strongly mixing process. Stat Pap, 2020, 61: 1465–1506
41 Zhang Y, Zhu C, Shao X. Change-point detection for object-valued time series. J Bus Econ Stat, 2026: 255–269
42 Zhou C, Zhang X, Zhou W, et al. A unified framework for testing high dimensional parameters: a data-adaptive approach. arXiv:1808.02648, 2018

Appendix A Supplementary numerical studies under a simplified model

In this appendix, we report supplementary simulation results for a simplified partially linear model with a single change-point and $\gamma(U) = 0$, which serves as a special case of the general setting considered in Section 4. Unless otherwise specified, the simulation design, tuning procedures, and implementation details follow exactly those used in the main text. The primary difference lies in the model specification and the number of replications. We consider the same partially linear model as in Section 4, except that the nonparametric component $\gamma(U)$ is set to zero. Specifically, the data are generated according to

$$Y_i = \mathbf{X}_i^\top \boldsymbol{\beta}^{(1)} \mathbb{I}\{i < l_0\} + \mathbf{X}_i^\top \boldsymbol{\beta}^{(2)} \mathbb{I}\{i \geqslant l_0\} + \alpha(U_i) Z_i + \varepsilon_i, \quad i = 1, \ldots, n. \tag{A.1}$$

The distributions of $(\mathbf{X}_i, Z_i, U_i)$, the error distributions, the regression coefficients, the covariance structures, and the tuning parameter selection are identical to those in Section 4. In this appendix, the number of replications is set to 1000, while the number of bootstrap resamples remains $B = 100$.

Appendix A.1 Size performance

This subsection reports the empirical size of the proposed tests under the simplified setting $\gamma(U) = 0$.

Table 4 reports the empirical sizes of the proposed tests under the simplified model with $\gamma(U) = 0$. The sizes are well controlled around the nominal level across different error distributions and covariance structures, which is consistent with the findings in Section 4.1.

Appendix A.2 Power performance

We next examine the finite-sample power of the proposed tests under the simplified model with $\gamma(U) = 0$.

Figure 6 and 7 summarize the empirical power results under the simplified setting $\gamma(U) = 0$. Overall, the power increases with the signal strength, and the tail-adaptive test maintains stable performance across different error distributions. The qualitative patterns are similar to those observed in Section 4.2.

Table 4 Empirical sizes of the individual test $\Theta_{\lambda,\alpha}$ with $\lambda \in \{0, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ and the tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$ under Toeplitz and Compound Symmetric covariance structures for $s_0 \in \{1, 2, 3, 4, 5\}$, at the significance level $\alpha = 0.05$. Results are based on 1000 replications with $B = 100$ bootstrap draws per replication.

Empirical sizes for Scenario 1									
Dist.	s_0	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 0.3$	$\lambda = 0.5$	$\lambda = 0.7$	$\lambda = 0.9$	$\lambda = 1$	Adaptive
t_3	1	0.048	0.058	0.052	0.050	0.044	0.047	0.040	0.040
	2	0.046	0.046	0.042	0.044	0.058	0.048	0.043	0.055
	3	0.046	0.047	0.070	0.054	0.050	0.044	0.054	0.046
	4	0.041	0.045	0.060	0.030	0.040	0.040	0.049	0.049
	5	0.047	0.048	0.050	0.049	0.041	0.052	0.048	0.053
t_5	1	0.052	0.051	0.053	0.049	0.044	0.045	0.047	0.053
	2	0.048	0.048	0.052	0.060	0.052	0.060	0.051	0.038
	3	0.069	0.049	0.054	0.051	0.048	0.049	0.057	0.050
	4	0.051	0.038	0.057	0.042	0.049	0.048	0.047	0.039
	5	0.042	0.058	0.051	0.054	0.050	0.063	0.038	0.039
$N(0,1)$	1	0.047	0.055	0.040	0.043	0.052	0.047	0.049	0.045
	2	0.055	0.063	0.051	0.054	0.055	0.039	0.043	0.049
	3	0.050	0.037	0.046	0.046	0.048	0.050	0.039	0.035
	4	0.053	0.045	0.060	0.046	0.047	0.050	0.049	0.050
	5	0.052	0.052	0.052	0.051	0.052	0.048	0.043	0.049
Empirical sizes for Scenario 2									
Dist.	s_0	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 0.3$	$\lambda = 0.5$	$\lambda = 0.7$	$\lambda = 0.9$	$\lambda = 1$	Adaptive
t_3	1	0.041	0.051	0.049	0.059	0.048	0.049	0.051	0.054
	2	0.037	0.041	0.049	0.051	0.053	0.051	0.055	0.046
	3	0.054	0.052	0.046	0.033	0.048	0.050	0.050	0.038
	4	0.051	0.051	0.043	0.058	0.047	0.044	0.047	0.051
	5	0.045	0.051	0.052	0.057	0.047	0.049	0.068	0.043
t_5	1	0.059	0.052	0.043	0.048	0.053	0.051	0.052	0.049
	2	0.053	0.054	0.045	0.046	0.051	0.052	0.051	0.050
	3	0.045	0.062	0.051	0.060	0.044	0.041	0.046	0.052
	4	0.062	0.054	0.043	0.043	0.061	0.045	0.044	0.054
	5	0.050	0.046	0.057	0.064	0.054	0.055	0.041	0.042
$N(0,1)$	1	0.064	0.033	0.046	0.050	0.049	0.048	0.047	0.045
	2	0.053	0.058	0.059	0.062	0.065	0.042	0.050	0.050
	3	0.051	0.041	0.049	0.046	0.050	0.042	0.040	0.042
	4	0.042	0.052	0.060	0.040	0.060	0.034	0.049	0.043
	5	0.051	0.040	0.051	0.043	0.055	0.043	0.048	0.045

Appendix B Extensions to a multiple change-points detection procedure

In practical applications, when the null hypothesis H_0 is rejected by our powerful tail-adaptive test for potential data structures, multiple change-points may exist in the data structure. Consequently, addressing the rapid estimation or detection of multiple change-points is crucial. In this section, we extend our single change-point detection method to the case of multiple change-points based on the principles of the seeded binary segmentation (SBS) method proposed by [23], aiming to effectively estimate the locations of all possible multiple change-points in the data structure.

Definition B.1 (Seeded intervals [23]). Let $\zeta \in [1/2, 1)$ denote a given decay parameter. Let $I_1 = (0, n]$. For $k = 2, \ldots, \left\lceil \log_{1/\zeta}(n) \right\rceil$ (i.e. logarithm with base $1/\zeta$), define the k-th layer $\mathcal{I}_k$ as the collection of n_k intervals of initial length l_k that are evenly shifted by the deterministic shift s_k as

$$\mathcal{I}_k := \bigcup_{i=1}^{n_k} \left\{ \left(\lfloor (i-1)s_k \rfloor, \lceil (i-1)s_k + l_k \rceil \right] \right\}, \tag{B.1}$$

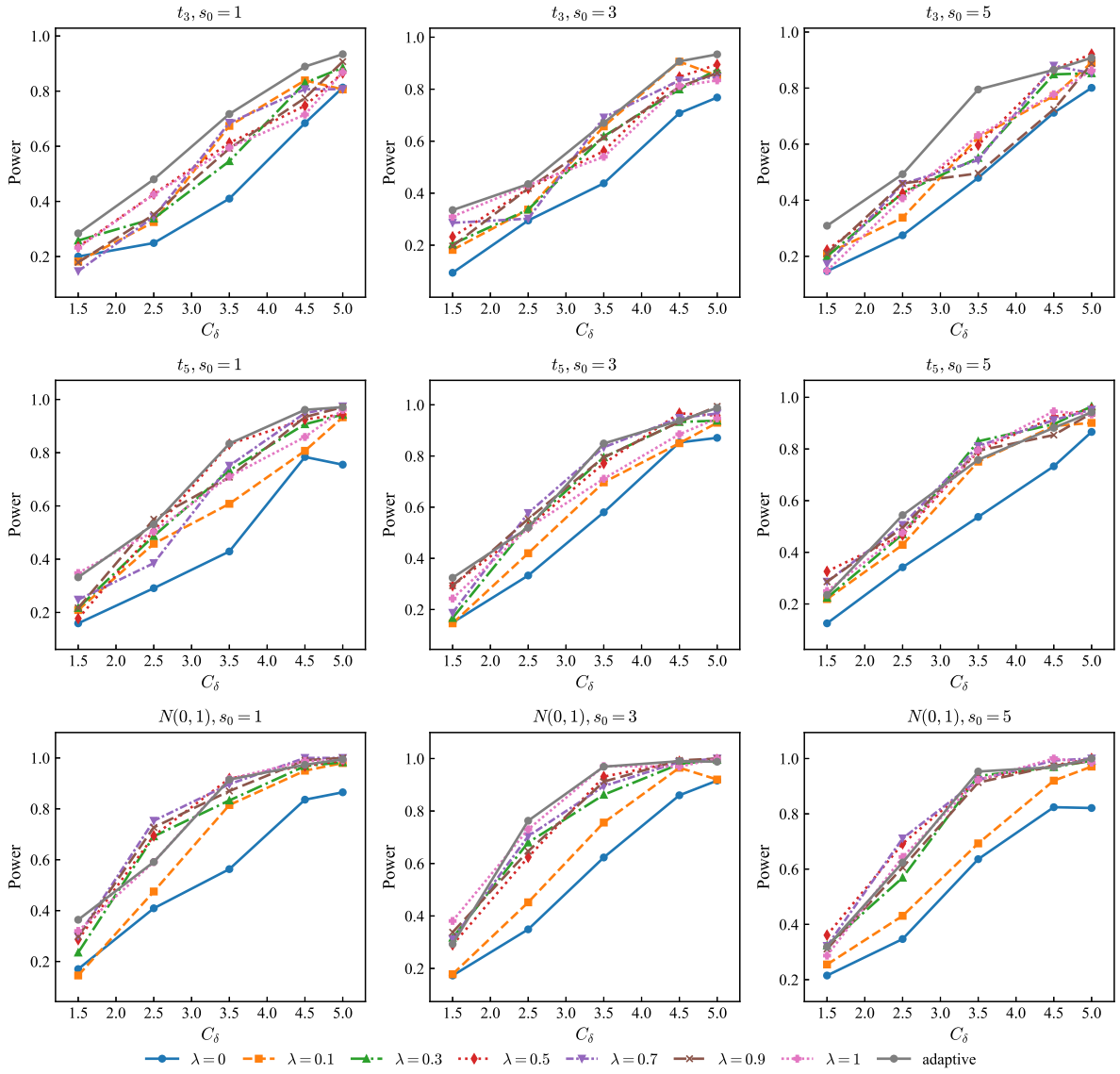

Figure 6 Empirical powers of the individual tests $\Theta_{\lambda,\alpha}$ with $\lambda \in \{0, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ and the tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$ under Toeplitz covariance structures for $s_0 \in \{1, 3, 5\}$ and $t_0 = 0.5$, at nominal level $\alpha = 0.05$. Results are based on 1000 replications with $B = 100$ bootstrap draws per replication. (Color online)

where $n_k := 2\left\lceil (1/\zeta)^{k-1} \right\rceil - 1, l_k := n\zeta^{k-1}$ and $s_k := (n - l_k) / (n_k - 1)$. Then, define the overall collection of seeded intervals as

$$\mathcal{I} := \bigcup_{k=1}^{\lceil \log_{1/\zeta}(n) \rceil} \mathcal{I}_k. \tag{B.2}$$

We propose an SBS-type tail-adaptive procedure for multiple change-point detection, hereafter referred to as multi-CPD, which is summarized in Algorithm A.1.

The procedure operates on a collection of N seeded intervals $\{(l_i, r_i]\}_{i=1}^N$, constructed following the seeded binary segmentation (SBS) principle, where the interval endpoints satisfy $\lfloor nq_0 \rfloor \leqslant l_i < r_i \leqslant \lceil n(1 - q_0) \rceil$. Only intervals with length at least m are retained to ensure reliable inference.

The algorithm proceeds as follows. First, the nonparametric components $\boldsymbol{\alpha}(\cdot)$ and $\gamma(\cdot)$ are estimated using the cross-fitted semiparametric CQR estimator in Algorithm 1. Based on these estimates, we construct the pseudo-responses and compute, for each seeded interval $(l_j, r_j]$, the corresponding tail-adaptive p-value $\widehat{P}_{\mathrm{ad}}(l_j, r_j]$ using the bootstrap procedure described in Section 2.3. Starting from the

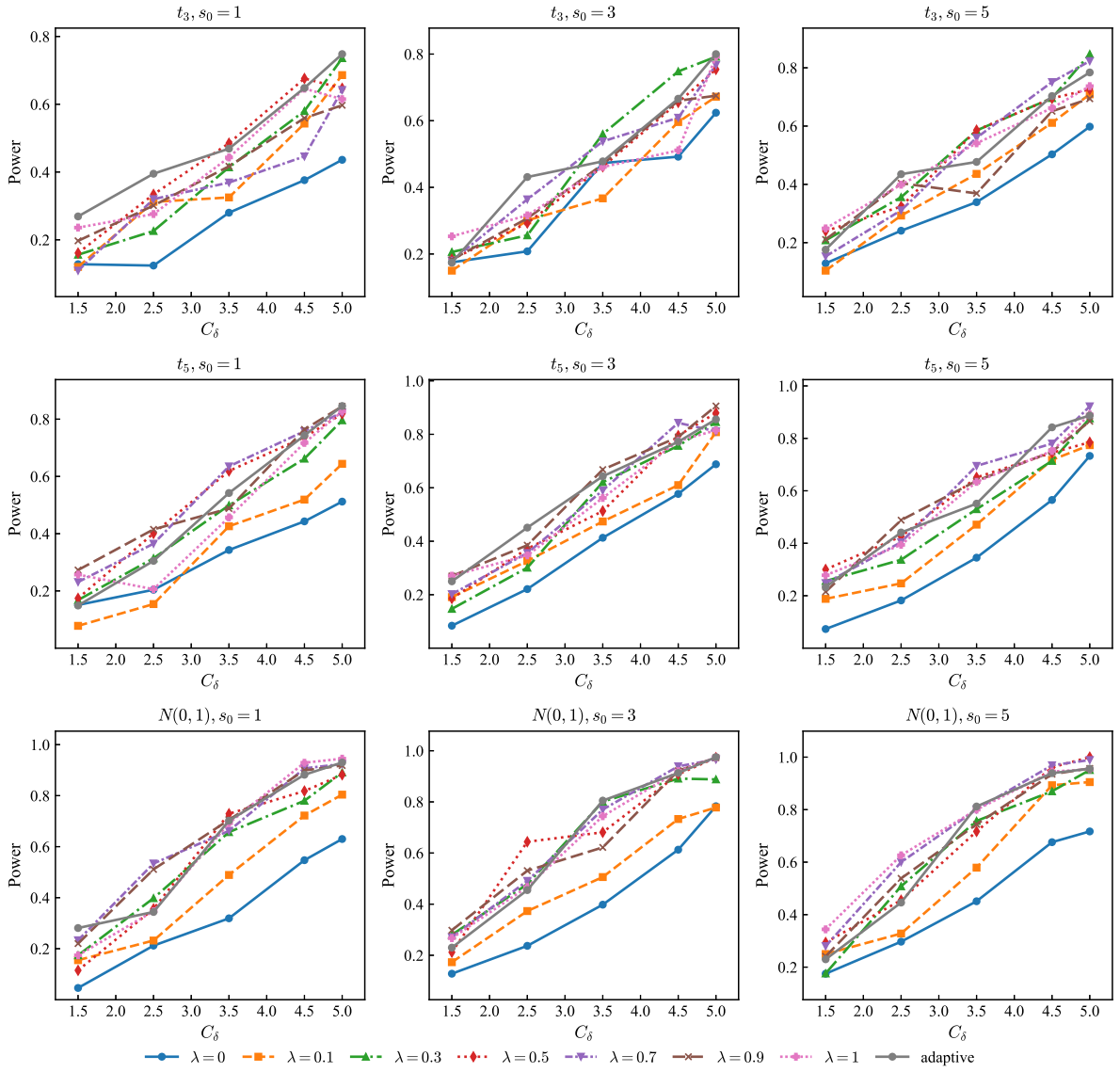

Figure 7 Empirical powers of the individual tests $\Theta_{\lambda,\alpha}$ with $\lambda \in \{0, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ and the tail-adaptive test $\Theta_{\mathrm{ad},\alpha}$ under Compound Symmetric covariance structures for $s_0 \in \{1, 3, 5\}$ and $t_0 = 0.5$, at nominal level $\alpha = 0.05$. Results are based on 1000 replications with $B = 100$ bootstrap draws per replication. (Color online)

initial search range $[L, R] = [q_0, 1 - q_0]$, the algorithm applies a recursive segmentation scheme. For a given interval $[L, R]$, let $\mathcal{I} = \{1 \leqslant j \leqslant N : [l_j, r_j] \subset [L, R]\}$ denote the index set of seeded intervals fully contained in $[L, R]$. Among all intervals in $\mathcal{I}$ with length at least m, we compute the tail-adaptive test statistic

$$\overline{P}_{\mathrm{ad}} = \min_{j \in \mathcal{I},\, m \leqslant r_j - l_j} \widehat{P}_{\mathrm{ad}}(l_j, r_j],$$

and denote by j^* the index that attains the minimum. If $\overline{P}_{\mathrm{ad}} > \alpha$, the algorithm terminates on the current interval $[L, R]$, indicating that no change-point is detected within this segment. Otherwise, a change-point $\widehat{t}_{j^*}$ associated with the interval $(l_{j^*}, r_{j^*}]$ is added to the estimated change-point set $\mathcal{C}$. The procedure is then recursively applied to the subintervals $[L, j^*]$ and $[j^*, R]$. The recursion continues until all remaining segments have length smaller than m or no further significant change-points are detected. The final output is the set $\mathcal{C}$ of estimated multiple change-points.

Algorithm A.1 An SBS-type tail-adaptive procedure for multiple change-point detection

Input: Given the data $(\mathcal{X}, \mathcal{Z}, \mathcal{U}, \mathcal{Y}) := \{(\boldsymbol{X}_i, \boldsymbol{Z}_i, U_i, Y_i)\}_{i=1}^n$, set the values for $\widetilde{\boldsymbol{\tau}}$, the significance level α, s_0, q_0, K, the bootstrap replication number B, the subset of candidate weights $\mathcal{D} \subset [0,1]$, and a collection $\{(l_i, r_i]\}_{i=1}^N$ of seeded intervals (see Definition B.1) with decay $\zeta \in [1/2, 1)$ and minimal segment length m. Initialize the set of change-points $\mathcal{C} = \emptyset$.

Step 1: Compute the estimators of the nonparametric components $\boldsymbol{\alpha}(\cdot)$ and $\gamma(\cdot)$ using Algorithm 1.

Step 2: For each $j = 1, \cdots, N$, compute $\widehat{P}_{\mathrm{ad}}(l_j, r_j]$ following Section 2.3.

Step 3: Perform the following function with $L = q_0$ and $R = 1 - q_0$.

(a) If $R - L < m$, **return**.

(b) Define $\mathcal{I} = \{1 \leqslant j \leqslant N \mid [l_j, r_j] \subset [L, R]\}$.

(c) Compute the tail-adaptive test statistics as

$$\overline{P}_{\mathrm{ad}} = \min_{j \in \mathcal{I},\, m \leqslant r_j - l_j} \widehat{P}_{\mathrm{ad}}(l_j, r_j),$$

and the corresponding optimal solution j^*.

(d) If $\overline{P}_{\mathrm{ad}} > \alpha$, **return**. Otherwise, Update $\mathcal{C}$ by adding $\widehat{t}_{j^*}$ to $\mathcal{C}$, and perform Function(L, j^*) and Function(j^*, R).

Output: The set of estimated multiple change-points $\mathcal{C}$.

Appendix C Verification of Assumption C under local-linear smoothing

Lemma C.1. *Suppose that $\eta(\cdot)$ has two bounded and continuous derivatives on the compact support $\mathcal{U} \subset \mathbb{R}$ of U, and that the density f_U of U is bounded away from zero and infinity on $\mathcal{U}$. Let $\widetilde{\eta}$ be the out-of-fold cross-fitted estimator of $\eta(\cdot)$ constructed via one-dimensional local-linear smoothing with a bounded symmetric kernel $\mathcal{K}$ satisfying $\int \mathcal{K}(u)\,du = 1$ and $\int u^2 \mathcal{K}(u)\,du < \infty$. If the bandwidth satisfies $h \asymp n^{-1/5}$, then*

$$\mathbb{E} \int_{\mathcal{U}} \{\widetilde{\eta}(u) - \eta(u)\}^2 du = O\left(h^4 + \frac{1}{nh}\right).$$

Consequently,

$$\|\widetilde{\eta} - \eta\|_{L_2} = O_p(n^{-2/5}),$$

and therefore

$$\sqrt{n}\,\|\widetilde{\eta} - \eta\|_{L_2}^2 \to 0.$$

*In particular, Assumption **C** holds.*

Proof. For one-dimensional local-linear smoothing, it is well known that the integrated mean squared error (IMSE) satisfies

$$\mathbb{E} \int_{\mathcal{U}} \{\hat{\eta}(u) - \eta(u)\}^2 du = O\left(h^4 + \frac{1}{nh}\right),$$

under the stated smoothness and kernel conditions; see [13]. The term h^4 corresponds to the integrated squared bias, while $(nh)^{-1}$ arises from the integrated variance.

Choosing $h \asymp n^{-1/5}$ balances the bias and variance terms and yields

$$\mathbb{E} \int_{\mathcal{U}} \{\hat{\eta}(u) - \eta(u)\}^2 du = O(n^{-4/5}).$$

An application of Markov's inequality then implies

$$\|\hat{\eta} - \eta\|_{L_2} = O_p(n^{-2/5}).$$

For the cross-fitted estimator $\widetilde{\eta}$, each local-linear fit is computed on a subsample whose size is proportional to n. Therefore, the first-order bias and variance expansions remain unchanged, and the same IMSE rate applies; see [8] for general discussions on rate preservation under cross-fitting.

Combining the above results gives

$$\|\widetilde{\eta} - \eta\|_{L_2} = O_p(n^{-2/5}),$$

which implies

$$\sqrt{n}\,\|\widetilde{\eta} - \eta\|_{L_2}^2 = O_p(n^{-3/10}) \to 0.$$

This verifies Assumption **C**. □

Appendix D Orthogonal offset for the CS-CQR procedure

We first present the cross-fitted orthogonal offset estimator and its implementation details in Appendix D.1. Building on this construction and the standing assumptions from the main text, Appendix D.2 then establishes the estimator's large-sample properties and provides the corresponding proofs.

Appendix D.1 Cross–fitted orthogonal offset estimator

Algorithm A.2 computes a fixed, cross-fitted orthogonal offset $\boldsymbol{\beta}^\dagger$ for use within CS-CQR. The offset is computed once and held fixed throughout. The estimator is obtained through three out-of-fold steps:

(1) preliminarily estimate $g(U,\mathbf{Z}) = \mathbf{Z}^\top \boldsymbol{\alpha}(U) + \gamma(U)$ using local-linear CQR with zero offset;

(2) estimate $\mathbf{m}_0(U,\mathbf{Z}) = \mathbb{E}[\mathbf{X} \mid U,\mathbf{Z}]$ by local-linear least squares and form residual covariates $\check{\mathbf{X}} = \mathbf{X} - \widehat{\mathbf{m}}$;

(3) regress the residual response $\widetilde{Y}^{(0)} = Y - \widehat{g}^{(0)}(U,\mathbf{Z})$ on $\check{\mathbf{X}}$ via CQR to obtain $\boldsymbol{\beta}^\dagger$.

Algorithm A.2 Cross-fitted orthogonal offset estimator

Input: Data $\{(\mathbf{X}_i,\mathbf{Z}_i,U_i,Y_i)\}_{i=1}^n$; quantile levels $\widetilde{\boldsymbol{\tau}} = (\tau_1,\dots,\tau_L)^\top$, kernel $\mathcal{K}(\cdot)$ with $\mathcal{K}_s(u) = \mathcal{K}(u/s)/s$, bandwidth grids $\mathcal{H}$ (for $g(u,z) = \mathbf{z}^\top\boldsymbol{\alpha}(u) + \gamma(u)$) and $\mathcal{B}$ (for $\mathbf{m}(u,\mathbf{z}) = \mathbb{E}[\mathbf{X} \mid U = u, \mathbf{Z} = \mathbf{z}]$), and the number of folds $K \geqslant 2$.

Randomly partition $\{1,\dots,n\}$ into K folds $\mathcal{I}_1,\dots,\mathcal{I}_K$. For $k = 1,\dots,K$, set $\mathcal{I}_k^c = \bigcup_{j\neq k}\mathcal{I}_j$.

Cross-validated local-linear CQR for a preliminary $g^{(0)}$ (zero offset).

for $h \in \mathcal{H}$ **do**

  for $k = 1,\dots,K$ **do**

    for each $i \in \mathcal{I}_k$ (set $u_0 := U_i$) **do**

      Solve the local-linear composite quantile program

$$(\widehat{\boldsymbol{\alpha}}_0,\widehat{\boldsymbol{\alpha}}_1,\widehat{\gamma}_0,\widehat{\gamma}_1) \leftarrow \underset{\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_1,\gamma_0,\gamma_1}{\operatorname{argmin}} \sum_{j\in\mathcal{I}_k^c}\sum_{\ell=1}^{L} \rho_{\tau_\ell}\Big(Y_j - \mathbf{Z}_j^\top\{\boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1(U_j-u_0)\} - \{\gamma_0 + \gamma_1(U_j-u_0)\}\Big)\mathcal{K}_h(U_j-u_0).$$

      Set $\widehat{g}_h^{(-k)}(U_i,\mathbf{Z}_i) = \mathbf{Z}_i^\top\widehat{\boldsymbol{\alpha}}_0 + \widehat{\gamma}_0$.

    end for

  end for

  Compute the CV loss

$$\mathrm{CV}_g(h) = \sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}\sum_{\ell=1}^{L}\rho_{\tau_\ell}\big(Y_i - \widehat{g}_h^{(-k)}(U_i,\mathbf{Z}_i)\big).$$

end for

Choose $\widehat{h}_0 \leftarrow \arg\min_{h\in\mathcal{H}} \mathrm{CV}_g(h)$ and set $\widehat{g}^{(0)}(U_i,\mathbf{Z}_i) = \widehat{g}_{\widehat{h}_0}^{(-k)}(U_i,\mathbf{Z}_i)$ for all $i \in \mathcal{I}_k$, $k = 1,\dots,K$.

Cross-validated local-linear least squares for $\mathbf{m}(u,\mathbf{z})$.

for $b \in \mathcal{B}$ **do**

  for $k = 1,\dots,K$ **do**

    for each $i \in \mathcal{I}_k$ (set $u_0 := U_i$) **do**

      For each coordinate $j = 1,\dots,d$, solve

$$(\widehat{\boldsymbol{a}}_{0j},\widehat{\boldsymbol{a}}_{1j},\widehat{c}_{0j},\widehat{c}_{1j}) \leftarrow \underset{\boldsymbol{a}_{0j},\boldsymbol{a}_{1j},c_{0j},c_{1j}}{\operatorname{argmin}} \sum_{r\in\mathcal{I}_k^c}\Big(X_{rj} - \mathbf{Z}_r^\top\{\boldsymbol{a}_{0j} + \boldsymbol{a}_{1j}(U_r-u_0)\} - \{c_{0j} + c_{1j}(U_r-u_0)\}\Big)^2\mathcal{K}_b(U_r-u_0).$$

      Set $\widehat{m}_{j,b}^{(-k)}(U_i,\mathbf{Z}_i) = \mathbf{Z}_i^\top\widehat{\boldsymbol{a}}_{0j} + \widehat{c}_{0j}$.

    end for

    Set $\widehat{\mathbf{m}}_b^{(-k)}(U_i,\mathbf{Z}_i) = (\widehat{m}_{1,b}^{(-k)},\dots,\widehat{m}_{d,b}^{(-k)})^\top$ and $\check{\mathbf{X}}_{i,b} = \mathbf{X}_i - \widehat{\mathbf{m}}_b^{(-k)}(U_i,\mathbf{Z}_i)$ for all $i \in \mathcal{I}_k$.

  end for

  Compute $\mathrm{CV}_m(b) = \sum_{k=1}^K\sum_{i\in\mathcal{I}_k}\|\mathbf{X}_i - \widehat{\mathbf{m}}_b^{(-k)}(U_i,\mathbf{Z}_i)\|_2^2$.

end for

Choose $\widehat{b} \leftarrow \arg\min_{b\in\mathcal{B}}\mathrm{CV}_m(b)$ and set $\check{\mathbf{X}}_i = \check{\mathbf{X}}_{i,\widehat{b}}$ for all i.

Residualized CQR for the offset. Construct $\widetilde{Y}_i^{(0)} = Y_i - \widehat{g}^{(0)}(U_i,\mathbf{Z}_i)$ and compute

$$\boldsymbol{\beta}^\dagger \leftarrow \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{\ell=1}^{L}\sum_{i=1}^{n}\rho_{\tau_\ell}\big(\widetilde{Y}_i^{(0)} - \check{\mathbf{X}}_i^\top\boldsymbol{\beta}\big).$$

Output: The final orthogonal offset estimator $\boldsymbol{\beta}^\dagger$.

Appendix D.2 Asymptotic properties of the orthogonal offset

Consider the following partially linear model:

$$Y \;=\; \mathbf{X}^\top \boldsymbol{\beta}_0 \;+\; \mathbf{Z}^\top \boldsymbol{\alpha}_0(U) \;+\; \gamma_0(U) \;+\; \varepsilon,$$

and write $W = (U, \mathbf{Z}, \mathbf{X})$, $g_0(U, \mathbf{Z}) = \mathbf{Z}^\top \boldsymbol{\alpha}_0(U) + \gamma_0(U)$, and $\mathbf{m}_0(U, \mathbf{Z}) = \mathbb{E}[\mathbf{X} \mid U, \mathbf{Z}]$. Let the composite quantile levels be $\widetilde{\boldsymbol{\tau}} = (\tau_1, \ldots, \tau_L)^\top$, and denote by r_ℓ^* the τ_ℓ-quantile of ε, i.e., $\mathbb{P}(\varepsilon \leqslant r_\ell^*) = \tau_\ell$. By Assumption **B.1** (independence of ε and W), these are also the conditional τ_ℓ-quantiles given W. Define

$$\psi_\tau(r) := \tau - \mathbb{I}\{r \leqslant 0\}, \qquad \bar{\psi}(r) := \frac{1}{L} \sum_{\ell=1}^{L} \psi_{\tau_\ell}\big(r - r_\ell^*\big),$$

and let $\bar{f}_0(W) := L^{-1} \sum_{\ell=1}^{L} f_{\varepsilon|W}(r_\ell^* \mid W)$ denote the average conditional density of ε at the alignment points. For any $(\boldsymbol{\beta}, g, \mathbf{m})$, define the following orthogonal moment:

$$\phi(Y, W; \boldsymbol{\beta}, g, \mathbf{m}) := \bar{\psi}\big(Y - g(U, \mathbf{Z}) - \mathbf{X}^\top \boldsymbol{\beta}\big)\,\big(\mathbf{X} - \mathbf{m}(U, \mathbf{Z})\big).$$

Algorithm A.2 returns the cross-fitted estimator $\boldsymbol{\beta}^\dagger$ that solves the following out-of-fold moment equation:

$$\frac{1}{n} \sum_{i=1}^{n} \phi\Big(Y_i, W_i;\, \boldsymbol{\beta}^\dagger,\, \widehat{g}^{(-k(i))},\, \widehat{\mathbf{m}}^{(-k(i))}\Big) = \mathbf{0}, \tag{A-1}$$

where $k(i)$ denotes the fold index of observation i.

To facilitate the theoretical analysis, we impose the following assumptions.

Assumption O1 (Error quantiles and smoothness.) For each $\tau_\ell \in \widetilde{\boldsymbol{\tau}}$, the conditional distribution of ε satisfies $p(\varepsilon \leqslant 0 \mid W) = \tau_\ell$. The conditional density $f_{\varepsilon|W}(\cdot \mid W)$ exists, is bounded, and Lipschitz in a neighborhood of 0, and there exist constants $0 < c_f \leqslant C_f < \infty$ such that $c_f \leqslant \bar{f}_0(W) \leqslant C_f$ almost surely.

Assumption O2 (Design and information matrix) Let $\check{\mathbf{X}} := \mathbf{X} - \mathbf{m}_0(U, \mathbf{Z})$. Assume that $\mathbb{E}\|\mathbf{X}\|^{2+\delta} < \infty$ for some $\delta > 0$, $\mathbb{E}\|\check{\mathbf{X}}\check{\mathbf{X}}^\top\| < \infty$, and the matrix $J := \mathbb{E}\big[\bar{f}_0(W)\,\check{\mathbf{X}}\check{\mathbf{X}}^\top\big]$ is positive definite.

Assumption O3 (Kernels and bandwidths) The kernel K is bounded, symmetric, and has a finite second moment. The bandwidths $h, b \to 0$ with $nh, nb \to \infty$ and $nh^5, nb^5 \to 0$.

Assumption O4 (Cross-fitted nuisance rates) The estimators $\widehat{g}^{(-k)}$ and $\widehat{\mathbf{m}}^{(-k)}$ are trained on folds independent of the evaluation observation and satisfies

$$\|\widehat{g}^{(-k)} - g_0\|_{L_2} = o_p(n^{-1/4}), \qquad \|\widehat{\mathbf{m}}^{(-k)} - \mathbf{m}_0\|_{L_2} = o_p(n^{-1/4}).$$

Proposition D.1. *Consider the model and notation in Appendix A. Let $\boldsymbol{\beta}^\dagger$ be the cross-fitted solution to the moment equation* (A-1)*. Suppose Assumptions **O1-O4** hold. Then, as $n \to \infty$, we have*

(i) (Asymptotic linearity)

$$\sqrt{n}\,(\boldsymbol{\beta}^\dagger - \boldsymbol{\beta}_0) \;=\; J^{-1} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \bar{\psi}(\varepsilon_i)\,\check{\mathbf{X}}_i \;+\; o_p(1). \tag{A-2}$$

(ii) (Asymptotic normality)

$$\sqrt{n}\,(\boldsymbol{\beta}^\dagger - \boldsymbol{\beta}_0) \;\xrightarrow{\;d\;}\; \mathcal{N}\big(\mathbf{0},\; J^{-1}\Sigma J^{-1}\big), \tag{A-3}$$

where $\Sigma := \mathrm{Var}\big(\bar{\psi}(\varepsilon)\,\check{\mathbf{X}}\big)$.

Proof. We proceed in four steps.

Step 1: Identification and orthogonality. Define the following population moment:

$$M(\boldsymbol{\beta}, g, \mathbf{m}) \;:=\; \mathbb{E}\big[\phi(Y, W; \boldsymbol{\beta}, g, \mathbf{m})\big], \quad \phi(Y, W; \boldsymbol{\beta}, g, \mathbf{m}) = \bar{\psi}\big(Y - g(U, \mathbf{Z}) - \mathbf{X}^\top \boldsymbol{\beta}\big)\,\big(\mathbf{X} - \mathbf{m}(U, \mathbf{Z})\big).$$

At the truth $(\boldsymbol{\beta}_0, g_0, \mathbf{m}_0)$ we have $Y - g_0(U, \mathbf{Z}) - \mathbf{X}^\top \boldsymbol{\beta}_0 = \varepsilon$. By Assumption **O1**, $\mathbb{E}[\bar{\psi}(\varepsilon) \mid W] = L^{-1} \sum_\ell \{\tau_\ell - F_{\varepsilon|W}(r_\ell^* \mid W)\} = 0$, and hence $M(\boldsymbol{\beta}_0, g_0, \mathbf{m}_0) = \mathbf{0}$. Directional derivatives at $(\boldsymbol{\beta}_0, g_0, \mathbf{m}_0)$ are

$$\partial_{\boldsymbol{\beta}} M(\boldsymbol{\beta}_0, g_0, \mathbf{m}_0) = -\,\mathbb{E}\big[\bar{f}_0(W)\,\check{\mathbf{X}}\check{\mathbf{X}}^\top\big] \;=\; -J,$$
$$\partial_g M(\boldsymbol{\beta}_0, g_0, \mathbf{m}_0)[h_g] = \mathbb{E}\big[\bar{f}_0(W)\,h_g(W)\,(\mathbf{X} - \mathbf{m}_0(W))\big] = \mathbb{E}\big[\bar{f}_0(W)\,h_g(W)\,\mathbb{E}(\check{\mathbf{X}} \mid W)\big] = \mathbf{0},$$
$$\partial_{\mathbf{m}} M(\boldsymbol{\beta}_0, g_0, \mathbf{m}_0)[h_m] = -\,\mathbb{E}\big[\bar{\psi}(\varepsilon)\,h_m(W)\big] = -\,\mathbb{E}\big[\mathbb{E}(\bar{\psi}(\varepsilon) \mid W)\,h_m(W)\big] = \mathbf{0}.$$

Thus, the moment is Neyman-orthogonal to first-order perturbations of $(g, \mathbf{m})$. By Assumption **O2**, J is positive definite.

Step 2: Sample moment decomposition. Let $k(i)$ be the fold of observation i, and abbreviate $\widehat{g}_i = \widehat{g}^{(-k(i))}(U_i, \mathbf{Z}_i)$, $\widehat{\mathbf{m}}_i = \widehat{\mathbf{m}}^{(-k(i))}(U_i, \mathbf{Z}_i)$, $\delta g_i = \widehat{g}_i - g_0(U_i, \mathbf{Z}_i)$, and $\delta \mathbf{m}_i = \widehat{\mathbf{m}}_i - \mathbf{m}_0(U_i, \mathbf{Z}_i)$. The sample moment equation (A-1) reads

$$\mathbf{0} = \mathbb{E}_n\big[\phi(Y, W; \boldsymbol{\beta}^\dagger, \widehat{g}, \widehat{\mathbf{m}})\big] = \frac{1}{n} \sum_{i=1}^{n} \phi(Y_i, W_i; \boldsymbol{\beta}^\dagger, \widehat{g}_i, \widehat{\mathbf{m}}_i).$$

A first-order expansion in $\boldsymbol{\beta}$ around $\boldsymbol{\beta}_0$ yields

$$\mathbf{0} = \underbrace{\mathbb{E}_n\big[\phi(Y, W; \boldsymbol{\beta}_0, g_0, \mathbf{m}_0)\big]}_{(I)} + \underbrace{\partial_{\boldsymbol{\beta}} M(\boldsymbol{\beta}_0, g_0, \mathbf{m}_0)\,(\boldsymbol{\beta}^\dagger - \boldsymbol{\beta}_0)}_{(II)} + \underbrace{R_n}_{(III)}, \tag{A-4}$$

where R_n collects all terms due to replacing $(g_0, \mathbf{m}_0)$ by $(\widehat{g}, \widehat{\mathbf{m}})$ and the higher-order remainder of the linearization.

Step 3: Control of the nuisance-induced remainder R_n. Write $r_i(\boldsymbol{\beta}, g) = Y_i - g(U_i, \mathbf{Z}_i) - \mathbf{X}_i^\top \boldsymbol{\beta}$. For each ℓ, by Assumption **O1**, the conditional CDF $F_{\varepsilon|W}$ is differentiable with a bounded derivative in a neighborhood of r_ℓ^*. A first-order Taylor expansion of $F_{\varepsilon|W}$ at r_ℓ^* gives, for $\vartheta = \delta g_i$,

$$F_{\varepsilon|W}(r_\ell^* + \vartheta \mid W_i) - F_{\varepsilon|W}(r_\ell^* \mid W_i) = f_{\varepsilon|W}(r_\ell^* \mid W_i)\,\vartheta + \rho_{i\ell}, \qquad |\rho_{i\ell}| \leqslant \tfrac{1}{2} \|f'_{\varepsilon|W}\|_\infty\, \vartheta^2.$$

Define the following centered fluctuation:

$$\psi_{i\ell} := \mathbb{I}\{\varepsilon_i \leqslant r_\ell^* + \delta g_i\} - \mathbb{I}\{\varepsilon_i \leqslant r_\ell^*\} - \big[F_{\varepsilon|W}(r_\ell^* + \delta g_i \mid W_i) - F_{\varepsilon|W}(r_\ell^* \mid W_i)\big],$$

so $\mathbb{E}[\psi_{i\ell} \mid W_i, \delta g_i] = 0$ and $\mathbb{E}[\psi_{i\ell}^2 \mid W_i, \delta g_i] \leqslant F_{\varepsilon|W}(r_\ell^* + \delta g_i \mid W_i) - F_{\varepsilon|W}(r_\ell^* \mid W_i) \leqslant C\,|\delta g_i|$ by the Lipschitz property implied by Assumption **O1**. A direct algebra (adding and subtracting the conditional means in each composite-quantile summand) then yields the decomposition

$$R_n = \underbrace{\mathbb{E}_n\Big[\bar{f}_0(W_i)\,\delta g_i\,\check{\mathbf{X}}_i\Big]}_{R_{n,1}} + \underbrace{\mathbb{E}_n\Big[\big\{\bar{\psi}(\varepsilon_i + \delta g_i) - \bar{\psi}(\varepsilon_i) + \bar{f}_0(W_i)\,\delta g_i\big\}\,\check{\mathbf{X}}_i\Big]}_{R_{n,2}}$$
$$- \underbrace{\mathbb{E}_n\Big[\bar{\psi}(\varepsilon_i)\,\delta \mathbf{m}_i\Big]}_{R_{n,3}} + \underbrace{\mathbb{E}_n\Big[\big\{\bar{\psi}(\varepsilon_i + \delta g_i) - \bar{\psi}(\varepsilon_i)\big\}\,(-\delta \mathbf{m}_i)\Big]}_{R_{n,4}}.$$

By cross-fitting in Assumption **O4**, δg_i and $\delta \mathbf{m}_i$ are independent of (Y_i, W_i) on the evaluation fold. Hence $\mathbb{E}[R_{n,1}] = 0$ and, by a conditional variance bound together with $\mathbb{E}\|\check{\mathbf{X}}\|_2^2 < \infty$ from Assumption **O2**, $R_{n,1} = O_p\big(\|\delta g\|_{L_2}/\sqrt{n}\big) = o_p(n^{-1/2})$. For $R_{n,2}$, the Taylor remainder above implies $\|\bar{\psi}(\varepsilon + \delta g) - \bar{\psi}(\varepsilon) + \bar{f}_0\,\delta g\|_{L_2} = O\big(\|\delta g\|_{L_2}^2\big)$, so $R_{n,2} = O_p\big(\|\delta g\|_{L_2}^2\big) = o_p(n^{-1/2})$ under Assumption **O4**. Similarly, $\mathbb{E}[R_{n,3}] = 0$ and $R_{n,3} = O_p\big(\|\delta \mathbf{m}\|_{L_2}/\sqrt{n}\big) = o_p(n^{-1/2})$; and $R_{n,4} = O_p\big(\|\delta g\|_{L_2}\,\|\delta \mathbf{m}\|_{L_2}\big) = o_p(n^{-1/2})$. Consequently,

$$\|R_n\| = o_p(n^{-1/2}). \tag{A-5}$$

Step 4: Conclusion. Plugging (A-5) into (A-4) and using $\partial_{\boldsymbol{\beta}} M(\boldsymbol{\beta}_0, g_0, \mathbf{m}_0) = -J$ gives

$$-J\,(\boldsymbol{\beta}^\dagger - \boldsymbol{\beta}_0) = \mathbb{E}_n\big[\bar{\psi}(\varepsilon_i)\,\check{\mathbf{X}}_i\big] + o_p(n^{-1/2}).$$

Multiplying by J^{-1} yields the linear expansion (A-2). Since $\{\bar{\psi}(\varepsilon_i)\check{\mathbf{X}}_i\}_{i=1}^n$ are i.i.d. with finite second moment (Assumption **O2**), by the Lindeberg-Feller central limit theorem, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \bar{\psi}(\varepsilon_i)\check{\mathbf{X}}_i \;\xrightarrow{\;d\;}\; \mathcal{N}(\mathbf{0}, \Sigma), \quad \Sigma = \mathrm{Var}\big(\bar{\psi}(\varepsilon)\check{\mathbf{X}}\big).$$

Combining this with (A-2) yields (A-3). □

Remark D.2. Under one-dimensional local-linear smoothing with bandwidths $h, b \asymp n^{-1/5}$, the out-of-fold errors obey $\|\widehat{g}^{(-k)} - g_0\|_{L_2} = \|\widehat{\mathbf{m}}^{(-k)} - \mathbf{m}_0\|_{L_2} = O_p(n^{-2/5})$, which satisfy Assumption **O4**. A plug-in estimator of $J^{-1}\Sigma J^{-1}$ can be constructed by replacing population quantities with their empirical counterparts and using the cross-fitted residuals $\widehat{\varepsilon}_i := Y_i - \widehat{g}^{(-k(i))}(U_i, \mathbf{Z}_i) - \mathbf{X}_i^\top \boldsymbol{\beta}^\dagger$ and $\check{\mathbf{X}}_i = \mathbf{X}_i - \widehat{\mathbf{m}}^{(-k(i))}(U_i, \mathbf{Z}_i)$.

Appendix E Proof of Theorem 3.3

Proof. Recall that

$$\Delta \boldsymbol{S}_\lambda(\mathbf{X}_i, \vartheta_i) = \lambda\,\mathbf{X}_i\,\vartheta_i + \frac{1-\lambda}{L}\sum_{\ell=1}^L \mathbf{X}_i\Big[\mathbb{I}\{\varepsilon_i \leqslant r_\ell + \vartheta_i\} - \mathbb{I}\{\varepsilon_i \leqslant r_\ell\}\Big],$$

and define

$$a_{i,n}(t) := \mathbb{I}\{i \leqslant \lfloor nt \rfloor\} - \frac{\lfloor nt \rfloor}{n},\ \mathcal{R}_n(t) := \frac{1}{\sqrt{n}}\sum_{i=1}^n a_{i,n}(t)\,\Delta \boldsymbol{S}_\lambda(\mathbf{X}_i, \vartheta_i), t \in [q_0, 1-q_0].$$

We show that $\sup_{t\in[q_0,1-q_0]} \|\mathcal{R}_n(t)\|_2 = o_p(1)$.

Let $m = \lfloor nt \rfloor$. Then

$$\sum_{i=1}^n a_{i,n}(t)^2 = m\Big(1 - \frac{m}{n}\Big)^2 + (n-m)\Big(\frac{m}{n}\Big)^2 = n \cdot \frac{m}{n}\Big(1 - \frac{m}{n}\Big) \leqslant \frac{n}{4}.$$

Hence, for any vectors $\{\boldsymbol{\xi}_i\} \subset \mathbb{R}^d$,

$$\sup_{t\in[q_0,1-q_0]} \left\| \frac{1}{\sqrt{n}}\sum_{i=1}^n a_{i,n}(t)\,\boldsymbol{\xi}_i \right\|_2 \;\leqslant\; \sup_t \Big(\frac{1}{n}\sum_{i=1}^n a_{i,n}(t)^2\Big)^{1/2} \Big(\sum_{i=1}^n \|\boldsymbol{\xi}_i\|_2^2\Big)^{1/2} \;\leqslant\; \frac{1}{2}\Big(\sum_{i=1}^n \|\boldsymbol{\xi}_i\|_2^2\Big)^{1/2}. \tag{B.1}$$

By Assumption **B**, f_ε is continuously differentiable with a bounded derivative. Applying a first-order Taylor expansion to F_ε at r_ℓ, we obtain

$$F_\varepsilon(r_\ell + \vartheta_i) - F_\varepsilon(r_\ell) = f_\varepsilon(r_\ell)\,\vartheta_i + \rho_{i\ell}, \qquad |\rho_{i\ell}| \leqslant \tfrac{1}{2}\|f_\varepsilon'\|_\infty\,\vartheta_i^2.$$

Define the following centered fluctuation:

$$\psi_{i\ell} := \mathbb{I}\{\varepsilon_i \leqslant r_\ell + \vartheta_i\} - \mathbb{I}\{\varepsilon_i \leqslant r_\ell\} - \big[F_\varepsilon(r_\ell + \vartheta_i) - F_\varepsilon(r_\ell)\big],$$

so that $\mathbb{E}[\psi_{i\ell} \mid U_i, \mathbf{Z}_i, \vartheta_i] = 0$. By independence in Assumption **B**, $\{\psi_{i\ell}\}_i$ are independent across i conditional on $\{(U_i, \mathbf{Z}_i, \vartheta_i)\}_i$. Then

$$\Delta \boldsymbol{S}_\lambda(\mathbf{X}_i, \vartheta_i) = \underbrace{\Big(\lambda + \frac{1-\lambda}{L}\sum_{\ell=1}^L f_\varepsilon(r_\ell)\Big)\mathbf{X}_i\vartheta_i}_{=:\,A} + \underbrace{\frac{1-\lambda}{L}\sum_{\ell=1}^L \mathbf{X}_i\,\psi_{i\ell}}_{=:\,B} + \underbrace{\frac{1-\lambda}{L}\sum_{\ell=1}^L \mathbf{X}_i\,\rho_{i\ell}}_{=:\,C}.$$

Accordingly, we have $\mathcal{R}_n(t) = \mathcal{R}_{n,A}(t) + \mathcal{R}_{n,B}(t) + \mathcal{R}_{n,C}(t)$.

For term $\mathcal{R}_{n,A}(t)$, by (B.1) and Assumption **A** (bounded $\|\mathbf{X}_i\|_\infty \leqslant M$),

$$\sup_t \|\mathcal{R}_{n,A}(t)\|_2 \;\leqslant\; \frac{C}{2}\Big(\sum_{i=1}^n \|\mathbf{X}_i\|_2^2\,\vartheta_i^2\Big)^{1/2} \;\leqslant\; \frac{C'}{2}\Big(\sum_{i=1}^n \vartheta_i^2\Big)^{1/2} = C'\sqrt{n}\,\Big(\frac{1}{n}\sum_{i=1}^n \vartheta_i^2\Big)^{1/2}.$$

Assumption **C** implies $\mathbb{E}[\vartheta_i^2] = r_n^2$, hence by Markov's inequality $\frac{1}{n}\sum_{i=1}^n \vartheta_i^2 = O_p(r_n^2)$, and therefore

$$\sup_t \|\mathcal{R}_{n,A}(t)\|_2 = O_p(r_n) = o_p(1).$$

For the term $\mathcal{R}_{n,B}(t)$, it follows from (B.1), conditional independence, and the Lipschitz bound $|F_\varepsilon(u) - F_\varepsilon(v)| \leqslant C_f|u-v|$ (a consequence of $\sup_t |f_\varepsilon(t)| \leqslant C_f$ in Assumption **B**) that

$$\mathbb{E}\big[\psi_{i\ell}^2 \mid U_i, \mathbf{Z}_i, \vartheta_i\big] \leqslant \big|F_\varepsilon(r_\ell + \vartheta_i) - F_\varepsilon(r_\ell)\big| \leqslant C_f\,|\vartheta_i|.$$

Therefore,

$$\begin{aligned}
\mathbb{E}\Big[\sup_t \|\mathcal{R}_{n,B}(t)\|_2^2\Big] &\leqslant \frac{1}{4}\sum_{i=1}^n \mathbb{E}\left\|\frac{1-\lambda}{L}\sum_{\ell=1}^L \mathbf{X}_i\,\psi_{i\ell}\right\|_2^2 \\
&\leqslant C\sum_{i=1}^n \mathbb{E}\left[\|\mathbf{X}_i\|_2^2\,\frac{1}{L}\sum_{\ell=1}^L \mathbb{E}(\psi_{i\ell}^2 \mid U_i, \mathbf{Z}_i, \vartheta_i)\right] \\
&\leqslant C'\sum_{i=1}^n \mathbb{E}\big[\|\mathbf{X}_i\|_2^2\,|\vartheta_i|\big] \;\leqslant\; C''n\,\big(\mathbb{E}\|\mathbf{X}_i\|_2^4\big)^{1/2}\big(\mathbb{E}\vartheta_i^2\big)^{1/2} \\
&\leqslant C'''n\,r_n,
\end{aligned}$$

where we used the Cauchy-Schwarz inequality and $\|\mathbf{X}_i\|_\infty \leqslant M$ to ensure $\mathbb{E}\|\mathbf{X}_i\|_2^4 < \infty$ (Assumptions **A**, **D**). Hence

$$\sup_{t\in[q_0,1-q_0]} \|\mathcal{R}_{n,B}(t)\|_2 = O_p(r_n^{1/2}) = o_p(1).$$

For the term $\mathcal{R}_{n,C}(t)$, it follows from $|\rho_{i\ell}| \leqslant \frac{1}{2}\|f_\varepsilon'\|_\infty\,\vartheta_i^2$ (Assumption **B**) and $\|\mathbf{X}_i\|_\infty \leqslant M$ that

$$\sup_t \|\mathcal{R}_{n,C}(t)\|_2 \leqslant \frac{1}{\sqrt{n}}\sum_{i=1}^n |a_{i,n}(t)|\,\frac{1-\lambda}{L}\sum_{\ell=1}^L \|\mathbf{X}_i\|_2\,|\rho_{i\ell}| \leqslant C\,\frac{1}{\sqrt{n}}\sum_{i=1}^n \vartheta_i^2 = C\,\sqrt{n}\,\left(\frac{1}{n}\sum_{i=1}^n \vartheta_i^2\right).$$

As above, $\frac{1}{n}\sum_i \vartheta_i^2 = O_p(r_n^2)$, and hence

$$\sup_t \|\mathcal{R}_{n,C}(t)\|_2 = O_p(\sqrt{n}\,r_n^2) = o_p(1)$$

by Assumption **C**.

Combining the bounds for $\mathcal{R}_{n,A}(t)$, $\mathcal{R}_{n,B}(t)$, and $\mathcal{R}_{n,C}(t)$, we have

$$\begin{aligned}
\sup_{t\in[q_0,1-q_0]} \|\mathcal{R}_n(t)\|_2 &\leqslant \sup_t \|\mathcal{R}_{n,A}(t)\|_2 + \sup_t \|\mathcal{R}_{n,B}(t)\|_2 + \sup_t \|\mathcal{R}_{n,C}(t)\|_2 \\
&= O_p(r_n) + O_p(r_n^{1/2}) + O_p(\sqrt{n}\,r_n^2) = o_p(1),
\end{aligned}$$

since $r_n \to 0$ and $\sqrt{n}\,r_n^2 \to 0$ by Assumption **C**. We finish the proof of Theorem 3.3. □