Simultaneous jump detection for multiple sequences via screening and multiple testing

Shengji Jia and Chunming Zhang

Shanghai Lixin University of Accounting and Finance

University of Wisconsin-Madison

Supplementary Material

S1 Conditions and proofs of main results

The following technical conditions are imposed. They are not the weakest possible, but facilitate the proofs. In Section 3.3, when we detect the simultaneous jump points in multiple sequences, we need to replace the conditions A1, A2 and B1 by the corresponding conditions A1′, A2′ and B1′ respectively. Without loss of generality, we assume $\mathcal{T} = [0, 1]$.

A1. $\alpha(\cdot)$ has a continuous second derivative;

A1′. $\alpha_k(\cdot)$ has a continuous second derivative, $k = 1, \ldots, m$;

A2. $\sigma^2(\cdot)$ has a continuous second derivative;

Corresponding author: Chunming Zhang. E-mail: *cmzhang@stat.wisc.edu*.

A2$'$. $\sigma_k^2(\cdot)$ has a continuous second derivative, $k = 1, \ldots, m$;

A3. The density function $f(\cdot)$ of $\{T_i : i = 1, \ldots, n\}$ is Lipschitz-continuous and bounded away from 0;

A4. $K^+(\cdot)$ is a right-continuous function with bounded variation on $[0, 1]$;

A5. $K^+(\cdot)$ is compactly supported with $K^+(0) > 0$;

A6. $h_1 \to 0$, $nh_1/\log(h_1^{-1}) \to \infty$ and $\log(h_1^{-1})/\log(\log(n)) \to \infty$ as $n \to \infty$;

A7. $J < \infty$ is fixed and the jumps $\{\tau_1 < \tau_2 < \cdots < \tau_J\}$ satisfy $\min_{1 \le j \le J+1}(\tau_j - \tau_{j-1}) > \xi$, for some $\xi > 0$, where $\tau_0 = 0$, $\tau_{J+1} = 1$.

Condition A7 means that the adjacent jumps cannot be too close to each other. This condition can be relaxed, as long as the number of jumps J diverges at a slower rate than the sample size n. The following technical conditions are imposed in Section 3.1 when we apply the profile likelihood estimation procedure in the partially linear model.

B1. $\sup_{t \in \mathcal{T}} \mathrm{E}\{|\varepsilon(t)|^{4+\delta_0}\} < \infty$ for some $\delta_0 > 0$;

B1$'$. $\sup_{t \in \mathcal{T}} \mathrm{E}\{|\varepsilon_k(t)|^{4+\delta_0}\} < \infty$ for some $\delta_0 > 0$, $k = 1, \ldots, m$;

B2. $nh_2^8 \to 0$ and $nh_2^2/\log^3(n) \to \infty$ as $n \to \infty$;

B3. $h_3 \to 0$ and $nh_3/\log(n) \to \infty$ as $n \to \infty$;

B4. $K(\cdot)$ is a symmetric continuously differentiable probability density function on $[-1,1]$.

To control the upper bound of the difference process $L(t)$ at continuity points, the strong uniform consistency result for the locally-linear regression estimator (Blondin, 2007) is employed in the proof of following Lemma 1, which is based upon modern empirical process theory.

Lemma 1. *Assume conditions* A1–A6, *and suppose that* $\mu(\cdot)$ *is a continuous function on the compact set* $\mathcal{C} \subseteq [0,1]$. *Then we have*

$$\left| \left\{ \frac{nh_1}{2\log(h_1^{-1})} \right\}^{1/2} \sup_{t\in\mathcal{C}} |\widehat{\mu}_+(t) - \mathrm{E}\{\widehat{\mu}_+(t)\}| - \Lambda \right| = o(1) \quad \text{a.s.}, \tag{S1.1}$$

where

$$\Lambda = \sup_{t\in\mathcal{C}} \left\{ \frac{\sigma^2(t)}{f(t)} \int (K^+)^2(u)\mathrm{d}u \right\}^{1/2}.$$

Proof: Replace $K(\cdot)$ by $K^+(\cdot)$ in Theorem 3.1 of Blondin (2007). ■

The following Lemma 2 shows the asymptotic normality of $\widehat{\mu}_+(t)$. Recall that $\mu_j = \int u^j K^+(u)\mathrm{d}u$ and $\nu_j = \int u^j (K^+)^2(u)\mathrm{d}u$.

Lemma 2. *Assume conditions* A1–A6. *If there is no change point on* $(\tau_j, \tau_j + h_1)$, *then we have*

$$\sqrt{nh_1}\Big\{\widehat{\mu}_+(\tau_j) - \mu_+(\tau_j) - \frac{1}{2}C\alpha''(\tau_j)h_1^2\Big\} \xrightarrow{\mathcal{D}} N\Big(0, \frac{V\sigma^2(\tau_j)}{f(\tau_j)}\Big), \quad j = 1,\ldots,J, \tag{S1.2}$$

where

$$V = \frac{\mu_2^2\nu_0 - 2\mu_1\mu_2\nu_1 + \mu_1^2\nu_2}{(\mu_0\mu_2 - \mu_1^2)^2}, \quad \text{and} \quad C = \frac{\mu_2^2 - \mu_1\mu_3}{\mu_0\mu_2 - \mu_1^2}.$$

Proof: Take $c = 0$ in Theorem 3.3 of Fan and Gijbels (1996). ■

Lemma 3. *Assume conditions* A1–A6. *Assume there is only one jump point* τ *(i.e.,* $J = 1$*). Denote the estimator of the unique jump point* τ *by the global maximizer* $\widehat{\tau}$ *of* $L(t)$*. Then we have*

$$nh_1\Big(\{\widehat{\mu}_+(\widehat{\tau}) - \widehat{\mu}_-(\widehat{\tau})\} - \{\widehat{\mu}_+(\tau) - \widehat{\mu}_-(\tau)\}\Big) = O_{\mathrm{P}}(1). \tag{S1.3}$$

Proof: Take $\alpha(n, h_1) = \beta(n, h_1) = nh_1$ in Theorem 3.1 and Lemma 3.1 of Grégoire and Hamrouni (2002). ■

Proof of Proposition 1. For any $\widetilde{\tau}_j \in \mathcal{S}_\lambda$, $1 \le j \le \widetilde{J}$, without loss of generality, let $\widetilde{\tau}_j = \omega_i$ for some $i \in \{1, \ldots, q\}$. Then according to (2.6) and (2.7), we have

$$\begin{aligned} L(\omega_i) &\ge \lambda > \lambda', \\ \omega_i &\notin \bigcup_{j=1}^{i-1}(\omega_j - h_1, \omega_j + h_1), \end{aligned}$$

which means $\omega_i \in \mathcal{S}_{\lambda'}$, and thus $\mathcal{S}_\lambda \subseteq \mathcal{S}_{\lambda'}$. ■

Proof of Theorem 1. It suffices to prove part (b); the proof of part (a) follows from (S1.12) directly. Let $\mathcal{I}_j := (\tau_j - h_1/2, \tau_j + h_1/2)$ be the neighborhood of τ_j, $j = 1, \ldots, J$, and define $\mathcal{C} := [0, 1] \cap \{\mathcal{I}_1 \cup \cdots \cup \mathcal{I}_J\}^c$, where

A^c denotes the complement of a set A. It is obvious that $\mu(\cdot)$ is continuous on the compact set $\mathcal{C}$. To avoid confusion, in the remaining proof, we will use $\lambda \equiv \lambda_n$ and $\widetilde{J} \equiv \widetilde{J}_n$ to emphasize the dependence on sample size n.

For any $t \in \mathcal{C}$, the estimator $\widehat{\mu}_+(t)$ does not involve any jumps, thus according to the results for standard locally-linear regression (Theorem 11.6.3 of Bickel and Doksum (2016)), $\sup_{t\in\mathcal{C}} |\mathrm{E}\{\widehat{\mu}_+(t)\} - \mu_+(t)| = O(h_1^2)$, together with (S1.1), we have

$$\sup_{t\in\mathcal{C}} |\widehat{\mu}_+(t) - \mu_+(t)| = O\Big(h_1^2 + \Big\{\frac{\log(h_1^{-1})}{nh_1}\Big\}^{1/2}\Big) \quad \text{a.s.,} \tag{S1.4}$$

and similarly,

$$\sup_{t\in\mathcal{C}} |\widehat{\mu}_-(t) - \mu_-(t)| = O\Big(h_1^2 + \Big\{\frac{\log(h_1^{-1})}{nh_1}\Big\}^{1/2}\Big) \quad \text{a.s..} \tag{S1.5}$$

Using $\mu_-(t) = \mu_+(t)$, $\forall t \in \mathcal{C}$, together with (S1.4), (S1.5) and the triangular inequality, we have

$$\sup_{t\in\mathcal{C}} L(t) = \sup_{t\in\mathcal{C}} |\widehat{\mu}_+(t) - \widehat{\mu}_-(t)| = O\Big(h_1^2 + \Big\{\frac{\log(h_1^{-1})}{nh_1}\Big\}^{1/2}\Big) = o(\lambda_n) \quad \text{a.s..} \tag{S1.6}$$

By the definition of $\widetilde{\tau}_j$, together with (S1.6), there exists $N_1 > 0$ such that

$$L(\widetilde{\tau}_j) \geq \lambda_n > \sup_{t\in\mathcal{C}} L(t) \quad \text{a.s.,} \quad j = 1, \ldots, \widetilde{J}_n,$$

when $n > N_1$, which implies

$$\{\widetilde{\tau}_1, \ldots, \widetilde{\tau}_{\widetilde{J}_n}\} \subseteq \mathcal{C}^c = \mathcal{I}_1 \cup \cdots \cup \mathcal{I}_J \quad \text{a.s.,} \tag{S1.7}$$

when $n > N_1$. (S1.7) together with the fact that $\min_{1\le i,j\le \widetilde{J}_n} |\widetilde{\tau}_i - \widetilde{\tau}_j| > h_1 = \max_{1\le j\le J} |\mathcal{I}_j|$ imply that each neighborhood $\mathcal{I}_j$ can not contain more than one estimator $\widetilde{\tau}_j$. Thus when $n > N_1$, we have

$$\widetilde{J}_n \le J \quad \text{a.s..} \tag{S1.8}$$

On the other hand, since $h_1 \to 0$ and $\lambda_n \to 0$, and by condition A7, there exists $N_2 > 0$ such that $\min_{1\le j\le J+1}(\tau_j - \tau_{j-1}) \ge \xi > 2h_1$ and $\min_{1\le j\le J} |\beta_j| > 2\lambda_n$ when $n > N_2$. Thus for $j = 1, \ldots, J$, we have

$$L(\tau_j) = |\widehat{\mu}_+(\tau_j) - \widehat{\mu}_-(\tau_j)| > \frac{1}{2}|\mu_+(\tau_j) - \mu_-(\tau_j)| > \lambda_n, \quad \text{a.s.,} \tag{S1.9}$$

when $n > N_2$. Define $\widetilde{\mathcal{I}}_j := (\widetilde{\tau}_j - h_1, \widetilde{\tau}_j + h_1)$ as the neighborhood of the estimator $\widetilde{\tau}_j$, $j = 1, \ldots, \widetilde{J}_n$, and let $\widetilde{\mathcal{C}} := [0,1] \cap (\widetilde{\mathcal{I}}_1 \cup \cdots \cup \widetilde{\mathcal{I}}_{\widetilde{J}_n})^c$. We claim that

$$\sup_{t\in\widetilde{\mathcal{C}}} L(t) \le \lambda_n, \tag{S1.10}$$

when $n > N_2$. Otherwise, if there is a local maximizer $\omega \in \widetilde{\mathcal{C}}$ such that $L(\omega) > \lambda_n$, according to the definition of $\widetilde{\mathcal{C}}$, we have $\min_{1\le j\le \widetilde{J}_n} |\omega - \widetilde{\tau}_j| > h_1$, and thus $\omega \in \mathcal{S}_\lambda$, which is impossible. (S1.9) and (S1.10) imply that

$$\sup_{t\in\widetilde{\mathcal{C}}} L(t) \le \lambda_n < L(\tau_j) \quad \text{a.s.,} \quad j = 1, \ldots, J,$$

when $n > N_2$, that means

$$\{\tau_1, \ldots, \tau_J\} \subseteq \widetilde{\mathcal{C}}^c = \widetilde{\mathcal{I}}_1 \cup \cdots \cup \widetilde{\mathcal{I}}_{\widetilde{J}_n}, \quad \text{a.s.,} \tag{S1.11}$$

when $n > N_2$. Again each $\widetilde{\mathcal{I}}_j$ can not contain more than one jump point since $\min_{1\leq j\leq J+1}(\tau_j - \tau_{j-1}) > 2h_1 = \max_{1\leq j\leq \widetilde{J}_n} |\widetilde{\mathcal{I}}_j|$. Thus we have

$$J \leq \widetilde{J}_n, \quad \text{a.s.,} \tag{S1.12}$$

when $n > N_2$. Now (2.9) follows from (S1.8) and (S1.12) if we take $N = \max\{N_1, N_2\}$. Next, to show (2.10), under the event $\{\widetilde{J}_n = J\}$, each $\mathcal{I}_j$ contains only one estimator $\widetilde{\tau}_j$. For the problem of detecting a single jump point on the fixed interval $[\tau_j - \xi/2, \tau_j + \xi/2]$, it's easy to check that all the conditions in Theorem 3.2 of Grégoire and Hamrouni (2002) are satisfied, and thus we have $n(\widetilde{\tau}_j - \tau_j) = O_{\mathrm{P}}(1)$. ■

Proof of Theorem 2. Since $K^-(u) = K^+(-u)$, we have

$$\int u^j K^-(u)\mathrm{d}u = (-1)^j \int u^j K^+(u)\mathrm{d}u, \quad \int u^j (K^-)^2(u)\mathrm{d}u = (-1)^j \int u^j (K^+)^2(u)\mathrm{d}u.$$

Therefore after replacing $K^+(u)$ by $K^-(u)$, the expressions of V and C in Lemma 2 will not change. Thus we can derive the asymptotic normality for $\widehat{\mu}_-(\tau_j)$:

$$\sqrt{nh_1}\Big\{\widehat{\mu}_-(\tau_j) - \mu_-(\tau_j) - \frac{1}{2}C\alpha''(\tau_j)h_1^2\Big\} \xrightarrow{\mathcal{D}} N\Big(0, \frac{V\sigma^2(\tau_j)}{f(\tau_j)}\Big). \tag{S1.13}$$

Since $\widehat{\mu}_+(\tau_j)$ and $\widehat{\mu}_-(\tau_j)$ only utilize the data points located on the right and left side of τ_j, they are independent. By (S1.2) and (S1.13), the asymptotic distribution of $\widehat{\mu}_+(\tau_j) - \widehat{\mu}_-(\tau_j)$ is

$$\sqrt{nh_1}\big\{\widehat{\mu}_+(\tau_j) - \widehat{\mu}_-(\tau_j) - \beta_j\big\} \xrightarrow{\mathcal{D}} N\Big(0, \frac{2V\sigma^2(\tau_j)}{f(\tau_j)}\Big). \qquad \text{(S1.14)}$$

For the problem of detecting a single jump point on the fixed interval $[\tau_j - \xi/2, \tau_j + \xi/2]$, it is easy to check that the conditions in Lemma 3 are satisfied, and thus we have

$$\sqrt{nh_1}\Big(\{\widehat{\mu}_+(\widetilde{\tau}_j) - \widehat{\mu}_-(\widetilde{\tau}_j)\} - \{\widehat{\mu}_+(\tau_j) - \widehat{\mu}_-(\tau_j)\}\Big) = O_{\mathrm{P}}\Big(\frac{1}{\sqrt{nh_1}}\Big) = o_{\mathrm{P}}(1). \qquad \text{(S1.15)}$$

Now Theorem 2 follows from (S1.14) and (S1.15). ■

Proof of Theorem 3. First we show

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{\mathcal{D}} N\Big(0, \frac{\mathrm{E}\big\{\sigma^2(T_1)\mathrm{var}(Z_1 \mid T_1)\big\}}{\big[\mathrm{E}\big\{\mathrm{var}(Z_1 \mid T_1)\big\}\big]^2}\Big). \qquad \text{(S1.16)}$$

By (3.18), we have

$$\sqrt{n}(\widehat{\beta} - \beta) = \sqrt{n}(\widetilde{\boldsymbol{Z}}^T\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}(\mathbf{I} - \mathbf{S})(\boldsymbol{m} + \boldsymbol{\varepsilon}). \qquad \text{(S1.17)}$$

By taking $p = 1$, $q = 1$, and $\mathbf{X} = (1, \ldots, 1)^T$ in Lemma A.2 of Fan and Huang (2005), and applying the law of large numbers and the property of conditional expectation, we have

$$\begin{aligned} n^{-1}\widetilde{\boldsymbol{Z}}^T\widetilde{\boldsymbol{Z}} &= n^{-1}\sum_{i=1}^{n}\big\{Z_i - \mathrm{E}(Z_i \mid T_i)\big\}^2\{1 + O_{\mathrm{P}}(c_n)\} \\ &\xrightarrow{\mathrm{P}} \mathrm{E}\big\{Z_1 - \mathrm{E}(Z_1 \mid T_1)\big\}^2 \end{aligned}$$

$$= \quad \mathrm{E}\big\{\mathrm{var}(Z_1 \mid T_1)\big\}, \tag{S1.18}$$

where $c_n = \big\{\log(1/h_2)/(nh_2)\big\}^{1/2} + h_2^2$. Lemma A.4 of Fan and Huang (2005) and condition B2 imply that

$$\sqrt{n}(\widetilde{\boldsymbol{Z}}^T\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}(\mathbf{I}-\mathbf{S})\boldsymbol{m} = O_{\mathrm{P}}(\sqrt{n}c_n^2) = o_{\mathrm{P}}(1). \tag{S1.19}$$

Using the derivation similar to that of (S1.18), and the central limit theorem, we have

$$\begin{aligned} \frac{1}{\sqrt{n}}\widetilde{\boldsymbol{Z}}^T(\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \big\{Z_i - \mathrm{E}(Z_i \mid T_i)\big\}\varepsilon(T_i)\{1+o_{\mathrm{P}}(1)\} \\ &\xrightarrow{\mathcal{D}} N\Big(0, \mathrm{E}\big\{\sigma^2(T_1)\mathrm{var}(Z_1 \mid T_1)\big\}\Big). \end{aligned} \tag{S1.20}$$

Therefore (S1.16) follows after we plug (S1.18)−(S1.20) into (S1.17). Next, we show Theorem 3. By (3.19), the Wald statistic W can be decomposed in the following way:

$$W = \frac{n\widehat{\beta}^2\big[\mathrm{E}\big\{\mathrm{var}(Z_1 \mid T_1)\big\}\big]^2}{\mathrm{E}\big\{\sigma^2(T_1)\mathrm{var}(Z_1 \mid T_1)\big\}} \cdot \frac{\big(n^{-1}\widetilde{\boldsymbol{Z}}^T\widetilde{\boldsymbol{Z}}\big)^2}{\big[\mathrm{E}\big\{\mathrm{var}(Z_1 \mid T_1)\big\}\big]^2} \cdot \frac{\mathrm{E}\big\{\sigma^2(T_1)\mathrm{var}(Z_1 \mid T_1)\big\}}{n^{-1}\widetilde{\boldsymbol{Z}}^T\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{Z}}}. \tag{S1.21}$$

According to (S1.16), under H_0: $\beta = 0$, we have

$$\frac{\big[\mathrm{E}\big\{\mathrm{var}(Z_1 \mid T_1)\big\}\big]^2}{\mathrm{E}\big\{\sigma^2(T_1)\mathrm{var}(Z_1 \mid T_1)\big\}}\big(\sqrt{n}\widehat{\beta}\big)^2 \xrightarrow{\mathcal{D}} \chi_1^2. \tag{S1.22}$$

An application of Proposition 1 of Li (2011) implies that

$$\max_{1\le i\le n} |\widehat{\sigma}^2(T_i) - \sigma^2(T_i)| = O_{\mathrm{P}}(d_n), \tag{S1.23}$$

where $d_n = \big\{\log(n)/(nh_2)\big\}^{1/2} + h_2^2 + \big\{\log(n)/(nh_3)\big\}^{1/2} + h_3^2$. Using the derivation similar to that of (S1.18), and the law of large numbers, by (S1.23) and the fact that $\widetilde{\boldsymbol{Z}}$ and $\sigma^2(\cdot)$ are bounded, we have

$$\begin{aligned} n^{-1}\widetilde{\boldsymbol{Z}}^T\widehat{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{Z}} &= n^{-1}\widetilde{\boldsymbol{Z}}^T\boldsymbol{\Sigma}\widetilde{\boldsymbol{Z}} + n^{-1}\widetilde{\boldsymbol{Z}}^T\big(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\big)\widetilde{\boldsymbol{Z}} \\ &= n^{-1}\sum_{i=1}^{n}\big\{Z_i - \mathrm{E}(Z_i \mid T_i)\big\}^2\sigma^2(T_i)\{1 + O_{\mathrm{P}}(c_n)\} + O_{\mathrm{P}}(d_n) \\ &\overset{\mathrm{P}}{\to} \mathrm{E}\Big[\big\{Z_1 - \mathrm{E}(Z_1 \mid T_1)\big\}^2\sigma^2(T_1)\Big] \\ &= \mathrm{E}\big\{\sigma^2(T_1)\mathrm{var}(Z_1 \mid T_1)\big\}, \end{aligned} \tag{S1.24}$$

where $\boldsymbol{\Sigma} = \mathrm{diag}\{\sigma^2(T_1), \ldots, \sigma^2(T_n)\}$. Theorem 3 follows after we plug (S1.18), (S1.22) and (S1.24) into (S1.21). ■

Proof of Theorem 4. The proof is similar to Theorem 1 except that (S1.6) is replaced by

$$\begin{aligned} \sup_{t\in\mathcal{C}} L^{\mathrm{multi}}(t) &= \sup_{t\in\mathcal{C}}\sum_{k=1}^{m}\big\{\widehat{\mu}_{k;+}(t) - \widehat{\mu}_{k;-}(t)\big\}^2 \\ &\le \sum_{k=1}^{m}\sup_{t\in\mathcal{C}}\big\{\widehat{\mu}_{k;+}(t) - \widehat{\mu}_{k;-}(t)\big\}^2 \\ &= O\Big(h_1^4 + \frac{\log(h_1^{-1})}{nh_1}\Big) = o(\lambda) \quad \text{a.s.} \end{aligned}$$

according to the conditions for λ in Theorem 4. ■

Proof of Proposition 2. First, we show part (a). The multiple sample Wald statistics $\{W_j^{\mathrm{multi}}\}_{j=1}^{\widetilde{J}}$ are independent because the sets $\mathcal{N}_j$'s defined in (3.22) are disjoint. According to Theorem 3, under the null hypothesis $H_{0,j}$ in (3.27), we have $W_{k,j} \xrightarrow{\mathcal{D}} \chi^2_1$, $k = 1, \ldots, m$. Since all the m sequences are independent, the statistics $\{W_{k,j}\}_{k=1}^m$ are also independent. Therefore under $H_{0,j}$, we have,

$$W_j^{\text{multi}} = \sum_{k=1}^{m} W_{k,j} \xrightarrow{\mathcal{D}} \chi^2_m.$$

Second, we show part (b). Similar to part (a), the single-index modulated p-values $\{p_j^{\text{SIM}}\}_{j=1}^{\widetilde{J}}$ are independent because the sets $\mathcal{N}_j$'s are disjoint. Under the null hypothesis $H_{0,j}$ in (3.27), the p-value $p_{k,j}$ asymptotically follows the uniform distribution on $[0,1]$, and thus $\Phi^{-1}(p_{k,j}) \xrightarrow{\mathcal{D}} N(0,1)$, $k = 1, \ldots, m$. Since all the m sequences are independent, $\{\Phi^{-1}(p_{k,j})\}_{k=1}^m$ are the independent standard normal random variables. Now $\sum_{k=1}^m w_k^2 = 1$ implies that the linear combination

$$\sum_{k=1}^{m} w_k \Phi^{-1}(p_{k,j}) \xrightarrow{\mathcal{D}} N(0,1)$$

under $H_{0,j}$. Thus $p_j^{\text{SIM}} = \Phi(\sum_{k=1}^m w_k \Phi^{-1}(p_{k,j})) \xrightarrow{\mathcal{D}} \text{unif}(0,1)$, which completes the proof. ■

S2 Additional simulation

In this simulation, we will investigate the finite sample performance of the Wald test statistic W in (3.19). First, we consider the partially linear model

$$Y(t) = \alpha(t) + \beta \mathrm{I}(t > t^*) + \varepsilon(t), \tag{S2.25}$$

where the sample size $n = 100$, and $T_i = i/n, i = 1, \ldots, n$. We set the location of the potential discontinuity point $t^* = 0.6$, the nonparametric component $\alpha(t) = e^{-t}$, and $\beta = 0.1 \times \theta$, $\theta \in \{0, 1, \ldots, 5\}$, and the case $\theta = 0$ corresponds to the null hypothesis H_0 in (3.14). Suppose that $\varepsilon(t)$ is a Gaussian random process with mean 0 and $\sigma(t) = 0.1(1 + t)$, $t \in [0, 1]$. We conduct the simulation 500 times, with the significance level $\alpha = 0.05$. For each simulated data, we take the Epanechnikov kernel $K(u) = 0.75(1 - u^2)_+$, $u \in [-1, 1]$, the bandwidth $h_2 \in \{0.09, 0.12, 0.15\}$, and h_3 is selected automatically by the R package "np" (Li and Racine, 2007).

Figure 1 shows the QQ plots and powers of the Wald test statistic (3.19) with different bandwidths over 500 simulations. The plots depict that the Wald test statistic W closely follows the χ_1^2 distribution, which is consistent with our asymptotic theory. The power functions increase rapidly as β increases, which in turn shows that the Wald test statistic proposed in Section 3.1 works well. Besides, the results are quite stable with different

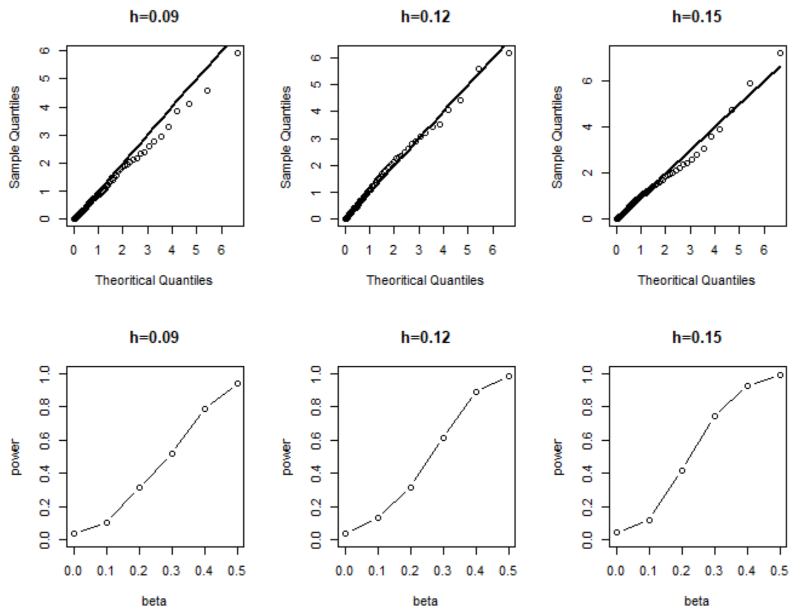

Figure 1: *QQ plots and powers of the Wald test statistic with different bandwidths.*

bandwidths, thus the proposed testing procedure is not sensitive to the choice of the bandwidth h_2, and there is a wide range for selecting the tuning parameters.

Next, we investigate the impact of estimation error in the potential jump estimator $\widetilde{\tau}$ on the performance of the Wald test statistic. Notably, $\widetilde{\tau}$ is derived from the identical dataset utilized for testing purposes. To accomplish this, we undertake a simulation study similar to our previous endeavor. However, a crucial distinction lies in the fact that the candidate $\widetilde{\tau}$ is initially obtained through the screening procedure. Subsequently, we proceed to test the hypothesis (3.12) with $t^* = \widetilde{\tau}$.

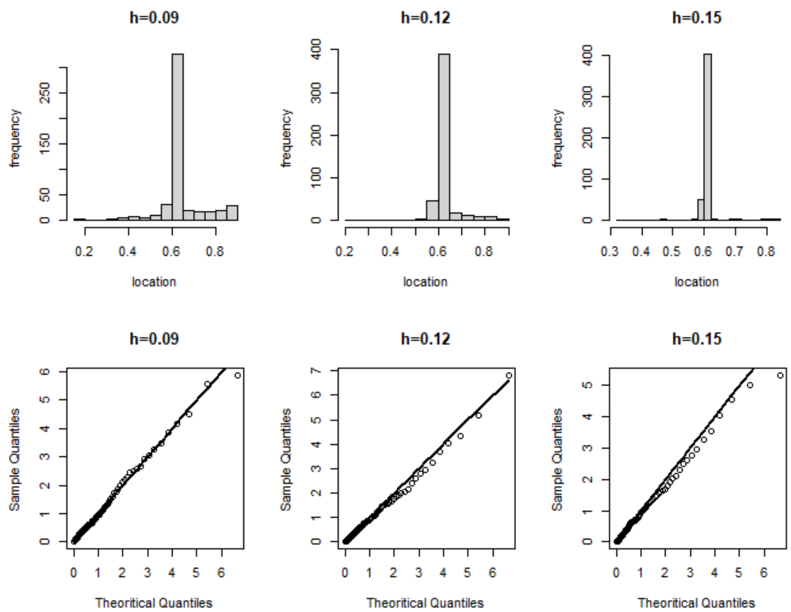

Figure 2: *Histograms of the jump estimator* $\widetilde{\tau}$ *in* (S2.25), *with* $\beta = 0.5$ *and QQ plots of the Wald test statistic for testing the hypothesis (3.12), with* $t^* = \widetilde{\tau}$.

Figure 2 depicts the histograms of the jump estimator $\widetilde{\tau}$ and QQ plots of the Wald test statistic across 500 simulations, with varying bandwidths. From the plots, we find that the jump estimators cluster around the true jump point 0.6. As the bandwidth h_2 increases, the estimator's precision also enhances since more observations are exploited in the neighborhood. Notably, compared with Figure 1, the influence of estimation error in $\widetilde{\tau}$ on the Wald test statistic appears to be minimal. This observation can be attributed to the convergence rate of the jump estimator $\widetilde{\tau}$, which is $O_{\mathrm{P}}(n^{-1})$

in Theorem 1, lower than the standard convergence rate $O_{\mathrm{P}}(n^{-1/2})$ of $\widehat{\beta}$ in the Wald test statistic (3.19). A similar argument of the post-selection inference for the classic screening procedure can be found in (S1.15).

References

[1] Bickel, P. J. and Doksum, K. A. (2016). *Mathematical Statistics: Basic Ideas and Selected Topics, Volume II.* Chapman and Hall, New York.

[2] Blondin, D. (2007). Rates of strong uniform consistency for local least squares kernel regression estimators. *Statist. Probab. Lett.* **77**, 1526–1534.

[3] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, New York.

[4] Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying coefficient partially linear models. *Bernoulli* **11**, 1031–1059.

[5] Grégoire, G. and Hamrouni, Z. (2002). Change point estimation by local linear smoothing. *J. Multivariate Anal.* **83**, 56–83.

[6] Li, Q. and Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice.* Princeton University Press, New Jersey.

[7] Li, Y. (2011). Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. *Biometrika* **98**, 355–370.