SCIENCE CHINA
Mathematics

• ARTICLES •

August 2014 Vol. 57 No. 8: 1701–1712
doi: 10.1007/s11425-014-4820-4

Estimator of a change point in single index models

JIN BaiSuo[1], DONG CuiLing[1,2], TAN ChangChun[3,*] & MIAO BaiQi[1]

[1]*Department of Statistics and Finance, University of Science and Technology of China, Hefei* 230026, *China;*
[2]*School of Mathematical Sciences, Xinjiang Normal University, Urumqi* 830054, *China;*
[3]*School of Mathematics, Hefei University of Technology, Hefei* 230009, *China*
Email: jbs@ustc.edu.cn, dcling@mail.ustc.edu.cn, cctan@ustc.edu.cn, bqmiao@ustc.edu.cn

Received February 6, 2013; accepted June 24, 2013; published online May 12, 2014

Abstract This paper considers the problem of change point in single index models. In order to obtain asymptotically valid confidence intervals for the estimation of the change point, the convergence rate and asymptotic distribution of the change point estimate is studied. Some simulation results are presented which show that the numerical performance of our estimator is satisfactory.

Keywords single index model, change point, convergence rate, asymptotic distribution

MSC(2010) 62F12, 62G20

Citation: Jin B S, Dong C L, Tan C C, et al. Estimator of a change point in single index models. Sci China Math, 2014, 57: 1701–1712, doi: 10.1007/s11425-014-4820-4

1 Introduction

The question of the instability of model parameters is a common phenomenon in the observational data analysis for signal processing, quality control, fault detection, finance, security, and clinical medicine. The statistical theory of change point was researched in many literatures (e.g., [5, 8, 9, 12, 13]). Model parametric changes can be initiated by policy decisions or permanent changes in resources, population or the society. Failure to take into account parameter changes can lead to huge forecasting errors and make the model unreliable.

The purpose of this paper is to study the parametric change problem in a single index model with an unknown change point. Cao et al. [6] studied the problem and obtained an asymptotic distribution of the test statistic under the null hypothesis, which focused on test rather than estimation of a change point in single index models. In this paper, we further research the asymptotic distribution of the estimation of a change point under an alternative hypothesis, which can be applied to obtain the mean and variance of the estimation of change point or to construct asymptotically valid confidence intervals for the estimator. Most previous studies involved econometrics (e.g., [1–3, 16]).

The single index model is a classical semi-parametric model that is widely used in many fields such as medicine, econometrics, and industry as a reasonable compromise between fully parametric and fully nonparametric modeling. A single index model with a change point parameter has the form

$$y_i = \begin{cases} g(X_i^{\mathrm{T}}\beta_1) + \varepsilon_i, & i = 1, 2, \dots, \lfloor n\tau^* \rfloor, \\ g(X_i^{\mathrm{T}}\beta_2) + \varepsilon_i, & i = \lfloor n\tau^* \rfloor + 1, \dots, n, \end{cases} \tag{1.1}$$

*Corresponding author

where $y_i, i = 1, 2, \ldots$ are the response variables, $X_i, i = 1, 2, \ldots$ are independent and identically distributed (i.i.d.) with a common density $f(X)$, $\varepsilon_i, i = 1, 2, \ldots$ are i.i.d with mean 0, variance σ^2 and $E(\varepsilon_i|X_i) = 0$, $\beta_1 \neq \beta_2$ are unknown index coefficient vectors of length p, g is a smooth unknown link function, and $\tau^* \in [\delta, 1-\delta]$ is an unknown change point where $\delta \in (0, 1/2)$ is a given constant. In order to ensure that β_1 and β_2 are identifiable, we set $\|\beta_1\| = \|\beta_2\| = 1$ where $\|(x_1, \ldots, x_p)^{\mathrm{T}}\| = (\sum_{k=1}^{p} x_k^2)^{1/2}$ and the first non-zero component of β_1 and β_2 are larger than zero.

There are three estimation problems for the single index model with a change point: the estimations of parameters β_1 and β_2, the estimation of the link function g, and the estimation of the change point τ^*. If τ^* is known, then there are direct and indirect methods to estimate the parameters β_1, β_2, and g for the single index model. Indirect methods, such as [7] and [14], use an iterative algorithm to simultaneously estimate parameters and the link function. Direct methods include the slice inverse regression [10] and the average derivative estimation [11, 15].

For the case where τ^* is unknown, Cao et al. [6] has proposed an estimator $\hat{\tau}_n$ based on the average derivative estimations of the parameters β_1 and β_2. In this paper, the consistency of $\hat{\tau}_n$ is discussed. For a consistent estimate $\hat{\tau}_n$, it is well known that the larger the difference $\rho = \beta_1 - \beta_2$ or the sample size n is, the more accurate the change point τ^* is estimated by $\hat{\tau}_n$. Theoretically, for a given small difference ρ, if the sample size n is large enough, then the consistent estimate $\hat{\tau}_n$ can also accurately estimate τ^*. But if $\rho \to 0$ as $n \to \infty$, then $\hat{\tau}_n$ might not accurately estimate τ^*, as stated in [1]. It is interesting and important that we can find a rate of convergence of $\rho \to 0$ such that the asymptotic distributions of $\hat{\tau}_n$ can be obtained in this paper.

Here is the structure of the paper. The main theorems about the consistency and the limiting distribution of the estimator are presented in Section 2. Section 3 contains the simulation results. Some conclusions and discussions are contained in Section 4. Proofs are presented in Appendix.

2 Main result

The density-weighted average derivative vectors are defined by Powell et al. [15] as

$$\beta_j^* = E\bigg(f(X)\frac{\partial g(X^{\mathrm{T}}\beta_j)}{\partial X}\bigg) = E\bigg(f(X)\frac{\partial g(X^{\mathrm{T}}\beta_j)}{\partial X^{\mathrm{T}}\beta_j}\bigg)\beta_j \equiv c_j\beta_j, \quad j = 1, 2.$$

Because $\|\beta_j\| = 1$, it is easy to see that $\beta_j = \beta_j^*/\|\beta_j^*\|, j = 1, 2$. We can use a U-statistic to estimate β_j^* as shown in [15], which is a consistent estimator of $\beta_j^*, j = 1, 2$. The three assumptions of Powell et al. [15] needed for the theoretical results are stated below:

A1. The support Ω of f is a convex (possibly unbounded) subset of $\mathbb{R}^p$ with a nonempty interior Ω_0. The density function f is continuous in the components of x for all $x \in \mathbb{R}^p$, so that $f(x) = 0$ for all $x \in \partial\Omega$, where $\partial\Omega$ denotes the boundary of Ω. Furthermore, f is continuously differentiable in the components of x for all $x \in \Omega_0$ and g is continuously differentiable in the components of x for all $x \in \bar{\Omega}$, where $\bar{\Omega}$ differs from Ω_0 by a set of measure zero. Let $p = (k+4)/2$ if k is even and $p = (k+3)/2$ if k is odd. All partial derivatives of $f(x)$ of order $p+1$ exist. The expectation $E[y(\partial^q f(x)/\partial x_{l_1} \cdots \partial x_{l_q})]$ exists for all $q < p+1$.

A2. The components of the random vector $\partial g(X^{\mathrm{T}}\beta)/\partial X$ and random matrix $\partial f(X)/\partial X[y, X^{\mathrm{T}}]$ have finite second moments. Additionally, $\partial f(X)/\partial X$ and $\partial(g(X^{\mathrm{T}}\beta)f(X))/\partial X$ satisfy the following Lipschitz conditions: for some $m(X)$,

$$\begin{aligned}
&\left\|\frac{\partial f(X+V)}{\partial X} - \frac{\partial f(X)}{\partial X}\right\| < m(X)\|V\|,\\
&\max_{t=1,2}\left\|\frac{\partial f(X+V)g((X+V)^{\mathrm{T}}\beta_t)}{\partial X} - \frac{\partial f(X)g(X^{\mathrm{T}}\beta_t)}{\partial X}\right\| < m(X)\|V\|
\end{aligned}$$

with $E[(1+|y_i|+\|X\|)m(X)]^2 < \infty$. Finally, $v(X) = E(Y^2|X)$ is continuous in X.

A3. The support Ω_K of $K(u)$ is a convex (possibly unbounded) subset of $\mathbb{R}^p$ with a nonempty interior and the origin as an interior point. $K(u)$ is a bounded differentiable function such that $K(u) = 0$ for all $u \in \partial\Omega_K$, where $\partial\Omega_K$ denotes the boundary of Ω_K. $K(u)$ is a symmetric function; $K(u) = K(-u)$ for all $u \in \Omega_K$. All moments of $K(u)$ of order p exist. The kernel function $K(\cdot)$ obeys

$$\int K(u)du = 1, \quad \int u_1^{l_1} \cdots u_k^{l_k} K(u)du = 0 \quad \text{for } 0 < l_1 + \cdots + l_k < p,$$

and

$$\int v_1^{l_1} \cdots v_k^{l_k} K(u)du \neq 0, \quad \text{for } l_1 + \cdots + l_k = p.$$

Remark 2.1. Assumption A1 restricts X to be a continuously distributed random vector and results in the smoothness conditions on f and g. Assumption A2 imposes the bounded moment conditions and dominance conditions. Assumption A3 restricts $K(\cdot)$ to be symmetric and whose "moments" of order $p-1$ or less are zero.

As Lemma 2.1 of [15] shows, under these assumptions,

$$\begin{aligned}\beta_t^* &= E\left(f(X)\frac{\partial g(X^{\mathrm{T}}\beta_t)}{\partial X}\right) = \int_\Omega \frac{\partial g(X^{\mathrm{T}}\beta_t)}{\partial X} f(X)^2 dX \\ &= -2\int_\Omega g(X^{\mathrm{T}}\beta_t)\frac{\partial f(X)}{\partial X} f(X)dX = -2Eg(X^{\mathrm{T}}\beta_t)\frac{\partial f(X)}{\partial X}, \quad t = 1, 2. \end{aligned} \tag{2.1}$$

Because $E(y_i|X_i) = g(X_i^{\mathrm{T}}\beta_1)$ if $i \leqslant \lfloor n\tau^* \rfloor$ and $E(y_i|X_i) = g(X_i^{\mathrm{T}}\beta_2)$ if $i > \lfloor n\tau^* \rfloor$, there are product-moment representations of the density-weighted average derivative:

$$\beta_1^* = -2Ey_i\frac{\partial f(X_i)}{\partial X_i}, \quad i \leqslant \lfloor n\tau^* \rfloor \quad \text{and} \quad \beta_2^* = -2Ey_i\frac{\partial f(X_i)}{\partial X_i}, \quad i > \lfloor n\tau^* \rfloor. \tag{2.2}$$

The estimators of β_1^* and β_2^* are obtained using the U-statistic form, which is similar to that of Powell et al. [15], $\hat{\beta}_1^* = T_{1n}(\lfloor n\tau^* \rfloor)$ and $\hat{\beta}_2^* = T_{2n}(\lfloor n\tau^* \rfloor)$, where

$$\begin{aligned} T_{1n}(k) &= -\frac{2}{k}\sum_{i=1}^{k}\left[\frac{1}{k-1}\sum_{j=1, j\neq i}^{k}\frac{1}{h^{p+1}}K'\left(\frac{X_i - X_j}{h}\right)\right]y_i = \binom{k}{2}^{-1}\sum_{1\leqslant i<j\leqslant k}\psi_{i,j}, \\ T_{2n}(k) &= \binom{n-k}{2}^{-1}\sum_{k+1\leqslant i<j\leqslant n}\psi_{i,j}, \end{aligned}$$

with $\psi_{i,j} = -(\frac{1}{h})^{p+1}K'(\frac{X_i - X_j}{h})(y_i - y_j)$, and $h = h_n$ is a window width that depends on the sample size n. From (3.13) of [15], we have

$$E\|\psi_{i,j}\|^2 = O(h^{-(p+2)}). \tag{2.3}$$

If τ^* is known, then the following results can be obtained from [15]. Define $Z_i = (X_i, y_i)$, $k^* = \lfloor n\tau^* \rfloor$, $g_1(X) = g(X^{\mathrm{T}}\beta_1)$, $g_2(X) = g(X^{\mathrm{T}}\beta_2)$,

$$\begin{aligned} r_1(Z_i) &= f(X_i)\frac{\partial g_1(X_i)}{\partial X_i} - [y_i - g_1(X_i)]\frac{\partial f(X_i)}{\partial X_i}, \quad r_{1n}(Z_i) = E(\psi_{i,j}|Z_i), \quad i \neq j, \ i \leqslant k^*, \ j \leqslant k^*, \\ r_2(Z_j) &= f(X_j)\frac{\partial g_2(X_j)}{\partial X} - [y_j - g_2(X_j)]\frac{\partial f(X_j)}{X}, \quad r_{2n}(Z_j) = E(\psi_{i,j}|Z_j), \quad i \neq j, \ i > k^*, \ j > k^*, \\ r_1^*(Z_i) &= f(X_i)\frac{\partial g_2(X_i)}{\partial X} - [y_i - g_2(X_i)]\frac{\partial f(X_i)}{\partial X_i}, \quad r_{1n}^*(Z_i) = E(\psi_{i,j}|Z_i), \quad i \leqslant k^*, \ j > k^*, \\ r_2^*(Z_j) &= f(X_j)\frac{\partial g_1(X_j)}{\partial X_j} - [y_j - g_1(X_j)]\frac{\partial f(X_j)}{\partial X_j}, \quad r_{2n}^*(Z_j) = E(\psi_{i,j}|Z_j), \quad i \leqslant k^*, \ j > k^*. \end{aligned}$$

In view of (2.1) and (2.2), we can show that $Er_1(Z_i) = \beta_1^*$, $Er_2(Z_j) = \beta_2^*$ and $Er_1^*(Z_i) = Er_2^*(Z_j) = (\beta_1^* + \beta_2^*)/2$. Using (3.15) from [15] and Assumption A2, we have, as $h \to 0$,

$$r_1(Z_i) - r_{1n}(Z_i) = O_p(h), \quad r_2(Z_j) - r_{2n}(Z_j) = O_p(h),$$

$$r_1^*(Z_i) - r_{1n}^*(Z_i) = O_p(h), \quad r_2^*(Z_j) - r_{2n}^*(Z_j) = O_p(h). \tag{2.4}$$

Define

$$\hat{r}_{1k}(Z_i) = \frac{1}{k-1}\sum_{j=1, j\neq i}^{k} \psi_{i,j}\ , \quad \hat{r}_{2k}(Z_j) = \frac{1}{n-k-1}\sum_{i=k+1, i\neq j}^{n} \psi_{i,j},$$
$$\hat{r}_{1k}^*(Z_i) = \frac{1}{n-k}\sum_{j=k+1}^{n} \psi_{i,j}\ , \quad i \leqslant k\ , \quad \hat{r}_{2k}^*(Z_j) = \frac{1}{k}\sum_{i=1}^{k}\psi_{i,j}, \quad j > k,$$
$$\Sigma_{\beta_1^*} = 4E[r_1(Z_i)r_1(Z_i)^{\mathrm{T}}] - 4\beta_1^*(\beta_1^*)^{\mathrm{T}}\ , \quad \Sigma_{\beta_2^*} = 4E[r_2(Z_j)r_2(Z_j)^{\mathrm{T}}] - 4\beta_2^*(\beta_2^*)^{\mathrm{T}},$$
$$\hat{\Sigma}_{\beta_1^*} = 4E[\hat{r}_{1k^*}(Z_i)\hat{r}_{1k^*}(Z_i)^{\mathrm{T}}] - 4\hat{\beta}_1^*(\hat{\beta}_1^*)^{\mathrm{T}} \quad \text{and} \quad \hat{\Sigma}_{\beta_2^*} = 4E[\hat{r}_{2k^*}(Z_j)\hat{r}_{2k^*}(Z_j)^{\mathrm{T}}] - 4\hat{\beta}_2^*(\hat{\beta}_2^*)^{\mathrm{T}}.$$

Using Theorems 3.3 and 3.4 from [15], under Assumptions A1–A3, $nh^{p+2} \to \infty$ and $nh^{2p} \to 0$ as $n \to \infty$, we have

$$\sqrt{n}(\hat{\beta}_t^* - \beta_t^*) \to_d N_p(0, \Sigma_{\beta_t^*}) \quad \text{and} \quad \hat{\Sigma}_{\beta_t^*} \to_p \Sigma_{\beta_t^*}, \quad \text{for } t = 1, 2. \tag{2.5}$$

Defining

$$E\bigg[-\bigg(\frac{1}{h}\bigg)^{p+1} K'\bigg(\frac{X_i - X_j}{h}\bigg)y_i\bigg] = \begin{cases} v_{1n}, & i \leqslant \lfloor n\tau^* \rfloor, \\ v_{2n}, & i > \lfloor n\tau^* \rfloor, \end{cases}$$

$\eta_n = \|\beta_1^* - \beta_2^*\|$ and $\rho_n = \|v_{1n} - v_{2n}\|$, by Theorem 3.2 from [15], it follows that

$$2v_{tn} - \beta_t^* = o(n^{-1/2}), \quad t = 1, 2 \tag{2.6}$$

and

$$2\rho_n - \eta_n = o(n^{-1/2}). \tag{2.7}$$

If τ^* is unknown, let $\hat{k}_n = \arg\max_{\delta n < k < (1-\delta)n} \|Q(k)\|$, where $Q(k) = T_{1n}(k) - T_{2n}(k)$. A natural estimator of τ^* is defined as $\hat{\tau}_n = \hat{k}_n/n$, as in [6]. By using the results (2.3)–(2.7), the following theorem is established.

Theorem 2.2. *Given Assumptions* A1–A3, *if* $\sqrt{n}\eta_n \to \infty$, *h satisfies* $nh^{p+2} \to \infty$ *and* $nh^{2p} \to 0$ *as* $n \to \infty$, *then we have* (1) $\hat{\tau}_n \to_p \tau^*$; (2) $|\hat{\tau}_n - \tau^*| = O_p(n^{-1}\eta_n^{-2})$; (3) $\omega_n(\hat{k}_n - k^*) \to_d \arg\max_s G(s)$, *where* $\omega_n = (\tau^*)^2\eta_n^2/\lambda_2^2$,

$$G(s) = \begin{cases} \dfrac{\lambda_1}{\lambda_2}W_1(s) - \dfrac{1-\tau^*}{\tau^*}|s|/2, & s > 0, \\ W_2(-s) - |s|/2, & s \leqslant 0 \end{cases}$$

and

$$\lambda_1 = \left(E\left\|\frac{(\beta_1^* - \beta_2^*)^{\mathrm{T}}}{\|\beta_1^* - \beta_2^*\|}\left[(1-\tau^*)\left(r_2^*(Z_j) - \frac{\beta_1^* + \beta_2^*}{2}\right) + \tau^*(r_2(Z_j) - \beta_2^*)\right]\right\|^2\right)^{1/2},$$
$$\lambda_2 = \left(E\left\|\frac{(\beta_1^* - \beta_2^*)^{\mathrm{T}}}{\|\beta_1^* - \beta_2^*\|}\left[\tau^*\left(r_1^*(Z_i) - \frac{\beta_1^* + \beta_2^*}{2}\right) + (1-\tau^*)(r_1(Z_i) - \beta_1^*)\right]\right\|^2\right)^{1/2}.$$

Remark 2.3. Here we give the estimators of λ_1 and λ_2. Because $(\hat{k}_n - k^*)/n \to_p 0$, by Makov's inequality, (2.3) and (2.4), we can obtain that $\hat{r}_{1k^*}^*(Z_i) - r_1^*(Z_i) \to_p 0$, $\hat{r}_{2k^*}^*(Z_j) - r_2^*(Z_j) \to_p 0$, $\hat{r}_{1k^*}(Z_i) - r_1(Z_i) \to_p 0$ and $\hat{r}_{2k^*}(Z_j) - r_2(Z_j) \to_p 0$. By (2.6) and the law of large numbers, we have $\frac{1}{k^*}\sum_{i=1}^{k^*}\hat{r}_{1k^*}(Z_i) \to_p \beta_1^*$, $\frac{1}{n-k^*}\sum_{j=k^*+1}^{n}\hat{r}_{2k^*}(Z_j) \to_p \beta_2^*$, $\frac{1}{k^*}\sum_{i=1}^{k^*}\hat{r}_{1k^*}^*(Z_i) \to_p (\beta_1^* + \beta_2^*)/2$ and $\frac{1}{n-k^*}\sum_{j=k^*+1}^{n}\hat{r}_{2k^*}^*(Z_j) \to_p (\beta_1^* + \beta_2^*)/2$. Denote

$$\hat{\delta}_n = \frac{1}{k^*}\sum_{i=1}^{k^*}\hat{r}_{1k^*}(Z_i) - \frac{1}{n-k^*}\sum_{j=k^*+1}^{n}\hat{r}_{2k^*}(Z_j),$$

$$\hat{r}_{1i} = \left(1 - \frac{\hat{k}_n}{n}\right)\left[\hat{r}_{1\hat{k}_n}(Z_i) - \frac{1}{\hat{k}_n}\sum_{i=1}^{\hat{k}_n}\hat{r}_{1\hat{k}_n}(Z_i)\right] + \frac{\hat{k}_n}{n}\left[\hat{r}^*_{1\hat{k}_n}(Z_i) - \frac{1}{\hat{k}_n}\sum_{i=1}^{\hat{k}_n}\hat{r}^*_{1\hat{k}_n}(Z_i)\right],$$

$$\hat{r}_{2j} = \left(1 - \frac{\hat{k}_n}{n}\right)\left[\hat{r}^*_{2\hat{k}_n}(Z_j) - \frac{1}{n-\hat{k}_n}\sum_{j=\hat{k}_n+1}^{n}\hat{r}^*_{2\hat{k}_n}(Z_j)\right] + \frac{\hat{k}_n}{n}\left[\hat{r}_{2\hat{k}_n}(Z_j) - \frac{1}{n-\hat{k}_n}\sum_{j=\hat{k}_n+1}^{n}\hat{r}_{2\hat{k}_n}(Z_j)\right].$$

Therefore, we can define $\hat{\eta}_n = \|\hat{\delta}_n\|$,

$$\hat{\lambda}_1 = \left(\frac{1}{n-\hat{k}_n}\sum_{j=\hat{k}_n+1}^{n}\frac{\hat{\delta}_n^{\mathrm{T}}\hat{r}_{2j}\hat{r}_{2j}^{\mathrm{T}}\hat{\delta}_n}{\hat{\eta}_n^2}\right)^{1/2} \quad \text{and} \quad \hat{\lambda}_2 = \left(\frac{1}{\hat{k}_n}\sum_{i=1}^{\hat{k}_n}\frac{\hat{\delta}_n^{\mathrm{T}}\hat{r}_{1i}\hat{r}_{1i}^{\mathrm{T}}\hat{\delta}_n}{\hat{\eta}_n^2}\right)^{1/2}$$

as the estimators of η_n, λ_1 and λ_2, respectively.

Remark 2.4. Using the result of Appendix B in [1], we can get the cumulative distribution function $F(s)$ of $\arg\max_{s>0} G(s)$ as follows:

$$\begin{aligned} F(s) = & -(2\pi)^{-1/2}|s|^{1/2}\exp(-8^{-1}|s|) - \frac{\phi(\phi+2\xi)}{\xi(\phi+\xi)}\exp\left(\frac{\xi(\phi+\xi)}{2\phi^2}|s|\right)\Phi\left(-\left(\frac{\phi+2\xi}{2\phi}\right)|s|^{1/2}\right) \\ & + \left(\frac{(\phi+2\xi)^2}{\xi(\phi+\xi)} - 2 + 2^{-1}|s|\right)\Phi(-2^{-1}|s|^{1/2}), \quad s \leqslant 0, \end{aligned}$$

and

$$\begin{aligned} F(s) = 1 & + \frac{\xi s^{1/2}}{\phi^{1/2}(2\pi)^{1/2}}\exp\left(-\frac{\xi^2}{8\phi}s\right) + \frac{\xi(2\phi+\xi)}{(\phi+\xi)\phi}\exp\left(\frac{(\phi+\xi)s}{2}\right)\Phi\left(-\frac{(2\phi+\xi)s^{1/2}}{2\phi^{1/2}}\right) \\ & - \left(\frac{(2\phi+\xi)^2}{(\phi+\xi)\phi} - 2 + \frac{\xi^2 s}{2\phi}\right)\Phi\left(-\frac{\xi s^{1/2}}{2\phi^{1/2}}\right), \quad s > 0, \end{aligned}$$

where $\phi = \lambda_1^2/\lambda_2^2$ and $\xi = (1-\tau^*)/\tau^*$.

Remark 2.5. Because $\|Q(k)\|$ is the estimate of η_n, it is natural to give the condition of η_n to replay $\|\beta_1 - \beta_2\|$ in the theorem which can be directly applied in the proof. The relationship between η_n and $\|\beta_1 - \beta_2\|$ is not very clear for $\eta_n = \|c_1\beta_1 - c_2\beta_2\|$, where $c_i, i = 1, 2$ are based on the unknown link function g and the explanatory variable X. We can easily see that if $\beta_2 - \beta_1 = 0$ then $\eta_n = 0$, and we can deduce $\eta_n \neq 0$ when $\beta_1 - \beta_2 \neq 0$ if $g'(t) + g'(-t) > 0$ for all t or $g'(t) + g'(-t) < 0$ for all t. Otherwise, we assume $\beta_1 \neq \beta_2$ but $c_1\beta_1 = c_2\beta_2$, and then only one case satisfies it that is $\beta_1 = -\beta_2$ and $c_1 = -c_2$. But, for $c_1 + c_2 = E(f(X)(g'(X^{\mathrm{T}}\beta_1)) + g'(-X^{\mathrm{T}}\beta_1)) \neq 0$ under the condition of $g'(t) + g'(-t) > 0$ for all t or $g'(t) + g'(-t) < 0$ for all t, the case of $\beta_1 \neq \beta_2$ and $\eta_n = 0$ does not exist.

Since we apply $\|Q(k)\|$ as the criterion to estimate the change point, Theorem 2.1 might also work in the model $y_i = g_1(X_i^{\mathrm{T}}\beta_1) + \epsilon_i, i \leqslant \lfloor n\tau^* \rfloor$ while $y_i = g_2(X_i^{\mathrm{T}}\beta_2) + \epsilon_i, i > \lfloor n\tau^* \rfloor$.

3 Simulation

In this section, we investigate the performance of our estimator of the parametric change point in a single index model. We consider the following model:

$$y_i = \begin{cases} g(X_i^{\mathrm{T}}\beta_1) + \varepsilon_i, & 1 \leqslant i < k^*, \\ g(X_i^{\mathrm{T}}\beta_2) + \varepsilon_i, & k^* \leqslant i \leqslant n. \end{cases}$$

For convenience, we assume that $\varepsilon_i, i = 1, 2, \ldots, n$ are independent and identically distributed (i.i.d.) with a standard normal distribution $N(0,1)$ and $g(v) = 10\exp(v)$. The normal density kernel function is used, and the bandwidth is set to be $h = n^{-1/5}$. Let $X = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)^{\mathrm{T}}$, where $\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,n})^{\mathrm{T}}$.

Consider the following four types of parameters:

Table 1 The true values and estimations of the parameters of $\hat{F}(\hat{k}_n)$

Models	True value				Mean of estimation value			
	k^*	η_n^2	λ_1	λ_2	$\hat{k}_n$	$\hat{\eta}_n^2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$
M1	100	71.57	66.37	18.02	109.34	79.45	64.27	23.80
M2	150	61.21	11.71	11.49	150.02	61.55	11.72	11.50
M3	100	61.00	13.24	10.85	95.01	62.12	13.59	11.34
M4	200	61.41	10.74	13.37	204.72	62.61	11.21	13.72

Table 2 The qunatiles of empirical distribution $F_n(\hat{k}_n)$ and theoretical distribution $\hat{F}(\hat{k}_n)$

Models	Qunatiles	2.5%	5%	10%	50%	90%	95%	97.5%
M1	$F_n(\hat{k}_n)$	91.97	95.00	98.00	103.00	139.00	146.00	148.00
	$\hat{F}(\hat{k}_n)$	87.56	92.25	96.35	102.98	131.14	148.41	167.70
M2	$F_n(\hat{k}_n)$	143.00	146.00	147.00	150.00	153.00	155.00	157.02
	$\hat{F}(\hat{k}_n)$	141.78	144.27	146.50	150.00	153.60	155.89	158.45
M3	$F_n(\hat{k}_n)$	57.00	71.95	85.00	99.00	100.00	101.00	102.00
	$\hat{F}(\hat{k}_n)$	95.48	96.79	97.96	99.87	100.38	100.68	101.01
M4	$F_n(\hat{k}_n)$	198.00	199.00	200.00	201.00	212.00	230.00	244.00
	$\hat{F}(\hat{k}_n)$	178.05	185.24	191.63	202.58	242.52	267.21	294.80

M1. Elements of $\boldsymbol{x}_1, \boldsymbol{x}_2$ and $\boldsymbol{x}_3$ are i.i.d. with $N(0, 0.5), N(1, 0.5)$ and $N(-1, 0.5)$, respectively, $\beta_1 = (1/\sqrt{2}, -1/\sqrt{2}, 0)$, $\beta_2 = (1/\sqrt{2}, 0, -1/\sqrt{2})$, $n = 200$ and $k^* = 100$.

M2, M3, M4. Elements of $\boldsymbol{x}_1, \boldsymbol{x}_2$ and $\boldsymbol{x}_3$ are i.i.d. with $N(0, 0.5), N(2, 0.5)$ and $N(-2, 0.5)$, respectively, $\beta_1 = (1/\sqrt{3}, -1/\sqrt{3}, -1/\sqrt{3})$, $\beta_2 = (1/\sqrt{3}, 1/, 1/\sqrt{3}, 1/\sqrt{3})$ and $n = 300$. Change points are set to be $k^* = 150, 100$ and 200 for M2, M3 and M4, respectively.

Each model is repeated 1000 times, and the results are shown in Tables 1 and 2. Table 1 displays the estimators of parameters of the distribution $F(k^*)$ as shown in Theorem 2.2. The true values η_n, λ_1 and λ_2 are obtained using the numerical approximate operation. The estimate values $\hat{\eta}_n$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are obtained by Remark 2.3. Table 1 shows that the differences between the true and estimated values are not large. Notice that η_n of M1 is larger than that of M2, but $\eta_n\sqrt{n}$ of M1 is smaller than that of M2, which correspond to the condition of Theorem 2.2 $\eta_n \to 0, \eta_n\sqrt{n} \to \infty$. For $\tau^* = k^*/n = 0.5$ in both M1 and M2, comparing the results of M1 and M2 in Tables 1 and 2, we find estimated values in M2 is more accurate.

Using the estimated values in Table 1, we obtain the quantiles of the distribution $\hat{F}(\hat{k}_n)$ as shown in Remark 2.4, which is displayed in Table 2. In Table 2, $F_n(\hat{k}_n)$ is the empirical distribution of $\hat{k}_n$ based on the 1000 estimated value. From Table 2, comparing with $F_n(\hat{k}_n)$, we can see that $\hat{F}(\hat{k}_n)$ is better when the change point is in the middle of data, the estimate too centralized when the change point is in the front of data (namely, $\tau^* < 0.5$), and the estimate is too decentralized when the change point is in the back of data (that is, $\tau^* > 0.5$).

4 Conclusion

In this paper, we proved the consistency, obtained the convergence rate, and derived the distribution function of the estimate of a change point in single index models. However, because there are many parameters in the distribution function, a description of how to find consistent estimators of these parameters is not included in this paper and will be the subject of a future paper.

Acknowledgements This work was supported by National Natural Science Foundation for Young Scientists of China (Grant Nos. 11101397, 11201108), the Humanities and Social Sciences Project from Ministry of Education of China (Grant No. 12YJC910007), Anhui Provincial Natural Science Foundation (Grant No. 1208085QA12), and the National Statistical Research Plan Project (Grant No. 2012LZ009).

References

1 Bai J. Estimation of a change point in multiple regression models. Rev Econ Stat, 1997, 79: 551–63

2 Bai J. Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices. Ann Econ Finance, 2000, 1: 303–339

3 Bai J, Lumsdaine R L, Stock J H. Testing for and dating common breaks in multivariate time series. Rev Econ Stud, 1998, 65: 395–432

4 Berk R H. Limiting behavior of posterior distributions when the model is incorrect. Ann Math Stat, 1966, 37: 51–58

5 Bhatti M I, Wang J. On testing for a change-point in variance of normal distribution. Biometrical J, 2000, 42: 1021–1032

6 Cao G Q, Wang Z F, Wu Y H, et al. Inference of change-point in single index models. Sci China Ser A, 2008, 51: 1855–1870

7 Carrolla R J, Fan J Q, Gijbels I, et al. Generalized partially linear single-index models. J Amer Stat Assoc, 1997, 92: 477–489

8 Chen X R. Inference in a simple change point model. Sci Sin Ser A, 1988, 31: 654–667

9 Csörgö M, Horváth L. Limit Theorems in Change-point Analysis. New York: Wiley, 1997

10 Duan N, Li K C. Slicing regression: A link-free regression method. Ann Stat,1991, 19: 505–530

11 Hristache M, Juditsky A, Spokoiny V. Direct estimation of the index coefficient in a single-index model. Ann Stat, 2001, 29: 593–623

12 Jin B S, Huang L M N, Miao B Q. Testing for variance changes in autoregressive models with unknown order. J Appl Stat, 2011, 38: 927–936

13 Krishnaiah P R, Miao B Q. Review about estimation of change point. Handbook Stat Quality Control Reliability, 1988, 7: 375–402

14 Peng H, Huang T. Penalized least squares for single index models. J Stat Plan Inference, 2011, 141: 1362–1379

15 Powell J C, Stock J H, Stoker T M. Semiparametric estimation of index coefficients. Econometrica, 1989, 57: 1403–1430

16 Qu Z, Perron P. Estimating and testing structural changes in multivariate regressions. Econometrica, 2007, 75: 459–502

Appendix

Lemma 4.1 (Rényi-Hájek-Chow inequality). *Denote $\{X_n, n \geqslant 1\}$ to be a martingale difference sequence and $S_n = \sum_{j=m}^n X_j$. Let $\sigma_n^2 = E\|X_n\|^2$ and $c_1 \geqslant c_2 \geqslant \cdots \geqslant c_n > 0$, then for any $x > 0$ and any positive integer m,*

$$P\Big(\max_{m\leqslant j\leqslant n} c_j\|S_j\| \geqslant x\Big) \leqslant \frac{1}{x^2}\sum_{j=m}^n c_j^2\sigma_j^2.$$

A.1 The proof of Theorem 2.2(1)

Proof. Since $\|Q(\hat{k}_n)\| \leqslant \|Q(\hat{k}_n)-EQ(\hat{k}_n)\|+\|EQ(\hat{k}_n)\|$ and $\|EQ(k^*)\| \leqslant \|Q(k^*)-EQ(k^*)\|+\|Q(k^*)\|$, noticing that $\|Q(\hat{k}_n)\| > \|Q(k^*)\|$, by the triangle inequality, it is easy to show that

$$\|EQ(k^*)\| - \|EQ(\hat{k}_n)\| \leqslant 2\max_{\delta n<k<(1-\delta)n}\|Q(k)-EQ(k)\|. \tag{A.1}$$

If $k > k^*$, by the simple decomposition for $Q(k)$, we have

$$Q(k) = \binom{k}{2}^{-1}\Bigg(\sum_{1\leqslant i<j\leqslant k^*}\psi_{i,j} + \sum_{k^*<i<j\leqslant k}\psi_{i,j} + \sum_{i=1}^{k^*}\sum_{j=k^*+1}^{k}\psi_{i,j}\Bigg) - \binom{n-k}{2}^{-1}\sum_{k<i<j\leqslant n}\psi_{i,j},$$

thus, we obtain

$$EQ(k) = \binom{k}{2}^{-1}\Bigg(\sum_{1\leqslant i<j\leqslant k^*}2v_{1n} + \sum_{k^*<i<j\leqslant k}2v_{2n} + \sum_{i=1}^{k^*}\sum_{j=k^*+1}^{k}(v_{1n}+v_{2n})\Bigg) - \binom{n-k}{2}^{-1}\sum_{k<i<j\leqslant n}2v_{2n}$$

$$= \frac{2(v_{1n} - v_{2n})k^*}{k}. \tag{A.2}$$

For $k \leqslant k^*$, we have

$$EQ(k) = \frac{2(v_{1n} - v_{2n})(n - k^*)}{n - k}.$$

This leads to

$$\|EQ(k^*)\| - \|EQ(\hat{k}_n)\| = \begin{cases} 2\rho_n \dfrac{\hat{k}_n - k^*}{\hat{k}_n} = 2\rho_n \dfrac{\hat{\tau}_n - \tau^*}{\hat{\tau}_n}, & \hat{k}_n > k^*, \\ 2\rho_n \dfrac{k^* - \hat{k}_n}{n - \hat{k}_n} = 2\rho_n \dfrac{\tau^* - \hat{\tau}_n}{1 - \hat{\tau}_n}, & \hat{k}_n \leqslant k^*. \end{cases} \tag{A.3}$$

Since $\delta < \hat{\tau}_n < 1 - \delta$, we obtain

$$\frac{2\rho_n|\tau^* - \hat{\tau}_n|}{1 - \delta} \leqslant \|EQ(k^*)\| - \|EQ(\hat{k}_n)\| \leqslant 2 \max_{\delta n < k < (1-\delta)n} \|Q(k) - EQ(k)\|.$$

Hence, we have

$$\begin{aligned} P\left(\frac{\rho_n|\tau^* - \hat{\tau}_n|}{1 - \delta} > \varepsilon\right) &\leqslant P\Big(\max_{\delta n < k < (1-\delta)n} \|Q(k) - EQ(k)\| > \varepsilon\Big) \\ &\leqslant P\Big(\max_{\delta n < k \leqslant k^*} \|Q(k) - EQ(k)\| > \varepsilon\Big) + P\Big(\max_{k^* < k < (1-\delta)n} \|Q(k) - EQ(k)\| > \varepsilon\Big). \end{aligned}$$

Because of the symmetry, we only show $P(\max_{k^* < k < (1-\delta)n} \|Q(k) - EQ(k)\| > \varepsilon) \to 0$ as $n \to \infty$. The remaining part $P(\max_{\delta n < k \leqslant k^*} \|Q(k) - EQ(k)\| > \varepsilon)$ is analogous and thus is omitted.

Now, we consider the $P(\max_{k^* < k < (1-\delta)n} \|Q(k) - EQ(k)\| > \varepsilon)$. By the decomposition for $Q(k) - EQ(k)$, we have, for $k^* < k < (1 - \delta)n$,

$$\begin{aligned} Q(k) - EQ(k) = &\binom{k}{2}^{-1} \sum_{1 \leqslant i < j \leqslant k^*} (\psi_{i,j} - E\psi_{i,j}) + \binom{k}{2}^{-1} \sum_{k^* < i < j \leqslant k} (\psi_{i,j} - E\psi_{i,j}) \\ &+ \binom{k}{2}^{-1} \sum_{i=1}^{k^*} \sum_{j=k^*+1}^{k} (\psi_{i,j} - E\psi_{i,j}) - \binom{n-k}{2}^{-1} \sum_{k < i < j \leqslant n} (\psi_{i,j} - E\psi_{i,j}) \\ \equiv &\, I + II + III - IV. \end{aligned} \tag{A.4}$$

Note that $n^{-1}h^{-(p+2)} \to 0$, and by (2.3), it follows that

$$\max_{k^* < k \leqslant (1-\delta)n} \|I\| \leqslant \binom{k^*}{2}^{-1} \sum_{1 \leqslant i < j \leqslant k^*} (\psi_{i,j} - E\psi_{i,j}) \to_p 0. \tag{A.5}$$

For the second part II in above $Q(k) - EQ(k)$, denote $U_{k-k^*} = \binom{k-k^*}{2}^{-1} \sum_{k^* < i < j \leqslant k} (\psi_{i,j} - E\psi_{i,j})$ and $V_{k-k^*} = \frac{4}{k-k^*} \sum_{i=k^*+1}^{k} (r_{2n}(Z_i) - E\psi_{i,j})$. Then $U_{k-k^*} - V_{k-k^*} = \binom{k-k^*}{2}^{-1} \sum_{k^* < i < j \leqslant k} \phi_{i,j}$, where $\phi_{i,j} = \psi_{i,j} - r_{2n}(Z_i) - r_{2n}(Z_j) + E\psi_{i,j}$. By (3.15) of [15] and Assumption A2, we have $E\|r_{2n}(Z_i) - E\psi_{i,j}\|^2 < \infty$. Thus, we can obtain

$$\begin{aligned} E\|\phi_{i,j}\|^2 &= E\|\psi\|^2 + E\|r_{2n}(Z_i) - E\psi_{i,j}\|^2 + E\|r_{2n}(Z_j) - E\psi_{i,j}\|^2 \\ &\quad - 4E[(\psi)^{\mathrm{T}}(r_{2n}(Z_i) - E\psi_{i,j})] \\ &= E\|\psi\|^2 - 2E\|r_{2n}(Z_i) - E\psi_{i,j}\|^2 = O(h^{-(p+2)}) \end{aligned} \tag{A.6}$$

and

$$\begin{aligned} E[\phi_{i,j}^{\mathrm{T}}\phi_{i,j'}] = &\, E[(\psi_{i,j} - E\psi_{i,j})^{\mathrm{T}}(\psi_{i,j'} - E\psi_{i,j'})] - E[(r_{2n}(Z_i) - E\psi_{i,j})^{\mathrm{T}}(\psi_{i,j'} - E\psi_{i,j'})] \\ &- E[(\psi_{i,j} - E\psi_{i,j})^{\mathrm{T}}(r_{2n}(Z_i) - E\psi_{i,j})] + E\|r_{2n}(Z_i) - E\psi_{i,j}\|^2 = 0. \end{aligned} \tag{A.7}$$

Similarly, we have $E[\phi_{i,j}^{\mathrm{T}}\phi_{i',j'}] = E[\phi_{i',j}^{\mathrm{T}}\phi_{i,j}] = E[\phi_{i,j}^{\mathrm{T}}\phi_{i,j'}] = 0$, where i, i', j, j' are four unequal numbers each other.

Let $\Phi_k = \sum_{k^*<i<j\leqslant k}\phi_{i,j}$. Since $E(\Phi_{k+1}|Z_i, i = k^*+1,\dots,k) = \Phi_k$, $\{\Phi_k, k = k^*+1,\dots,n\}$ is a discrete-time martingale process. By Lemma 4.1 and $\binom{k}{2}^{-1}\binom{k-k^*}{2}\frac{1}{k-k^*} = \frac{(k-k^*-1)}{k(k-1)} < \frac{1}{k}$, we have

$$\begin{aligned}
P\Big(\max_{k^*<k\leqslant(1-\delta)n}\|II\| \geqslant \varepsilon\Big) &\leqslant P\Big(\max_{k^*<k\leqslant n}\Big\|\frac{2}{k}\sum_{i=k^*+1}^{k}(r_{2n}(Z_i) - E(\psi_{i,j}))\Big\| \geqslant \varepsilon/2\Big)\\
&\quad + P\Big(\max_{k^*<k\leqslant n}\binom{k}{2}^{-1}\Big\|\sum_{k^*<i<j\leqslant k}\phi_{i,j}\Big\| \geqslant \varepsilon/2\Big)\\
&\leqslant \frac{16E\|(r_{2n}(Z_i) - E(\psi_{i,j})\|^2}{\varepsilon^2}\sum_{k=k^*+1}^{n}\frac{1}{k^2}\\
&\quad + \sum_{k=k^*+2}^{n}\frac{16(k-1-k^*)E\|\phi_{i,j}\|^2}{\varepsilon^2k^2(k-1)^2}\\
&= O(k^{*-1}\varepsilon^{-2}) + O(k^{*-2}h^{-(p+2)}\varepsilon^{-2}) = O(n^{-1}\varepsilon^{-2}).
\end{aligned}\tag{A.8}$$

If $1 \leqslant i \leqslant k^*$ and $k^* < j \leqslant k$, denote $U_{k^*,k-k^*} = \frac{1}{k^*(k-k^*)}\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{k}(\psi_{i,j} - E\psi_{i,j})$ and

$$\begin{aligned}
V_{k^*,k-k^*} &= \sum_{i=1}^{k^*}E(U_{k^*,k-k^*}|Z_i) + \sum_{j=k^*+1}^{k}E(U_{k^*,k-k^*}|Z_j)\\
&\equiv \frac{1}{k^*}\sum_{i=1}^{k^*}(r_{1n}^*(Z_i) - v_{1n} - v_{2n}) + \frac{1}{k-k^*}\sum_{j=k^*+1}^{k}(r_{2n}^*(Z_j) - v_{1n} - v_{2n}).
\end{aligned}$$

Notice that $U_{k^*,k-k^*} - V_{k^*,k-k^*} = \frac{1}{k^*(k-k^*)}\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{k}\tilde\phi_{i,j}$, where $\tilde\phi_{i,j} = \psi_{i,j} - r_{1n}^*(Z_i) - r_{2n}^*(Z_j) + v_{1n} + v_{2n}$. By Assumption A2, we have $E\|r_{1n}^*(Z_i)\|^2 < \infty$ and $E\|r_{2n}^*(Z_j)\|^2 < \infty$. Similarly to the proofs of (A.6) and (A.7), we have $E\|\tilde\phi_{i,j}\|^2 = O(h^{-(p+2)})$ and $E[\tilde\phi_{i,j}^{\mathrm{T}}\tilde\phi_{i',j'}] = E[\tilde\phi_{i',j}^{\mathrm{T}}\tilde\phi_{i,j}] = E[\tilde\phi_{i,j}^{\mathrm{T}}\tilde\phi_{i,j'}] = 0$, where i, i', j, j' are four unequal numbers. By Chebyshev inequality and Lemma 4.1, we have

$$\begin{aligned}
&P\Big(\max_{k^*<k\leqslant(1-\delta)n}\|III\| \geqslant 3\varepsilon\Big)\\
&\leqslant P\Big(\max_{k^*<k\leqslant n}\binom{k}{2}^{-1}k^*(k-k^*)\frac{1}{k^*}\Big\|\sum_{i=1}^{k^*}(r_{1n}^*(Z_i) - v_{1n} - v_{2n})\Big\| > \varepsilon\Big)\\
&\quad + P\Big(\max_{k^*<k\leqslant n}\binom{k}{2}^{-1}k^*(k-k^*)\frac{1}{k-k^*}\Big\|\sum_{j=k^*+1}^{k}(r_{2n}^*(Z_j) - v_{1n} - v_{2n})\Big\| > \varepsilon\Big)\\
&\quad + P\Big(\max_{k^*<k\leqslant n}\binom{k}{2}^{-1}k^*(k-k^*)\frac{1}{k^*(k-k^*)}\Big\|\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{k}\phi_{i,j}\Big\| > \varepsilon\Big)\\
&\leqslant P\Big(\frac{1}{k^*}\Big\|\sum_{i=1}^{k^*}[r_{1n}^*(Z_i) - v_{1n} - v_{2n}]\Big\| > \varepsilon\Big)\\
&\quad + P\Big(\max_{k^*<k\leqslant(1-\delta)n}\frac{2k^*}{k(k-1)}\Big\|\sum_{j=k^*+1}^{k}[r_{2n}^*(Z_j) - v_{1n} - v_{2n}]\Big\| > \varepsilon\Big)\\
&\quad + P\Big(\max_{k^*<k\leqslant(1-\delta)n}\frac{2}{k(k-1)}\Big\|\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{k}\phi_{i,j}\Big\| > \varepsilon\Big)\\
&\leqslant \frac{E\|r_{1n}^*(Z_i)\|^2}{\varepsilon^2k^*} + \frac{E\|r_{2n}^*(Z_j)\|^2}{\varepsilon^2}\sum_{k=k^*+1}^{n}\Big(\frac{2k^*}{k(k-1)}\Big)^2 + \frac{E\|\phi_{i,j}\|^2}{\varepsilon^2}\sum_{k=k^*+1}^{n}\frac{4(k^*-1)}{k^2(k-1^2)}
\end{aligned}$$

$$= O(n^{-1}\varepsilon^{-2}) + O(n^{-1}\varepsilon^{-2}) + O(k^{*-2}h^{-(p+2)}\varepsilon^{-2}) = O(n^{-1}\varepsilon^{-2}). \tag{A.9}$$

As IV is a U-statistic which is a reversed martingale (see [4]), by the basic inequality of reversed martingale, we have

$$P\Big(\max_{k^*<k\leqslant(1-\delta)n} \|IV\| \geqslant \varepsilon\Big) \leqslant \frac{E\|\binom{(1-\delta)n}{2}^{-1}\sum_{(1-\delta n<i<j\leqslant n)}(\psi_{i,j}-E\psi_{i,j})\|^2}{\varepsilon^2} = O(n^{-1}\varepsilon^{-2}). \tag{A.10}$$

Thus, combining (A.4), (A.5) and (A.8)–(A.10), there exits a constant c such that

$$P\bigg(\frac{\rho_n|\tau^*-\hat{\tau}_n|}{1-\delta} > \varepsilon\bigg) \leqslant P\Big(\max_{\delta n<k<(1-\delta)n} \|Q(k)-EQ(k)\| > \varepsilon\Big) \leqslant \frac{c}{n\varepsilon^2}.$$

By (2.7), $n\eta_n^2 \to \infty$ implies $n\rho_n^2 \to \infty$. Therefore, we have $P(|\tau^*-\hat{\tau}_n| > \varepsilon) \leqslant \frac{(1-\delta)^2 c}{n\varepsilon^2\rho_n^2} \to 0$ as $n \to \infty$. The proof is completed. □

A.2 The proof of Theorem 2.2(2)

Proof. Noticing that $\|Q(\hat{k}_n)\|^2 \geqslant \|Q(k^*)\|^2$ and

$$\begin{aligned}\|Q(\hat{k}_n)\|^2 - \|Q(k^*)\|^2 &= \|Q(\hat{k}_n)-EQ(\hat{k}_n)\|^2 - \|Q(k^*)-EQ(k^*)\|^2 + \|EQ(\hat{k}_n)\|^2 - \|EQ(k^*)\|^2\\ &\quad + 2(Q(\hat{k}_n)-EQ(\hat{k}_n))^{\mathrm{T}}EQ(\hat{k}_n) - 2(Q(k^*)-EQ(k^*))^{\mathrm{T}}EQ(k^*),\end{aligned}$$

we have

$$\begin{aligned}\|EQ(k^*)\|^2 - \|EQ(\hat{k}_n)\|^2 &\leqslant \|Q(\hat{k}_n)-EQ(\hat{k}_n)\|^2 - \|Q(k^*)-EQ(k^*)\|^2\\ &\quad + 2(Q(\hat{k}_n)-EQ(\hat{k}_n))^{\mathrm{T}}E(Q(\hat{k}_n)) - 2(Q(k^*)-EQ(k^*))^{\mathrm{T}}EQ(k^*)\\ &= \|Q(\hat{k}_n)-EQ(\hat{k}_n)-Q(k^*)+EQ(k^*)\|^2\\ &\quad + 2(Q(k^*)-EQ(k^*))^{\mathrm{T}}(Q(\hat{k}_n)-EQ(\hat{k}_n)-Q(k^*)+EQ(k^*))\\ &\quad + 2(Q(\hat{k}_n)-EQ(\hat{k}_n))^{\mathrm{T}}(EQ(\hat{k}_n)-EQ(k^*))\\ &\quad + 2(Q(\hat{k}_n)-EQ(\hat{k}_n)-Q(k^*)+EQ(k^*))^{\mathrm{T}}EQ(k^*) \equiv A.\end{aligned}$$

By (A.2), it follows that

$$\|EQ(k^*)\|^2 - \|EQ(\hat{k}_n)\|^2 = \begin{cases} 4\rho_n^2(\hat{\tau}_n-\tau^*)(\hat{\tau}_n+\tau^*)/(\hat{\tau}_n)^2, & \hat{k}_n > k^*,\\ 4\rho_n^2(\tau^*-\hat{\tau}_n)(2-\hat{\tau}_n-\tau^*)/(1-\hat{\tau}_n)^2, & \hat{k}_n \leqslant k^*,\end{cases}$$

therefore, we have $\|EQ(k^*)\|^2 - \|EQ(\hat{k}_n)\|^2 \geqslant 8\rho_n^2|\hat{\tau}_n-\tau^*|\delta/(1-\delta)^2$.

Now we consider the case $\hat{k}_n > k^*$. By the simple decomposition, we have

$$\begin{aligned}&Q(\hat{k}_n)-EQ(\hat{k}_n)-Q(k^*)+EQ(k^*)\\ &\quad= \left(\binom{\hat{k}_n}{2}^{-1} - \binom{k^*}{2}^{-1}\right)\sum_{1\leqslant i<j\leqslant k^*}(\psi_{i,j}-E\psi_{i,j}) + \left(\binom{\hat{k}_n}{2}^{-1} + \binom{n-k^*}{2}^{-1}\right)\sum_{k^*<i<j\leqslant\hat{k}_n}(\psi_{i,j}-E\psi_{i,j})\\ &\qquad + \binom{\hat{k}_n}{2}^{-1}\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{\hat{k}_n}(\psi_{i,j}-E\psi_{i,j}) + \binom{n-k^*}{2}^{-1}\sum_{i=k^*+1}^{\hat{k}_n}\sum_{j=\hat{k}_n+1}^{n}(\psi_{i,j}-E\psi_{i,j})\\ &\qquad + \left(\binom{n-k^*}{2}^{-1} - \binom{n-\hat{k}_n}{2}^{-1}\right)\sum_{\hat{k}_n<i<j\leqslant n}(\psi_{i,j}-E\psi_{i,j})\\ &\quad\equiv \tilde{I} + \tilde{II} + \tilde{III} + \tilde{IV} + \tilde{V}.\end{aligned} \tag{A.11}$$

By the asymptotic normality of U-statistic, we obtain

$$\begin{aligned}\tilde{I} &= \left(\binom{\hat{k}_n}{2}^{-1}\binom{k^*}{2}-1\right)\binom{k^*}{2}^{-1}\sum_{1\leqslant i<j\leqslant k^*}(\psi_{i,j}-E\psi_{i,j})\\ &= O_p\left(\left(\binom{\hat{k}_n}{2}^{-1}\binom{k^*}{2}-1\right)/\sqrt{k^*}\right)=O_p\left(\frac{\tau^*-\hat{\tau}_n}{\sqrt{n}}\right).\end{aligned} \tag{A.12}$$

Similarly, we can obtain $\tilde{II}=O_p(\frac{(\tau-\tau^*)^{3/2}}{\sqrt{n}})$ and $\tilde{V}=O_p(\frac{\tau^*-\hat{\tau}_n}{\sqrt{n}})$.

Recall the proof of theorem 2.2(1), we have

$$\begin{aligned}I\tilde{I}I &= \binom{\hat{k}_n}{2}^{-1}\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{\hat{k}_n}(\psi)=\binom{\hat{k}_n}{2}^{-1}k^*(\hat{k}_n-k^*)\frac{1}{k^*}\sum_{i=1}^{k^*}(r_{1n}^*(Z_i)-v_{1n}-v_{2n})\\ &\quad+\binom{\hat{k}_n}{2}^{-1}k^*(\hat{k}_n-k^*)\frac{1}{\hat{k}_n-k^*}\sum_{j=k^*+1}^{\hat{k}_n}(r_{2n}^*(Z_j)-v_{1n}-v_{2n})\\ &\quad+\binom{\hat{k}_n}{2}^{-1}k^*(\hat{k}_n-k^*)\frac{1}{k^*(\hat{k}_n-k^*)}\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{\hat{k}_n}\tilde{\phi}_{i,j}\\ &\equiv \tilde{I}.1+\tilde{I}.2+\tilde{I}.3.\end{aligned} \tag{A.13}$$

By the central limit theory, we have $\tilde{I}.1=O_p(\frac{\hat{\tau}_n-\tau^*}{\sqrt{n}})$, $\tilde{I}.2=O_p(\frac{(\hat{\tau}_n-\tau^*)^{1/2}}{\sqrt{n}})$ and $\tilde{I}.3=o_p(\frac{(\hat{\tau}_n-\tau^*)^{1/2}}{\sqrt{n}})$. Thus, we obtain that $I\tilde{I}I=O_p(\frac{(\hat{\tau}_n-\tau^*)^{1/2}}{\sqrt{n}})$. Similar arguments give that $I\tilde{V}=O_p(\frac{(\hat{\tau}_n-\tau^*)^{1/2}}{\sqrt{n}})$. By (A.11)–(A.13) and Theorem 2.2(1), $\hat{\tau}_n-\tau^*\to_p 0$, we have that $Q(\hat{k}_n)-EQ(\hat{k}_n)-Q(k^*)+EQ(k^*)=O_p(\frac{(\hat{\tau}_n-\tau^*)^{1/2}}{\sqrt{n}})$.

Similarly, by the asymptotic normality of U-statistic, we have $Q(\hat{k}_n)-EQ(\hat{k}_n)=O_p(\frac{1}{\sqrt{n}})$ and $Q(k^*)-EQ(k^*)=O_p(\frac{1}{\sqrt{n}})$. By (A.6), we obtain that $EQ(\hat{k}_n)-EQ(k^*)=\frac{2(v_{1n}-v_{2n})(k^*-\hat{k}_n)}{\hat{k}_n}=O_p(\rho_n(\tau^*-\hat{\tau}_n))$ and $EQ(k^*)=O_p(\rho_n)$. Thus, note that $\sqrt{n}\rho_n\to\infty$ and $\hat{\tau}_n-\tau^*\to_p 0$, we can obtain that

$$\begin{aligned}A &= O_p\left(\frac{\hat{\tau}_n-\tau^*}{n}\right)+O_p\left(\frac{(\hat{\tau}_n-\tau^*)^{1/2}}{n}\right)+O_p\left(\frac{\rho_n(\tau^*-\hat{\tau}_n)}{\sqrt{n}}\right)+O_p\left(\frac{(\hat{\tau}_n-\tau^*)^{1/2}\rho_n}{\sqrt{n}}\right)\\ &= O_p\left(\frac{(\hat{\tau}_n-\tau^*)^{1/2}\rho_n}{\sqrt{n}}\right).\end{aligned}$$

Similar arguments also yield that, for $\hat{k}_n\leqslant k^*$,

$$A=O_p\left(\frac{(\tau^*-\hat{\tau}_n)^{1/2}\rho_n}{\sqrt{n}}\right).$$

Therefore, for any $\varepsilon>0$, there exists $M>0$ such that

$$P\left(\frac{\sqrt{n}}{|\hat{\tau}_n-\tau^*|^{1/2}\rho_n}8\rho_n^2|\hat{\tau}_n-\tau^*|\frac{\delta}{(1-\delta)^2}\leqslant M\right)\geqslant P\left(\frac{\sqrt{n}}{|\hat{\tau}_n-\tau^*|^{1/2}\rho_n}A\leqslant M\right)\geqslant 1-\varepsilon.$$

which leads to $|\hat{\tau}_n-\tau^*|=O_p(n^{-1}\rho_n^{-2})=O_p(n^{-1}\eta_n^{-2})$. The proof is completed. □

A.3 The proof of Theorem 2.2(3)

Proof. By Theorem 2.2(2), it follows that $\hat{k}_n=k^*+O_p(\rho_n^{-2})$. For any given $M>0$, we shall derive the limiting process of $V_n(s)=n\tau^*(1-\tau^*)(\|Q(k^*+\lfloor s\rho_n^{-2}\rfloor)\|^2-\|Q(k^*)\|^2)/8$, where $s\in[-M,M]$. For $s>0$, by the proof of Theorem 2.2(2), we have

$$V_n(s)=n\tau^*(1-\tau^*)(\|EQ(k^*+\lfloor s\rho_n^{-2}\rfloor)\|^2-\|EQ(k^*)\|^2)/8$$

$$+ 2n\tau^*(1-\tau^*)(EQ(k^*))^{\mathrm{T}}\binom{k^*+\lfloor s\rho_n^{-2}\rfloor}{2}^{-1}k^*\sum_{j=k^*+1}^{k^*+\lfloor s\rho_n^{-2}\rfloor}(r_{2n}^*(Z_j)-v_{1n}-v_{2n})/8$$
$$+ 2n\tau^*(1-\tau^*)(EQ(k^*))^{\mathrm{T}}\binom{n-k^*}{2}^{-1}(n-k^*-\lfloor s\rho_n^{-2}\rfloor)\sum_{j=k^*+1}^{k^*+\lfloor s\rho_n^{-2}\rfloor}(r_{2n}(Z_j)-2v_{2n})/8$$
$$+ o_p(1). \tag{A.14}$$

In view of (A.2), we obtain that, as $n\to\infty$,

$$\begin{aligned}
&n\tau^*(1-\tau^*)(\|EQ(k^*+\lfloor s\rho_n^{-2}\rfloor)\|^2-\|EQ(k^*)\|^2)/8\\
&\quad= n\tau^*(1-\tau^*)\left(4\|v_{1n}-v_{2n}\|^2\frac{(k^*)^2}{(k^*+\lfloor s\rho_n^{-2}\rfloor)^2}-4\|v_{1n}-v_{2n}\|^2\right)/8\\
&\quad= -n\tau^*(1-\tau^*)\rho_n^2\frac{\lfloor s\rho_n^{-2}\rfloor(2k^*+\lfloor s\rho_n^{-2}\rfloor)}{2(k^*+\lfloor s\rho_n^{-2}\rfloor)^2}\\
&\quad\to -(1-\tau^*)s.
\end{aligned} \tag{A.15}$$

Combining (2.4), (A.2), (A.6), (A.7) and functional central limit theorem, we get

$$\begin{aligned}
&2n\tau^*(1-\tau^*)(EQ(k^*))^{\mathrm{T}}\binom{k^*+\lfloor s\rho_n^{-2}\rfloor}{2}^{-1}k^*\sum_{j=k^*+1}^{k^*+\lfloor s\rho_n^{-2}\rfloor}(r_{2n}^*(Z_j)-v_{1n}-v_{2n})/8\\
&\qquad+2n\tau^*(1-\tau^*)(EQ(k^*))^{\mathrm{T}}\binom{n-k^*}{2}^{-1}(n-k^*-\lfloor s\rho_n^{-2}\rfloor)\sum_{j=k^*+1}^{k^*+\lfloor s\rho_n^{-2}\rfloor}(r_{2n}(Z_j)-2v_{2n})/8\\
&\quad=\rho_n\sum_{j=k^*+1}^{k^*+\lfloor s\rho_n^{-2}\rfloor}\frac{(\beta_1^*-\beta_2^*)^{\mathrm{T}}}{\|\beta_1^*-\beta_2^*\|}\left[(1-\tau^*)\left(r_2^*(Z_j)-\frac{\beta_1^*+\beta_2^*}{2}\right)+\tau^*(r_2(Z_j)-\beta_2^*)\right]+o_p(1)\\
&\quad\Rightarrow\lambda_1W_1(s),
\end{aligned} \tag{A.16}$$

where $W_1(s)$ is a standard Wiener process on $[0,\infty)$ and

$$\lambda_1=\left(E\left\|\frac{(\beta_1^*-\beta_2^*)^{\mathrm{T}}}{\|\beta_1^*-\beta_2^*\|}\left[(1-\tau^*)\left(r_2^*(Z_j)-\frac{\beta_1^*+\beta_2^*}{2}\right)+\tau^*(r_2(Z_j)-\beta_2^*)\right]\right\|^2\right)^{1/2}.$$

Thus, we obtain that, for $s>0$, $V_n(s)\Rightarrow\lambda_1W_1(s)-(1-\tau^*)s$. For the case $s\leqslant 0$, similar arguments also lead to $V_n(-s)\Rightarrow\lambda_2W_2(-s)+\tau^*s$, where $W_2(\cdot)$ is another Wiener process on $[0,\infty)$ independent of $W_1(\cdot)$ and

$$\lambda_2=\left(E\left\|\frac{(\beta_1^*-\beta_2^*)^{\mathrm{T}}}{\|\beta_1^*-\beta_2^*\|}\left[\tau^*\left(r_1^*(Z_i)-\frac{\beta_1^*+\beta_2^*}{2}\right)+(1-\tau^*)(r_1(Z_i)-\beta_1^*)\right]\right\|^2\right)^{1/2}.$$

Since $W_1(s)$ is a function of $\{X_i, i\leqslant k^*\}$ and $W_2(s)$ is a function of $\{X_i, i>k^*\}$, $W_1(s)$ and $W_2(s)$ are independent. By the continuous mapping theorem, we have $\rho_n^2(\hat{k}_n-k^*)\to_d \arg\max_s V(s)$, where

$$V(s)=\begin{cases}\lambda_1W_1(s)-(1-\tau^*)s, & s>0,\\ \lambda_2W_2(-s)+\tau^*s, & s\leqslant 0.\end{cases}$$

Setting $s=\lambda_2^2/(2\tau^*)^2v$, it can be shown that $\arg\max_s V(s)=\lambda_2^2/(2\tau^*)\arg\max_u G(u)$, where

$$G(u)=\begin{cases}\frac{\lambda_1}{\lambda_2}W_1(u)-\frac{1-\tau^*}{\tau^*}|u|/2, & u>0,\\ W_2(-u)-|u|/2, & u\leqslant 0.\end{cases}$$

This implies that $\frac{4(\tau^*)^2\rho_n^2}{\lambda_2^2}(\hat{k}_n-k^*)\to_d \arg\max_u G(u)$.

□