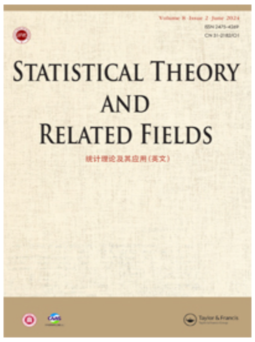

Statistical Theory and Related Fields

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tstf20

Estimation and inference for multi-kink expectile regression with nonignorable dropout

Dongyu Li & Lei Wang

To cite this article: Dongyu Li & Lei Wang (2024) Estimation and inference for multi-kink expectile regression with nonignorable dropout, Statistical Theory and Related Fields, 8:2, 136-151, DOI: 10.1080/24754269.2024.2302763

To link to this article: https://doi.org/10.1080/24754269.2024.2302763

STATISTICAL THEORY AND RELATED FIELDS
2024, VOL. 8, NO. 2, 136–151
https://doi.org/10.1080/24754269.2024.2302763

Estimation and inference for multi-kink expectile regression with nonignorable dropout

Dongyu Li and Lei Wang

School of Statistics and Data Science, KLMDASR, LEBPS and LPMC, Nankai University, Tianjin, People's Republic of China

ABSTRACT

In this paper, we consider parameter estimation, kink points testing and statistical inference for a longitudinal multi-kink expectile regression model with nonignorable dropout. In order to accommodate both within-subject correlations and nonignorable dropout, the bias-corrected generalized estimating equations are constructed by combining the inverse probability weighting and quadratic inference function approaches. The estimators for the kink locations and regression coefficients are obtained by using the generalized method of moments. A selection procedure based on a modified BIC is applied to estimate the number of kink points. We theoretically demonstrate the number selection consistency of kink points and the asymptotic normality of all estimators. A weighted cumulative sum type statistic is proposed to test the existence of kink effects at a given expectile, and its limiting distributions are derived under both the null and the local alternative hypotheses. Simulation studies show that the proposed estimators and test have desirable finite sample performance in both homoscedastic and heteroscedastic errors. An application to the Nation Growth, Lung and Health Study dataset is also presented.

ARTICLE HISTORY
Received 16 July 2023
Revised 29 December 2023
Accepted 3 January 2024

KEYWORDS
Dropout propensity; inverse probability weighting; missing not at random; nonresponse instrument; quadratic inference function

1. Introduction

Longitudinal data frequently arise in many fields where repeated measurements within the same subject are correlated, such as epidemiology, medical science and socioeconomic panel studies. In most longitudinal regression models, the impacts of covariates on responses are often assumed to be constant on the whole domain of the covariates, which may not be valid in some applications. For example, before and during puberty, a child's height would increase rapidly but stop increasing in the late teens, which shows there exists one change point. D. Li et al. (2022) investigated the relationship between the bike rental count and the time of the day using Capital Bike sharing data set in Washington D.C. and found that there are four kink points splitting the domain of the 24 hours into five periods: before sunrise, morning, early afternoon, late afternoon and evening. In these examples, the traditional linear regression may not fit well. Compared with the traditional regression models, the kink or change point regression models (B. Hansen, 2017) can achieve better performance and provide complementary information.

1.1. Related work

There exists a vast amount of literature related to kink regression models. For example, Lee et al. (2011), Lee et al. (2016), Fong (2019) and many others investigated a single unknown kink or change point estimation and inference problems. Bai and Perron (2003), Perron and Qu (2006) and Matteo et al. (2018) proposed the testing and estimation methods for multiple kink regression models. Alternatively, quantile regression (QR; Koenker & Bassett, 1978) and expectile regression (ER; Aigner et al., 1976; Newey & Powell, 1987) models are useful statistical tools for modelling and inferring the relationship between the response and covariates in some studies about the weights in child growth, high expenses in medical cost and so on. Compared with the traditional regression methods, both QR and ER can capture a complete picture of the relationship between the response and predictors. C. Li et al. (2011), Oka and Qu (2011) and L. Zhang et al. (2014) considered the estimation and testing problems for the single kink QR models while Zhong et al. (2022) investigated the multi-kink QR models and Wan et al. (2023) investigated it for the longitudinal data. Unlike QR, ER enjoys the computation efficiency, not only because of its differentiable L_2 loss, but also because the asymptotic covariance matrix of its estimator does not involve the density

CONTACT Lei Wang lwangstat@nankai.edu.cn, leiwang.stat@gmail.com School of Statistics and Data Science, KLMDASR, LEBPS and LPMC, Nankai University, Tianjin 300071, People's Republic of China

Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/24754269.2024.2302763.

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

function of the errors. Because of these advantages, F. Zhang and Li (2017) studied the single threshold ER models, and D. Li et al. (2022) considered the multi-kink ER models for the longitudinal data. However, it should be pointed out that all these existing methods are based on fully observed data and little knowledge is available on analysing multi-kink regression models in the presence of missing data.

Our study is motivated by the longitudinal body mass index (BMI) and blood pressure (BP) data on the Nation Growth, Lung and Health Study (NGHS) website. However, due to adverse events, the desire to seek other therapies and some other reasons, patients often drop out prior to the end of the study and the dropout rates of the follow-up times are 91.5%, 84.1%, 73.7%, 64.9%, 58.9%, 52.6%, 45.6%, 39.3%, 37.0% and 33.0%, respectively. In addition, previous experiences from doctors found that a steep rise in the BP indicates the disease progression, and patients with high BP values are more likely to drop out from the scheduled study visits as compared to patients with normal BP values, i.e., the nonresponse of the BP is likely related to itself and is nonignorable (Shao & Wang, 2016). Moreover, it can be checked that the distribution of BP is skewed such that mean regression may not appropriately assess the longitudinal change in BP data. Therefore, the existing methods may produce misleading results when applied to this NGHS data.

1.2. Our contributions

In this paper, we consider the situation where the covariates are always observed, but subjects may drop out before the end of the longitudinal study, which results in incomplete data. Dropout is ignorable if the dropout propensity depends on the observed values. Zhou and Liang (2021) investigated one change point estimation in the regression model when the response is missing at random. However, in practice the missing not at random (MNAR; L. Wang et al., 2019) or nonignorable dropout is more common; see L. Wang et al. (2019). It is well known that complete case (CC) analysis can not be trusted under nonignorable missing responses. Furthermore, L. Wang et al. (2019) and many others showed that developing valid methodologies for statistical analysis with nonignorable dropout is challenging, because the population parameters are not identifiable (Shao & Wang, 2016) if there is no assumption imposed, and the existing methods based on the assumption of ignorable dropouts may have large biases. To address the identifiability issue, S. Wang et al. (2014), Shao and Wang (2016), Miao and Tchetgen Tchetgn (2018), L. Wang et al. (2019) and many others proposed an instrumental variable approach for the parameter estimation with nonignorable nonresponse. Chen et al. (2022) studied Bayesian change-point joint models for multivariate longitudinal and time-to-event data, and discussed its application to nonignorable missing data.

The existence of nonignorable dropout and multiple kinks, and their impacts on the estimation and inference motivate us to search for an efficient and unified approach. To the best of our knowledge, the longitudinal multiple kink expectile regression (MKER) model with nonignorable dropout has not been investigated when the number of kink points and their locations are both unknown.

(1) In order to account for nonignorable dropout, we impose a parametric model on the dropout propensity and use a nonresponse instrument for the identifiability to consistently estimate the dropout propensity (L. Wang et al., 2019). The bias-corrected generalized estimating equations (GEEs; Liang & Zeger, 1986) by the inverse propensity weighting (IPW) are used to incorporate the within-subject correlations through a working correlation matrix. However, it should be pointed out that it is difficult to describe and specify the underlying within-subject covariance matrix. Motivated by the matrix expansion idea, the quadratic inference function (QIF; Qu et al., 2000) is applied to estimate parameters, which neither assumes the exact knowledge of the within-subject correlation matrix nor estimates the parameters of the within-subject correlation matrix. We estimate the number of kink points based on a modified BIC information criterion. The selection consistency of the number of kink points and the asymptotic normality of parameters are derived.
(2) A two-stage testing procedure for the existence of kink points at a given expectile level for longitudinal data with nonignorable dropout based on a weighted CUSUM type statistic is developed. The limiting distribution of the test statistic is also established. This two-stage testing procedure only requires fitting the ordinary ER model under the null hypothesis in the absence of kink points, avoids the estimation of the dropout propensity based on resampled data and is computationally much more efficient. Moreover, we propose a modified blockwise wild bootstrap to approximate the P-value.
(3) It is worthwhile to point out that the proposed estimation and testing methods can be used in balanced/imbalanced longitudinal data. Simulation results show that the proposed estimators of the regression coefficients and kink locations have good finite sample performance. In addition, compared with the CC test, our proposed test has better control of Type I error and higher powers in a wide range of scenarios. We also apply the proposed method to the longitudinal NGHS data. All the estimation and testing procedures are implemented in R codes, which are available when requested by readers.

1.3. Organization

The remainder of the article is organized as follows. In Section 2, we describe the longitudinal MKER model with nonignorable dropout, develop a parameter estimation procedure and study their asymptotic properties. In Section 3, we propose a testing procedure at a given expectile level and derive the limiting distribution of the statistic. In Section 4, we conduct simulation studies. A real data analysis is shown in Section 5. Section 6 concludes the remarks. The technique proofs are presented in the Appendix.

2. Methodology

2.1. Model and parameter estimation

Let $Y_i = (Y_{i1}, \ldots, Y_{im_i})^\top \in \mathbb{R}^{m_i}$ denote the ith subject's response and $\boldsymbol{D}_{ij} = (X_{ij}, \boldsymbol{Z}_{ij}^\top)^\top \in \mathbb{R}^{p+1}$ denote the corresponding covariance vector associated with Y_i, where $X_{ij} \in \Lambda$ is a bounded scalar covariate with multiple kink effects and $\boldsymbol{Z}_{ij}$ is a p-dimensional additional covariate for $j = 1, \ldots, m_i$ and $i = 1, \ldots, n$.

Given $\tau \in (0,1)$, define the τth expectile of Y as $\mathcal{E}_Y(\tau) = \arg\min_{\mathcal{E}} E\{\rho_\tau(Y - \mathcal{E})\}$ with $\rho_\tau(u) = u^2|\tau - I(u < 0)|$. In this paper, we consider a longitudinal MKER model with an undetermined number of kink points,

$$\mathcal{E}_{Y_{ij}}(\tau;\boldsymbol{\theta}|\boldsymbol{D}_{ij}) = \alpha_{0,\tau} + \alpha_{1,\tau}X_{ij} + \sum_{k=1}^{K} \beta_{k,\tau}(X_{ij} - \delta_{k,\tau})I(X_{ij} > \delta_{k,\tau}) + \boldsymbol{\gamma}_\tau^\top \boldsymbol{Z}_{ij}, \tag{1}$$

where $\mathcal{E}_{Y_{ij}}(\tau;\boldsymbol{\theta}|\boldsymbol{D}_{ij})$ is the τth conditional expectile of Y_{ij} given covariates $\boldsymbol{D}_{ij}$, $\beta_{k,\tau} \neq 0$ implies the existence of a kink effect at $X_{ij} = \delta_{k,\tau}$ and $\delta_{1,\tau} < \delta_{2,\tau} < \cdots < \delta_{K,\tau}$ are K unknown kink points. Denote $\boldsymbol{\nu}_\tau = (\alpha_{0,\tau}, \alpha_{1,\tau}, \beta_{1,\tau}, \ldots, \beta_{K,\tau}, \boldsymbol{\gamma}_\tau)^\top$ as the vector of the corresponding regression coefficients, $\boldsymbol{\delta}_\tau = (\delta_{1,\tau}, \ldots, \delta_{K,\tau})^\top$ as the vector of kink points and $\boldsymbol{\theta}_\tau = (\boldsymbol{\nu}_\tau^\top, \boldsymbol{\delta}_\tau^\top)^\top$. We omit the subscript τ for ease of notations.

In this paper, we consider the situation where $\boldsymbol{D}_{ij}$ is always observed, but subjects may drop out prior to the end of the study, which results in incomplete $\boldsymbol{Y}_i$ data. Let r_{ij} be the response indicator, i.e., $r_{ij} = 1$ if Y_{ij} is observed or $r_{ij} = 0$ otherwise, and define $\pi_{ij} = \Pr(r_{ij} = 1|\boldsymbol{D}_i, \boldsymbol{Y}_i)$ with $\boldsymbol{D}_i = (\boldsymbol{D}_{i1}, \ldots, \boldsymbol{D}_{im_i})^\top$. In the presence of missing data, we consider the following bias-corrected objective function to estimate $\boldsymbol{\theta}$, i.e.,

$$S_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{r_{ij}}{\pi_{ij}} \rho_\tau\{Y_{ij} - \mathcal{E}_{Y_{ij}}(\boldsymbol{\theta}|\boldsymbol{D}_{ij})\}. \tag{2}$$

Denote $\boldsymbol{W}_i = \mathrm{diag}(r_{i1}/\pi_{i1}, \ldots, r_{im_i}/\pi_{im_i})$, $\boldsymbol{\Psi}_\tau(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta})) = \mathrm{diag}(\Psi_\tau(Y_{i1} - \mathcal{E}_{Y_{i1}}(\boldsymbol{\theta})), \ldots, \Psi_\tau(Y_{im_i} - \mathcal{E}_{Y_{im_i}}(\boldsymbol{\theta})))$ with $\Psi_\tau(u) = |\tau - I(u < 0)|$ and $\mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}) = (\mathcal{E}_{Y_{i1}}(\boldsymbol{\theta}), \ldots, \mathcal{E}_{Y_{im_i}}(\boldsymbol{\theta}))^\top$. Motivated by Liang and Zeger (1986), the bias-corrected generalized estimating equations (GEEs) can be written as

$$\sum_{i=1}^{n} \boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta})^\top \boldsymbol{V}_{i\tau}^{-1} \boldsymbol{W}_i \boldsymbol{\Psi}_\tau(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}))(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta})) = \boldsymbol{0}, \tag{3}$$

where $\boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta}) = (\boldsymbol{\mathcal{X}}_{i1}(\boldsymbol{\theta}), \ldots, \boldsymbol{\mathcal{X}}_{im_i}(\boldsymbol{\theta}))^\top$ with $\boldsymbol{\mathcal{X}}_{ij}(\boldsymbol{\theta}) = (1, X_{ij}, (X_{ij} - \delta_1)_+, \ldots, (X_{ij} - \delta_K)_+, \boldsymbol{Z}_{ij}^\top, -\beta_1 I(X_{ij} > \delta_1), \ldots, -\beta_K I(X_{ij} > \delta_K))^\top$ and $\boldsymbol{V}_{i\tau}$ is the true covariance matrix of $\boldsymbol{\Psi}_\tau(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}_0))(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}_0))$ with the true vector $\boldsymbol{\theta}_0$. It should be pointed out that the following decomposition holds, i.e., $\boldsymbol{V}_{i\tau}^{-1} = \boldsymbol{A}_{i\tau}^{-1/2}\boldsymbol{R}_{i\tau}^{-1}\boldsymbol{A}_{i\tau}^{-1/2}$, with $\boldsymbol{A}_{i\tau}$ and $\boldsymbol{R}_{i\tau}$ being the true marginal variance matrix and correlation matrix of $\boldsymbol{\Psi}_\tau(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}_0))(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}_0))$ respectively.

To estimate the unknown parameter $\boldsymbol{\theta}$, a consistent estimator of π_{ij} should be obtained firstly. Using the instrumental estimating equations proposed by Shao and Wang (2016) and L. Wang et al. (2019), assume $\boldsymbol{D}_{ij}$ can be decomposed as two parts $\boldsymbol{U}_{ij}$ and $\boldsymbol{U}_{ij}^\perp$, i.e., $\boldsymbol{D}_{ij} = (\boldsymbol{U}_{ij}^\top, (\boldsymbol{U}_{ij}^\perp)^\top)^\top$. Denote $\overrightarrow{\boldsymbol{U}}_{ij} = (\boldsymbol{U}_{i1}^\top, \ldots, \boldsymbol{U}_{ij}^\top)^\top$, $\overrightarrow{\boldsymbol{Y}}_{ij} = (Y_{i1}, \ldots, Y_{ij})^\top$ and further assume that

$$\begin{aligned}
&\Pr(r_{ij} = 1|r_{i(j-1)} = 1, \boldsymbol{D}_i, \boldsymbol{Y}_i) = \Pr(r_{ij} = 1|r_{i(j-1)} = 1, \overrightarrow{\boldsymbol{U}}_{ij}, \overrightarrow{\boldsymbol{Y}}_{ij}) = \psi(\boldsymbol{O}_{ij}^\top \boldsymbol{\phi}_j),\\
&\Pr(r_{ij} = 1|r_{i(j-1)} = 0, \boldsymbol{D}_i, \boldsymbol{Y}_i) = 0, \quad j = 1, \ldots, m_i,\ i = 1, \ldots, n,
\end{aligned} \tag{4}$$

where $\boldsymbol{O}_{ij} = (1, \overrightarrow{\boldsymbol{U}}_{ij}^\top, \overrightarrow{\boldsymbol{Y}}_{ij}^\top)^\top$, $\boldsymbol{\phi}_j$ is an unknown parameter vector with the true value $\boldsymbol{\phi}_{j0}$, ψ defined on $[0,1]$ is a known monotone function and $r_{i0} = 1$. Under the model (4),

$$\pi_{ij} = \prod_{t=1}^{j} \Pr(r_{it} = 1 | r_{i(t-1)} = 1, \boldsymbol{D}_i, \boldsymbol{Y}_i) = \prod_{t=1}^{j} \psi(\boldsymbol{O}_{it}^\top \boldsymbol{\phi}_{t0}).$$

We then define the instrumental estimating equations

$$\boldsymbol{g}_{ij}(\boldsymbol{Y}_i, \boldsymbol{D}_i, \boldsymbol{r}_i, \boldsymbol{\phi}_j) = r_{i(j-1)} \left\{ \frac{r_{ij}}{\psi(\boldsymbol{O}_{ij}^\top \boldsymbol{\phi}_j)} - 1 \right\} \mathbb{S}(\overrightarrow{\boldsymbol{D}}_{ij}, \overrightarrow{\boldsymbol{Y}}_{i(j-1)}), \tag{5}$$

where $\boldsymbol{r}_i = (r_{i1}, \ldots, r_{im_i})^\top$ and $\mathbb{S}(\overrightarrow{\boldsymbol{D}}_{ij}, \overrightarrow{\boldsymbol{Y}}_{i(j-1)})$ is a known vector-valued function (L. Wang et al., 2019). The consistent estimator $\widehat{\boldsymbol{\phi}} = (\widehat{\boldsymbol{\phi}}_1^\top, \ldots, \widehat{\boldsymbol{\phi}}_{\max_i m_i}^\top)^\top$ of the true parameter vector $\boldsymbol{\phi}_0 = (\boldsymbol{\phi}_{10}^\top, \ldots, \boldsymbol{\phi}_{\max_i m_i 0}^\top)^\top$ can be obtained by two-step generalized method of moments (GMM; L. Hansen, 1982) in Shao and Wang (2016), L. Wang et al. (2019) and D. Li and Wang (2022). Once the estimator $\widehat{\boldsymbol{\phi}}$ is obtained, π_{ij} and $\boldsymbol{W}_i$ can be estimated by $\hat{\pi}_{ij} = \prod_{t=1}^{j} \psi(\boldsymbol{O}_{it}^\top \widehat{\boldsymbol{\phi}}_t)$ and $\boldsymbol{W}_i(\widehat{\boldsymbol{\phi}}) = \operatorname{diag}(r_{i1}/\hat{\pi}_{i1}, \ldots, r_{im_i}/\hat{\pi}_{im_i})$, respectively. Subsequently, a consistent estimator $\widehat{\boldsymbol{A}}_{i\tau}(\boldsymbol{\theta}, \widehat{\boldsymbol{\phi}})$ of $\boldsymbol{A}_{i\tau}(\boldsymbol{\theta})$ can be obtained by computing the marginal sample variance based on the IPW approach.

In practice, $\boldsymbol{R}_{i\tau}^{-1}$ is always unknown and we borrow the matrix expansion idea of the quadratic inference functions (QIF) method (Qu et al., 2000) to approximate $\boldsymbol{R}_{i\tau}^{-1} = \sum_{l=1}^{L} b_{l\tau} \boldsymbol{M}_{li}$, where $\boldsymbol{M}_{li}$'s are some given symmetric basic matrices and $b_{l\tau}$'s are unknown coefficients. Thereafter, the Equation (3) can be approximated by a linear combination of $\widehat{\boldsymbol{S}}_i(\boldsymbol{\theta})$, for $i = 1, \ldots, n$, as follows,

$$\widehat{\boldsymbol{S}}_i(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta})^\top \widehat{\boldsymbol{A}}_{i\tau}^{-1/2}(\boldsymbol{\theta}, \widehat{\boldsymbol{\phi}}) \boldsymbol{M}_{i1} \widehat{\boldsymbol{A}}_{i\tau}^{-1/2}(\boldsymbol{\theta}, \widehat{\boldsymbol{\phi}}) \boldsymbol{W}_i(\widehat{\boldsymbol{\phi}}) \boldsymbol{\Psi}_\tau(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}))(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta})) \\ \vdots \\ \boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta})^\top \widehat{\boldsymbol{A}}_{i\tau}^{-1/2}(\boldsymbol{\theta}, \widehat{\boldsymbol{\phi}}) \boldsymbol{M}_{iL} \widehat{\boldsymbol{A}}_{i\tau}^{-1/2}(\boldsymbol{\theta}, \widehat{\boldsymbol{\phi}}) \boldsymbol{W}_i(\widehat{\boldsymbol{\phi}}) \boldsymbol{\Psi}_\tau(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}))(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta})) \end{pmatrix}. \tag{6}$$

Since the number of estimation equations is greater than the number of parameters, we estimate $\boldsymbol{\theta}$ as follows:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \widehat{\boldsymbol{S}}(\boldsymbol{\theta})^\top \widehat{\boldsymbol{\Omega}}^{-1}(\boldsymbol{\theta}) \widehat{\boldsymbol{S}}(\boldsymbol{\theta}), \tag{7}$$

where $\widehat{\boldsymbol{S}}(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \widehat{\boldsymbol{S}}_i(\boldsymbol{\theta})$ and $\widehat{\boldsymbol{\Omega}}(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \widehat{\boldsymbol{S}}_i(\boldsymbol{\theta}) \widehat{\boldsymbol{S}}_i(\boldsymbol{\theta})^\top$. The estimation procedure for (7) can be summarized as the following Algorithm 1.

Algorithm 1 Parameter estimation procedure

(1) Obtain $\widehat{\boldsymbol{\phi}}$, $\hat{\pi}_{ij} = \prod_{t=1}^{j} \psi(\boldsymbol{O}_{it}^\top \widehat{\boldsymbol{\phi}}_t)$ and $\widehat{\boldsymbol{A}}_{i\tau}(\boldsymbol{\theta}, \widehat{\boldsymbol{\phi}})$.

(2) Given $\hat{\boldsymbol{\theta}}^{(r-1)}$ from the $r-1$ step, update

$$\widehat{\boldsymbol{A}}_{i\tau}(\hat{\boldsymbol{\theta}}^{(r-1)}, \widehat{\boldsymbol{\phi}}) = \operatorname{diag}\left(\frac{1}{|T_1|} \sum_{t \in |T_1|} \frac{r_{t1}}{\hat{\pi}_{t1}} \Psi_\tau(\hat{\varepsilon}_{t1})^2 \hat{\varepsilon}_{t1}^2, \ldots, \frac{1}{|T_{m_i}|} \sum_{t \in |T_{m_i}|} \frac{r_{tm_i}}{\hat{\pi}_{tm_i}} \Psi_\tau(\hat{\varepsilon}_{tm_i})^2 \hat{\varepsilon}_{tm_i}^2 \right),$$

$$\dot{\widehat{\boldsymbol{S}}}_{il}(\hat{\boldsymbol{\theta}}^{(r-1)}) = -\boldsymbol{\mathcal{X}}_i(\hat{\boldsymbol{\theta}}^{(r-1)})^\top \widehat{\boldsymbol{A}}_{i\tau}^{-1/2}(\hat{\boldsymbol{\theta}}^{(r-1)}, \widehat{\boldsymbol{\phi}}) \boldsymbol{M}_{il} \widehat{\boldsymbol{A}}_{i\tau}^{-1/2}(\hat{\boldsymbol{\theta}}^{(r-1)}, \widehat{\boldsymbol{\phi}}) \boldsymbol{W}_i(\widehat{\boldsymbol{\phi}}) \boldsymbol{\Psi}_\tau(\hat{\boldsymbol{\varepsilon}}_i) \boldsymbol{\mathcal{X}}_i(\hat{\boldsymbol{\theta}}^{(r-1)}),$$

with $\hat{\varepsilon}_{tj} = Y_{tj} - \mathcal{E}_{Y_{tj}}(\hat{\boldsymbol{\theta}}^{(r-1)} | \boldsymbol{D}_{tj})$, $\hat{\boldsymbol{\varepsilon}}_i = (\hat{\varepsilon}_{i1}, \ldots, \hat{\varepsilon}_{im_i})^\top$ and $T_j = \{t | m_t \geq j\}$.

(3) Obtain $\dot{\widehat{\boldsymbol{S}}}_i(\hat{\boldsymbol{\theta}}^{(r-1)}) = (\dot{\widehat{\boldsymbol{S}}}_{i1}(\hat{\boldsymbol{\theta}}^{(r-1)})^\top, \ldots, \dot{\widehat{\boldsymbol{S}}}_{iL}(\hat{\boldsymbol{\theta}}^{(r-1)})^\top)^\top$ and $\dot{\widehat{\boldsymbol{S}}}(\hat{\boldsymbol{\theta}}^{(r-1)}) = n^{-1} \sum_{i=1}^{n} \dot{\widehat{\boldsymbol{S}}}_i(\hat{\boldsymbol{\theta}}^{(r-1)})$. Update $\hat{\boldsymbol{\theta}}^{(r)}$ by

$$\hat{\boldsymbol{\theta}}^{(r)} = \hat{\boldsymbol{\theta}}^{(r-1)} - \{\dot{\widehat{\boldsymbol{S}}}(\hat{\boldsymbol{\theta}}^{(r-1)})^\top \widehat{\boldsymbol{\Omega}}^{-1}(\hat{\boldsymbol{\theta}}^{(r-1)}) \dot{\widehat{\boldsymbol{S}}}^{-1}(\hat{\boldsymbol{\theta}}^{(r-1)})\}^{-1} \dot{\widehat{\boldsymbol{S}}}(\hat{\boldsymbol{\theta}}^{(r-1)})^\top \widehat{\boldsymbol{\Omega}}^{-1}(\hat{\boldsymbol{\theta}}^{(r-1)}) \widehat{\boldsymbol{S}}(\hat{\boldsymbol{\theta}}^{(r-1)}).$$

(4) Repeat Steps (2)–(3) until convergence.

2.2. Number selection of kink points

It should be pointed out that the true number of kink points K_0 is always unknown in practice. In order to implement the estimation Algorithm 1, we need to identify the number of kink points K first. Following Zhong et al. (2022) and

D. Li et al. (2022), we propose to choose K by minimizing the following modified Bayesian information criterion:

$$\mathrm{BIC}(K) = \log\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{r_{ij}}{\hat{\pi}_{ij}}\rho_\tau\{Y_{ij} - \mathcal{E}_{Y_{ij}}(\hat{\boldsymbol{\theta}}_K|\boldsymbol{D}_{ij})\}\right\} + N_K\frac{\log(n)}{2n}C_n, \tag{8}$$

where $\hat{\boldsymbol{\theta}}_K$ is the estimator computed by the algorithm above with K kink points, N_K is the number of parameters, and C_n is a positive constant. Thus, given the maximum number of kink points K_*, the estimator for K_0 is

$$\widehat{K} = \underset{k=0,1,\ldots,K_*}{\arg\min}\ \mathrm{BIC}(k).$$

According to Zhong et al. (2022) and D. Li et al. (2022), the selection results are not very sensitive to the choice of the C_n value satisfying $C_n\log n/n = o(1)$ and they recommended to use $C_n = \log(n)$ to estimate K based on the BIC in the simulations.

2.3. Asymptotic properties

Next, we investigate the theoretical properties of the proposed estimator and the number selection consistency of the kink points. Let

$$\begin{aligned}
&\boldsymbol{\mathcal{W}}_j = E[\boldsymbol{g}_{ij}(\boldsymbol{Y}_i,\boldsymbol{D}_i,\boldsymbol{r}_i,\boldsymbol{\phi}_j)\boldsymbol{g}_{ij}(\boldsymbol{Y}_i,\boldsymbol{D}_i,\boldsymbol{r}_i,\boldsymbol{\phi}_j)^\top], \quad \boldsymbol{\pi}_i(\boldsymbol{\phi}_0) = \mathrm{diag}(\pi_{i1},\ldots,\pi_{im_i}),\\
&\boldsymbol{T}_l = E[-\boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta}_0)^\top\boldsymbol{A}_{i\tau}^{-1/2}\boldsymbol{M}_{il}\boldsymbol{A}_{i\tau}^{-1/2}\boldsymbol{\Psi}_\tau(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}_0))\boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta}_0)], \quad \boldsymbol{\Gamma}_j = E[\dot{\boldsymbol{g}}_{ij}(\boldsymbol{Y}_i,\boldsymbol{D}_i,\boldsymbol{r}_i,\boldsymbol{\phi}_j)],\\
&\boldsymbol{H}_l = E[-\boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta}_0)^\top\boldsymbol{A}_{i\tau}^{-1/2}\boldsymbol{M}_{il}\boldsymbol{A}_{i\tau}^{-1/2}\boldsymbol{\Psi}_\tau(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}_0))\mathrm{diag}(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}_0))\boldsymbol{\pi}_i^{-1}(\boldsymbol{\phi}_0)\dot{\boldsymbol{\pi}}_i(\boldsymbol{\phi}_0)],\\
&\boldsymbol{S}_{il}^0 = \boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta}_0)^\top\boldsymbol{A}_{i\tau}^{-1/2}\boldsymbol{M}_{il}\boldsymbol{A}_{i\tau}^{-1/2}\boldsymbol{W}_i(\boldsymbol{\phi}_0)\boldsymbol{\Psi}_\tau(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}_0))(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}_0)), \quad \boldsymbol{S}_i^0 = (\boldsymbol{S}_{i1}^{0\top},\ldots,\boldsymbol{S}_{iL}^{0\top})^\top.
\end{aligned}$$

We then make the following assumptions.

- (A1) $\{\boldsymbol{D}_i,\boldsymbol{Y}_i\}_{i=1}^n$ are independent and identically distributed (i.i.d.) random vectors. The response probability function $\pi_{ij}(\boldsymbol{\phi})$ for all j satisfies: (a) it is strictly monotone and twice differentiable with respect to $\boldsymbol{\phi}$; (b) $0 < c_0 < \pi_{ij}(\boldsymbol{\phi}) < 1$ for a positive constant c_0 and $\dot{\boldsymbol{\pi}}_{ij}(\boldsymbol{\phi})$ is uniformly bounded for any $\boldsymbol{\phi}$ in a neighbourhood of $\boldsymbol{\phi}_0$.
- (A2) $E[\|\boldsymbol{D}_i\|^2] < \infty$ and there exists a neighbourhood $\boldsymbol{\mathcal{N}}_j$ of $\boldsymbol{\phi}_j$ such that $E[r_{ij}\sup_{\boldsymbol{\vartheta}} \in \boldsymbol{\mathcal{N}}_j\{(1+\|\boldsymbol{D}_i\|^2)\pi_{ij}(\boldsymbol{\vartheta})^2 + \|\mathbb{S}(\boldsymbol{D}_i,\overrightarrow{\boldsymbol{Y}}_{i(j-1)})\|_1\|\mathbb{S}(\boldsymbol{D}_i,\overrightarrow{\boldsymbol{Y}}_{ij})\|_1|\dot{\pi}_{ij}(\boldsymbol{\vartheta})| + \|\mathbb{S}(\boldsymbol{D}_i,\overrightarrow{\boldsymbol{Y}}_{ij})\|_2|\ddot{\pi}_{ij}(\boldsymbol{\vartheta})|\}] < \infty$, where $\|\cdot\|$ is the L_2-norm and $j = 1,\ldots,\max_i m_i$.
- (A3) The true parameter $\boldsymbol{\theta}_0$ is in a compact subset $\boldsymbol{\Theta}$ of $\mathbb{R}^{2+p+2K_0}$.
- (A4) The true number of kink points K_0 and the dimension p of $\boldsymbol{Z}_{ij}$ are fixed; T_j goes to infinity as n goes to infinity for $j = 1,\ldots,\max_i m_i$ with $\max_i m_i < \infty$.
- (A5) The bounded scalar variable X_{ij} has a continuous distribution with density function f_X, which is strictly positive, bounded and continuous for any δ in a neighbourhood of δ_0, $E\|\boldsymbol{Z}_{ij}\|^4 < \infty$, $\max_{i,j}\|\boldsymbol{Z}_{ij}\| = o_p(n^{1/2})$ and $E|Y_{ij}|^4 < \infty$.
- (A6) $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\Gamma}_1,\ldots,\boldsymbol{\Gamma}_{\max_i m_i})$ is of full rank.
- (A7) There exist two positive constants c_1 and c_2 such that $c_1 \le \lambda_{\min}(\frac{1}{n}\sum_{i=1}^n\boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta}_0)^\top\boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta}_0)) \le \lambda_{\max}(\frac{1}{n}\sum_{i=1}^n \boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta}_0)^\top\boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta}_0)) \le c_2$, where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the smallest and largest eigenvalues of a matrix.
- (A8) The eigenvalues for each $\boldsymbol{M}_l$ for $l = 1,\ldots,L$ are bounded away from 0 and ∞.
- (A9) $\boldsymbol{H} = (\boldsymbol{H}_1^\top,\ldots,\boldsymbol{H}_L^\top)^\top$ is of full rank.
- (A10) There exists a positive definite matrix $S_v = E[\sum_j|\tau - I(Y_{ij} \le \boldsymbol{\alpha}^\top\boldsymbol{\mathcal{A}}_{ij})|\boldsymbol{\mathcal{A}}_{ij}\boldsymbol{\mathcal{A}}_{ij}^\top]$.

Assumptions (A1)–(A2) and (A6) are common assumptions and restrictions on the propensity function, which also can be found in missing data literature (L. Wang et al., 2019; S. Wang et al., 2014). Assumption (A4) is a condition for the longitudinal data and implies that the dimension of the unknown parameter we are interested in is finite. Assumptions (A3) and (A5) are some common conditions for studying asymptotic properties of ER models, which also can be found in F. Zhang and Li (2017), and Assumption (A5) is a condition on threshold variable X_{ij} and covariates $\boldsymbol{Z}_{ij}$. Assumptions (A7)-(A10) are imposed for some invertible matrices to illustrate the consistency and asymptotic normality property.

Theorem 2.1: *Under* (A1)–(A5), $C_n\log n/n = o(1)$ *and* $\widehat{K} = \arg\min_{k=0,\ldots,K_*}\mathrm{BIC}(k)$, *we have* $P(\widehat{K} = K_0) \to 1$ *as* $n \to \infty$.

Theorem 2.1 shows that the modified BIC can consistently select the true number of kink points, which plays a fundamental role in statistical inference. Our simulation results in Section 4 show that the correct selection rate is close to one when a proper C_n is given. The following theorem studies the limiting distribution of the proposed estimator $\hat{\boldsymbol{\theta}}$ obtained by (7) given the true number of kink points.

Theorem 2.2: *Under* (A1)–(A9), *as* $n \to \infty$, *we have*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, (\boldsymbol{T}^\top \boldsymbol{\Xi}^{-1} \boldsymbol{T})^{-1}),$$

where $\boldsymbol{T} = (\boldsymbol{T}_1^\top, \ldots, \boldsymbol{T}_L^\top)^\top$, $\boldsymbol{\Xi} = \boldsymbol{B} + \boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^\top$, $\boldsymbol{B} = E[\boldsymbol{S}_i^0 \boldsymbol{S}_i^{0\top}]$, $\boldsymbol{\Sigma} = (\boldsymbol{\Gamma}^\top \boldsymbol{\mathcal{W}}^{-1} \boldsymbol{\Gamma})^{-1}$ *and* $\boldsymbol{\mathcal{W}} = \mathrm{diag}(\boldsymbol{\mathcal{W}}_1, \ldots, \boldsymbol{\mathcal{W}}_{\max_i m_i})$.

By the plug-in method, we obtain the estimators $\widehat{\boldsymbol{\Xi}}$ and $\widehat{\boldsymbol{T}}$ of $\boldsymbol{\Xi}$ and $\boldsymbol{T}$ as follows:

$$\widehat{\boldsymbol{\Xi}} = \widehat{\boldsymbol{B}} + \widehat{\boldsymbol{H}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{H}}^\top, \quad \widehat{\boldsymbol{B}} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{S}}_i(\hat{\boldsymbol{\theta}})\widehat{\boldsymbol{S}}_i(\hat{\boldsymbol{\theta}})^\top, \quad \widehat{\boldsymbol{\mathcal{W}}} = \mathrm{diag}(\widehat{\boldsymbol{\mathcal{W}}}_1, \ldots, \widehat{\boldsymbol{\mathcal{W}}}_{\max_i m_i}),$$

$$\widehat{\boldsymbol{\Gamma}} = \mathrm{diag}(\widehat{\boldsymbol{\Gamma}}_1, \ldots, \widehat{\boldsymbol{\Gamma}}_{\max_i m_i}), \quad \widehat{\boldsymbol{\Gamma}}_j = \frac{1}{|T_j|}\sum_{i \in T_j} \dot{\boldsymbol{g}}_{ij}(\boldsymbol{Y}_i, \boldsymbol{D}_i, \boldsymbol{r}_i, \widehat{\boldsymbol{\phi}}_j), \quad \widehat{\boldsymbol{\Sigma}} = (\widehat{\boldsymbol{\Gamma}}^\top \widehat{\boldsymbol{\mathcal{W}}}^{-1} \widehat{\boldsymbol{\Gamma}})^{-1},$$

$$\widehat{\boldsymbol{\mathcal{W}}}_j = \frac{1}{|T_j|}\sum_{i \in T_j} \boldsymbol{g}_{ij}(\boldsymbol{Y}_i, \boldsymbol{D}_i, \boldsymbol{r}_i, \widehat{\boldsymbol{\phi}}_j)\boldsymbol{g}_{ij}(\boldsymbol{Y}_i, \boldsymbol{D}_i, \boldsymbol{r}_i, \widehat{\boldsymbol{\phi}}_j)^\top, \quad \widehat{\boldsymbol{T}} = (\widehat{\boldsymbol{T}}_1^\top, \ldots, \widehat{\boldsymbol{T}}_L^\top)^\top,$$

$$\widehat{\boldsymbol{T}}_l = -\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\mathcal{X}}_i(\hat{\boldsymbol{\theta}})^\top \widehat{\boldsymbol{A}}_{i\tau}^{-1/2}(\hat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}})\boldsymbol{M}_{il}\widehat{\boldsymbol{A}}_{i\tau}^{-1/2}(\hat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}})\boldsymbol{W}_i(\widehat{\boldsymbol{\phi}})\boldsymbol{\Psi}_\tau(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\hat{\boldsymbol{\theta}}))\boldsymbol{\mathcal{X}}_i(\hat{\boldsymbol{\theta}}),$$

$$\begin{aligned}\widehat{\boldsymbol{H}}_l = &-\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\mathcal{X}}_i(\hat{\boldsymbol{\theta}})^\top \widehat{\boldsymbol{A}}_{i\tau}^{-1/2}(\hat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}})\boldsymbol{M}_{il}\widehat{\boldsymbol{A}}_{i\tau}^{-1/2}(\hat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}})\boldsymbol{W}_i(\widehat{\boldsymbol{\phi}})\boldsymbol{\Psi}_\tau(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\hat{\boldsymbol{\theta}}))\\ &\times \mathrm{diag}(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\hat{\boldsymbol{\theta}}))\boldsymbol{\pi}_i^{-1}(\widehat{\boldsymbol{\phi}})\dot{\boldsymbol{\pi}}_i(\widehat{\boldsymbol{\phi}}), \quad \widehat{\boldsymbol{H}} = (\widehat{\boldsymbol{H}}_1^\top, \ldots, \widehat{\boldsymbol{H}}_L^\top)^\top.\end{aligned}$$

Once $\widehat{\boldsymbol{\Xi}}$ and $\widehat{\boldsymbol{T}}$ are obtained, $(\widehat{\boldsymbol{T}}^\top \widehat{\boldsymbol{\Xi}}^{-1} \widehat{\boldsymbol{T}})^{-1}$ is used to estimate the variance matrix $(\boldsymbol{T}^\top \boldsymbol{\Xi}^{-1} \boldsymbol{T})^{-1}$. Hence, one can build a normal approximation based confidence region from Theorem 2.2.

Remark 2.1: When missing is not at random, it should be pointed out that the estimator based on the CC method is not consistent. On the other hand, it can be verified that $n^{-1}\sum_{i=1}^{n} \widehat{\boldsymbol{S}}_i(\boldsymbol{\theta}_0)\widehat{\boldsymbol{S}}_i(\boldsymbol{\theta}_0)^\top \xrightarrow{p} \boldsymbol{B}$ while $n^{-1/2}\sum_{i=1}^{n} \widehat{\boldsymbol{S}}_i(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Xi})$ as $n \to \infty$. Due to additional estimation of the nuisance parameter vector $\boldsymbol{\phi}$, $\boldsymbol{\Xi} > \boldsymbol{B}$ such that the CC estimator may generally have a smaller variance compared with our proposed estimators, which can be seen in our simulation results. When there is no missing data, i.e., $\pi_{ij} = 1$ for $j = 1, \ldots, m_i$ and $i = 1, \ldots, n$, it can be verified that $\boldsymbol{H} = \mathbf{0}$, $\boldsymbol{\Xi} = \boldsymbol{B}$ with

$$\begin{aligned}(\boldsymbol{B})_{ll'} = &E[\boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta}_0)^\top \boldsymbol{A}_{i\tau}^{-1/2}\boldsymbol{M}_{il}\boldsymbol{A}_{i\tau}^{-1/2}\boldsymbol{\Psi}_\tau(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}_0))(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}_0))\\ &\times (\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}_0))^\top \boldsymbol{\Psi}_\tau(\boldsymbol{Y}_i - \mathcal{E}_{\boldsymbol{Y}_i}(\boldsymbol{\theta}_0))\boldsymbol{A}_{i\tau}^{-1/2}\boldsymbol{M}_{il'}\boldsymbol{A}_{i\tau}^{-1/2}\boldsymbol{\mathcal{X}}_i(\boldsymbol{\theta}_0)],\end{aligned}$$

being the (l, l')th block of $\boldsymbol{B}$ with $l = 1, \ldots, L$ and $l' = 1, \ldots, L$.

3. Statistical inference for kink effects and location parameters

In this section, we are interested in testing whether there exist kink points in the presence of nonignorable dropout, rather than the specific number of kink points. Consider the following null (H_0) and alternative (H_1) hypotheses,

$$H_0 : \beta_k = 0 \text{ for all } k = 1, \ldots, K, \quad \text{v.s. } H_1 : \beta_k \neq 0 \text{ for some } k = 1, \ldots, K.$$

Let $\boldsymbol{\mathcal{A}}_{ij} = (1, X_{ij}, \boldsymbol{Z}_{ij}^\top)^\top$ and $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \boldsymbol{\gamma}^\top)^\top$. We define the following statistic

$$R_n(\delta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{r_{ij}}{\hat{\pi}_{ij}}|\tau - I(Y_{ij} \leq \hat{\boldsymbol{\alpha}}^\top \boldsymbol{\mathcal{A}}_{ij})|(Y_{ij} - \hat{\boldsymbol{\alpha}}^\top \boldsymbol{\mathcal{A}}_{ij})(X_{ij} - \delta)I(X_{ij} \leq \delta),$$

where $\hat{\boldsymbol{\alpha}} = \arg\min \sum_{i,j} r_{ij}/\hat{\pi}_{ij}\rho_\tau(Y_{ij} - \alpha^\top \boldsymbol{\mathcal{A}}_{ij})$ is the IPW estimator under the null hypothesis H_0. The proposed test statistic is

$$T_n = \sup_{\delta\in\Delta} |R_n(\delta)|,$$

where Δ is the compact set of δ. Intuitively, when there is no kink point, the statistic T_n will be very small since the estimated residuals show a random pattern against the variable X_{ij}, and when there exists at least one kink point, the estimated residuals fail to be consistent with the real residuals such that T_n is large. However, it can be verified that the CC estimator $\hat{\boldsymbol{\alpha}} = \arg\min \sum_{i,j} r_{ij}\rho_\tau(Y_{ij} - \alpha^\top \boldsymbol{\mathcal{A}}_{ij})$ is not consistent with $\boldsymbol{\alpha}$ in the presence of the nonignorable dropout. Hence, we directly use the test of D. Li et al. (2022), we will obtain a much larger test statistic even under the null hypothesis. Then, to derive the limiting distribution of T_n, we consider the following local alternative model,

$$Y_{ij} = \alpha_0 + \alpha_1 X_{ij} + n^{-1/2}\beta(X_{ij} - \delta)_+ + \boldsymbol{\gamma}^\top \boldsymbol{Z}_{ij} + \varepsilon_{ij}, \tag{9}$$

where the τth expectile of ε_{ij} is zero, and introduce some notations:

$$\widehat{S}_{1n}(\delta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{r_{ij}}{\hat{\pi}_{ij}}|\tau - I(Y_{ij} \le \hat{\boldsymbol{\alpha}}^\top \boldsymbol{\mathcal{A}}_{ij})|\boldsymbol{\mathcal{A}}_{ij}(X_{ij} - \delta)I(X_{ij} \le \delta),$$

$$S_1(\delta) = E\left[\sum_{j=1}^{m_i}|\tau - I(Y_{ij} \le \boldsymbol{\alpha}^\top \boldsymbol{\mathcal{A}}_{ij})|\boldsymbol{\mathcal{A}}_{ij}(X_{ij} - \delta)I(X_{ij} \le \delta)\right],$$

$$\widehat{S}_{2n}(\delta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{r_{ij}}{\hat{\pi}_{ij}}|\tau - I(Y_{ij} \le \hat{\boldsymbol{\alpha}}^\top \boldsymbol{\mathcal{A}}_{ij})|\boldsymbol{\mathcal{A}}_{ij}\beta(X_{ij} - \delta)I(X_{ij} \ge \delta),$$

$$S_2(\delta) = E\left[\sum_{j=1}^{m_i}|\tau - I(Y_{ij} \le \boldsymbol{\alpha}^\top \boldsymbol{\mathcal{A}}_{ij})|\boldsymbol{\mathcal{A}}_{ij}\beta(X_{ij} - \delta)I(X_{ij} \ge \delta)\right],$$

$$\widehat{S}_{vn} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{r_{ij}}{\hat{\pi}_{ij}}|\tau - I(Y_{ij} \le \hat{\boldsymbol{\alpha}}^\top \boldsymbol{\mathcal{A}}_{ij})|\boldsymbol{\mathcal{A}}_{ij}\boldsymbol{\mathcal{A}}_{ij}^\top,$$

$$S_v = E\left[\sum_{j=1}^{m_i}|\tau - I(Y_{ij} \le \boldsymbol{\alpha}^\top \boldsymbol{\mathcal{A}}_{ij})|\boldsymbol{\mathcal{A}}_{ij}\boldsymbol{\mathcal{A}}_{ij}^\top\right], \quad q(\delta) = S_1(\delta)^\top S_v^{-1} S_2(\delta).$$

Theorem 3.1: *Under* (A1)–(A6) *and* (A10), *for the local alternative model* (9), $R_n(\delta)$ *has the asymptotic representation*

$$\begin{aligned} R_n(\delta) = {} & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{r_{ij}}{\hat{\pi}_{ij}}\varepsilon_{ij}|\tau - I(Y_{ij} \le \boldsymbol{\alpha}^\top \boldsymbol{\mathcal{A}}_{ij})|[(X_{ij} - \delta)I(X_{ij} \le \delta) - S_1(\delta)^\top S_v^{-1}\boldsymbol{\mathcal{A}}_{ij}] \\ & - q(\delta) + o_p(1) := R(\delta) - q(\delta) + o_p(1). \end{aligned}$$

Furthermore, T_n *converges weakly to the process* $\sup_\delta |R(\delta) - q(\delta)|$, *where* $R(\delta)$ *is the Gaussian process with mean zero and covariance function*

$$\begin{aligned} & E\Bigg[\sum_{j,j'}\frac{r_{ij}}{\pi_{ij}}\varepsilon_{ij}\big|\tau - I(Y_{ij} \le \boldsymbol{\alpha}^\top \boldsymbol{\mathcal{A}}_{ij})\big|\frac{r_{ij'}}{\pi_{ij'}}\varepsilon_{ij'}\big|\tau - I(Y_{ij'} \le \boldsymbol{\alpha}^\top \boldsymbol{\mathcal{A}}_{ij'})\big|\{(X_{ij} - \delta_1)I(X_{ij} \le \delta_1) \\ & - S_1(\delta_1)^\top S_v^{-1}\boldsymbol{\mathcal{A}}_{ij}\}\{(X_{ij'} - \delta_2)I(X_{ij'} \le \delta_2) - S_1(\delta_2)^\top S_v^{-1}\boldsymbol{\mathcal{A}}_{ij'}\}\Bigg] + \boldsymbol{\mathcal{H}}(\delta_1)\boldsymbol{\Sigma}\boldsymbol{\mathcal{H}}(\delta_2)^\top, \end{aligned}$$

with $\boldsymbol{\mathcal{H}}(\delta) = E\left[-\sum_j \frac{\dot{\boldsymbol{\pi}}_{ij}^\top}{\pi_{ij}}\varepsilon_{ij}|\tau - I(Y_{ij} \le \boldsymbol{\alpha}^\top \boldsymbol{\mathcal{A}}_{ij})|\{(X_{ij} - \delta)I(X_{ij} \le \delta) - S_1(\delta)^\top S_v^{-1}\boldsymbol{\mathcal{A}}_{ij}\}\right]$.

According to Theorem 3.1, when $\beta = 0$, i.e., under the null hypothesis, $q(\delta) = 0$ and $R_n(\delta)$ would converge to a Gaussian process $R(\delta)$ with mean zero. However, under the alternative hypotheses, i.e., $\beta \neq 0$, it can be seen that

$q(\delta) \neq 0$ and $T_n(\delta)$ would be significantly larger than zero, which provides evidence to reject the null hypothesis. When there is no missing data, i.e., $\pi_{ij} = 1$ for $j = 1, \ldots, m_i$ and $i = 1, \ldots, n$, we have $\boldsymbol{\mathcal{H}}(\delta) = \mathbf{0}$ and the covariance function of $R(\delta)$ is same as that of D. Li et al. (2022). Hence, the additional part $\boldsymbol{\mathcal{H}}(\delta_1)\boldsymbol{\Sigma}\boldsymbol{\mathcal{H}}(\delta_2)^\top$ reflects the cost of the estimation for the nuisance parameter vector $\boldsymbol{\phi}$ in the nonignorable propensity, which is different from $R(\delta)$ of D. Li et al. (2022) when there is no missing data. Because the limiting null distribution of T_n is nonstandard, we propose to approximate the P-values using a two-stage modified blockwise wild bootstrap and the procedure is described in Algorithm 2.

Algorithm 2 Modified blockwise wild bootstrap method

(1) Generate i.i.d. standard normal samples $\{v_1, v_2, \ldots, v_n\}$ and $\boldsymbol{\Delta}_{\boldsymbol{\phi}} \sim N(\mathbf{0}, \widehat{\boldsymbol{\Sigma}})$.

(2) Compute the test statistic $T_n^* = \sup_{\delta \in \Delta} |R_n^*(\delta)|$ with residuals $\hat{\varepsilon}_{ij}$ under null hypothesis, where

$$R_n^*(\delta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_i \sum_{j=1}^{m_i} \frac{r_{ij}}{\hat{\pi}_{ij}} \hat{\varepsilon}_{ij} |\tau - I(\hat{\varepsilon}_{ij} < 0)| [(X_{ij} - \delta) I(X_{ij} \leq \delta) - \widehat{S}_{1n}(\delta)^\top \widehat{S}_{nv}^{-1} \boldsymbol{\mathcal{A}}_{ij}]$$
$$- \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{r_{ij} \dot{\hat{\boldsymbol{\pi}}}_{ij}^\top \boldsymbol{\Delta}_{\boldsymbol{\phi}}}{\hat{\pi}_{ij}^2} \hat{\varepsilon}_{ij} |\tau - I(\hat{\varepsilon}_{ij} < 0)| [(X_{ij} - \delta) I(X_{ij} \leq \delta) - \widehat{S}_{1n}(\delta)^\top \widehat{S}_{nv}^{-1} \boldsymbol{\mathcal{A}}_{ij}].$$

(3) Repeat Steps (1)–(2) B times and calculate the P-values

$$\hat{p}_n = B^{-1} \sum_{t=1}^{B} I(T_n^{*(t)} \geq T_n).$$

Note that this modified bootstrap procedure avoids the estimation of π_{ij} based on the resampled data and only requires the estimators under the null hypothesis. Hence, this procedure is computationally efficient and can be directly applied to test the existence of multiple kink points as well. When there is no missing data, $R_n^*(\delta)$ is the same as that of D. Li et al. (2022). The following theorem shows the validity of the proposed blockwise wild bootstrap scheme.

Theorem 3.2: *Under the assumptions in Theorem 3.1, $R_n^*(\delta)$ defined in Algorithm 2 converges to the Gaussian process $R(\delta)$ as $n \to \infty$ under both the null and the local alternative hypotheses.*

4. Simulation studies

In this section, we conduct simulation studies to assess the finite-sample performance of the following estimators based on the Equation (6):

(i) the complete case (CC) estimator using the compound symmetry (CS) structure, i.e., $\boldsymbol{M}_{i1}$ is an identity matrix and $\boldsymbol{M}_{i2}$ is a symmetric matrix with 0 on the diagonal and 1 elsewhere, with $\boldsymbol{W}_i(\widehat{\boldsymbol{\phi}}) = \mathrm{diag}(r_{i1}, \ldots, r_{im_i})$;

(ii) the proposed MNAR estimator ($\mathrm{MNAR_{Ind}}$) using the independent (Ind) structure, i.e., $\widehat{\boldsymbol{A}}_{i\tau}^{-1/2}(\boldsymbol{\theta}, \widehat{\boldsymbol{\phi}})\boldsymbol{M}_{il}\widehat{\boldsymbol{A}}_{i\tau}^{-1/2}(\boldsymbol{\theta}, \widehat{\boldsymbol{\phi}}) = \boldsymbol{I}_{m_i}$;

(iii) the proposed MNAR estimator ($\mathrm{MNAR_{CS}}$) using the CS structure;

(iv) the proposed MNAR estimator ($\mathrm{MNAR_{AR1}}$) using the first-order autoregressive (AR1) structure, i.e., $\boldsymbol{M}_{i1}$ is an identity matrix, $\boldsymbol{M}_{i2}$ has 1 on the two mapin subdiagonals and 0 elsewhere, and $\boldsymbol{M}_{i3}$ has 1 on $(1, 1)$ and (m_i, m_i) components and 0 elsewhere;

(v) the full sample (Full) estimator using the CS structure with $\boldsymbol{W}_i(\widehat{\boldsymbol{\phi}}) = \boldsymbol{I}_{m_i}$.

In estimators (i) and (v), the true structure of $\boldsymbol{R}_i$ is used to obtain their best results. In particular, we evaluate the accuracy of estimation, the converge probabilities, the Type I error and the power of the proposed test. We generate

$$Y_{ij} = \alpha_0 + \alpha_1 X_{ij} + \sum_{k=1}^{K} \beta_k (X_{ij} - \delta_k)_+ + \gamma Z_{ij} + (1 + \ell |Z_{ij}|)\varepsilon_{ij},$$

where $m_i = 4$, $X_{ij} \sim U(-5, 5)$, $Z_{ij} \sim N(1, 0.5^2)$, X_{ij} and Z_{ij} are independent. We set $(\alpha_0, \alpha_1, \gamma) = (1, 1, 1)$ and consider three different numbers of kink points:

(i) $K = 1$, $\beta_1 = -3$ and $\delta_1 = 0.5$;
(ii) $K = 2$, $(\beta_1, \beta_2) = (-3, 4)$ and $(\delta_1, \delta_2) = (-1, 2)$;
(iii) $K = 3$, $(\beta_1, \beta_2, \beta_3) = (-3, 4, 4)$ and $(\delta_1, \delta_2, \delta_3) = (-2, 1, 3)$.

The random errors $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}, \varepsilon_{i4})^\top$ are generated from four different distributions: (a) Homoscedastic normal errors: $\ell = 0$ and $\boldsymbol{\varepsilon}_i \sim N(0, \Sigma)$, where $\Sigma_{jj'} = \rho$ for $j \neq j'$, and $\Sigma_{jj} = 1$ for $j = 1, \ldots, 4$; (b) Homoscedastic t errors: $\ell = 0$ and $\boldsymbol{\varepsilon}_i \sim t_{10}(0, \Sigma)$; (c) Heteroscedastic normal errors: $\ell = 0.1$ and $\boldsymbol{\varepsilon}_i \sim N(0, \Sigma)$; (d) Heteroscedastic t errors: $\ell = 0.1$ and $\boldsymbol{\varepsilon}_i \sim t_{10}(0, \Sigma)$. We generate these errors with $\rho = 0.5$ such that elements of $\boldsymbol{\varepsilon}_i$ are correlated.

The missing indicators $(r_{i1}, r_{i2}, r_{i3}, r_{i4})$ are generated from the following propensity model:

$$\Pr(r_{ij} = 1 | r_{i(j-1)} = 1, \overrightarrow{\boldsymbol{D}}_{ij}, \overrightarrow{\boldsymbol{Y}}_{ij}) = 1/(1 + \exp(\boldsymbol{h}_{ij}^\top \boldsymbol{\phi}_{j0})), \tag{10}$$

where $\boldsymbol{h}_{ij} = (1, X_{ij}, Y_{ij})^\top$ and $\boldsymbol{\phi}_{j0}$ is a part of $\boldsymbol{\phi}_0$. It can be seen that $\boldsymbol{U}_{ij} = X_{ij}$ and the instrumental variable $\boldsymbol{U}_{ij}^{\perp} = Z_{ij}$. Two different choices of $\boldsymbol{\phi}_0 = (\boldsymbol{\phi}_{10}^\top, \boldsymbol{\phi}_{20}^\top, \boldsymbol{\phi}_{30}^\top, \boldsymbol{\phi}_{40}^\top)^\top$ are considered.

(M1) $\boldsymbol{\phi}_{j0} = (-1.5, -0.5, 0.5)^\top$ for $K = 1, 2$ and $\boldsymbol{\phi}_{j0} = (-2.0, -0.5, 0.5)^\top$ for $K = 3$.
(M2) For $K = 1, 2$, $\boldsymbol{\phi}_{j0} = (-3.0, 0.1, 1.0)^\top$ for $j = 1, 2$ and $\boldsymbol{\phi}_{j0} = (-3.0, -0.1, 1.0)^\top$ for $j = 3, 4$. For $K = 3$, $\boldsymbol{\phi}_{j0} = (-7.0, 0.1, -1.0)^\top$ for $j = 1, 2$ and $\boldsymbol{\phi}_{j0} = (-7.0, -0.1, -1.0)^\top$ for $j = 3, 4$.

For $j = 1, \ldots, 4$, the approximately unconditional probabilities π_{ij} for four time points are about 75.3%, 56.7%, 42.6% and 32.2% for $K = 1$, 80.6%, 65.0%, 52.4% and 42.2% for $K = 2$, 85.1%, 72.3%, 61.4% and 52.2% for $K = 3$ under (M1); 90.8%, 82.4%, 74.5% and 67.4% for $K = 1$, 95.7%, 91.6%, 87.7% and 84.0% for $K = 2$, 95.0%, 90.3%, 86.1%, 82.1% for $K = 3$ under (M2). It should be pointed out that (M2) generates missing data mainly around one change point while (M1) generates relatively balanced missing data. For example, when the missing indicators are generated from (M2) under $K = 1$, it can be seen that the responses mainly have missing data around δ_1 such that the CC estimator may have a large bias for the estimation of α_0, α_1 and β_1, which is consistent with our simulation results in Table S7 in the Supplementary Material. Hence, even the unconditional dropout percentages of the four time points in (M2) are much smaller than those of (M1), and the CC estimator still has much larger biases. Moreover, when the third element of $\boldsymbol{\phi}_{j0}$, i.e., the coefficient of Y_{ij}, is positive, the parameter estimation results may become better if τ becomes smaller, since the larger Y_{ij} is more likely to be missing. When $K = 3$ and the missing indicators are generated from (M2), it can be seen that the coefficient of Y_{ij} is negative, and then the parameter estimation results may become better when τ becomes larger.

4.1. Number selection consistency of K

Two different scenarios are considered in the following simulations. Under Scenario 1, we consider two dropout settings (M1) and (M2) with four different errors, the true numbers $K = 1, 2, 3$ and show the rates of correct number selecting of kink points in Table 1 for $n = 200$ based on 1000 Monte Carlo replications. It can be seen that the CC method has comparable or better performance with our proposed three methods. The possible reason is that the complete case data have significant turning points in these cases and the CC estimator has smaller variances (see Tables 3 and S4 in the Supplementary Material). Hence CC method has better performance. Under Scenario 2, we consider $K = 3$ with $\boldsymbol{\phi}_{j0} = (-1.5, 0.1, 1.0)^\top$ for $j = 1$, 2 and $\boldsymbol{\phi}_{j0} = (-1.5, -0.1, 1.0)^\top$ for $j = 3$, 4, where the missing rates are extremely high when $X_{ij} \geq \delta_3$. The percentages of correctly selecting kink points based on $n = 200$, 300 and 500 are reported in Table 2. As shown in the table, the rates of the CC method are lower than the rates of our proposed three methods, especially when $n = 200$. In summary, our proposed methods work reasonably well in all the simulation studies.

4.2. Parameter estimation and coverage probability

After the selection of the number K, we evaluate the finite sample performance of parameter estimators to check the validity of Theorem 2.2. For $\tau = 0.25, 0.5, 0.75$ and $n = 500, 1000$, $\boldsymbol{\theta}_0 = (\alpha_0, \alpha_1, \beta_1, \ldots, \beta_K, \gamma, \delta_1, \ldots, \delta_K)^\top \in \mathbb{R}^{3+2K_0}$ and we compute (1) the simulated absolute bias, i.e., $\mathbf{AB} = T^{-1} \sum_{t=1}^{T} \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0\|_1$, and standard deviation, i.e., $\mathbf{SD} = \sum_{j=1}^{3+2K_0} \mathcal{SD}(\{\hat{\theta}_{tj}\}_{t=1}^{T})$; (2) the simulated mean square error, i.e., $\mathbf{MSE} = T^{-1} \sum_{t=1}^{T} \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0\|_2^2$, and the

Table 1. The percentages of correctly selecting $\widehat{K} = K$ under Scenario 1.

		$K=1$					$K=2$					$K=3$				
			MNAR					MNAR					MNAR			
π	τ	CC	Ind	CS	AR	Full	CC	Ind	CS	AR	Full	CC	Ind	CS	AR	Full
							Homoscedastic normal errors									
M1	0.25	0.998	0.995	0.995	0.997	1	1	0.996	0.997	0.999	1	1	0.999	1	1	1
	0.50	0.999	0.989	0.996	0.999	1	0.999	0.995	0.999	0.998	1	1	1	0.999	1	1
	0.75	0.999	0.985	0.996	0.993	1	1	0.997	0.999	0.999	1	1	0.998	0.999	0.999	1
M2	0.25	1	0.997	1	1	1	1	1	1	1	1	1	1	1	1	1
	0.50	1	0.998	1	1	1	1	1	1	1	1	1	1	1	1	1
	0.75	1	0.998	0.997	0.998	1	1	0.999	1	1	1	1	1	1	0.999	0.997
							Homoscedastic t errors									
M1	0.25	0.997	0.985	0.980	0.987	1	0.995	0.990	0.992	0.994	1	0.997	0.998	0.998	0.996	0.999
	0.50	1	0.994	0.997	0.996	1	1	0.995	0.999	0.998	1	0.999	1	1	0.999	1
	0.75	0.997	0.972	0.992	0.994	1	0.998	0.986	0.996	0.998	1	0.999	0.992	0.999	0.998	0.999
M2	0.25	1	0.999	0.998	0.998	1	0.999	0.999	0.999	0.999	0.999	1	1	1	0.999	0.999
	0.50	1	0.998	0.999	0.999	1	1	0.999	0.999	0.999	1	1	0.999	1	1	1
	0.75	0.998	0.991	0.996	0.997	1	1	0.997	0.998	0.999	1	1	1	1	1	1
							Heteroscedastic normal errors									
M1	0.25	0.998	0.996	0.993	0.992	1	1	0.998	1	0.999	1	0.999	1	1	1	1
	0.50	1	0.990	0.994	0.997	1	1	0.993	1	0.999	1	1	0.999	0.998	1	1
	0.75	0.997	0.974	0.986	0.994	1	1	0.986	0.998	0.998	1	1	0.998	0.999	1	0.999
M2	0.25	1	0.998	1	1	1	1	1	1	0.999	1	1	0.999	1	1	1
	0.50	1	0.996	0.998	0.999	1	1	1	1	1	1	1	0.999	1	1	1
	0.75	1	0.998	1	1	1	1	0.999	1	0.999	1	1	1	1	1	1
							Heteroscedastic t errors									
M1	0.25	0.998	0.986	0.989	0.986	0.998	0.999	0.990	0.996	0.996	0.999	0.997	0.995	0.998	0.995	0.998
	0.50	0.999	0.981	0.992	0.995	1	1	0.996	1	1	1	1	0.998	0.999	1	0.998
	0.75	0.997	0.961	0.991	0.990	1	1	0.979	0.994	0.996	1	0.999	0.995	0.999	0.999	0.999
M2	0.25	1	0.999	1	0.999	1	0.999	0.996	0.999	0.998	1	1	0.996	0.998	0.999	0.999
	0.50	0.999	0.995	0.999	0.997	1	1	0.998	0.998	0.997	1	1	1	1	1	0.999
	0.75	1	0.988	0.999	0.998	1	1	0.998	1	1	1	1	1	1	1	1

Table 2. The percentages of correctly selecting $\widehat{K} = K$ under Scenario 2.

		$n=200$					$n=300$					$n=400$				
			MNAR					MNAR					MNAR			
Error	τ	CC	Ind	CS	AR	Full	CC	Ind	CS	AR	Full	CC	Ind	CS	AR	Full
(a)	0.25	0.886	0.961	0.936	0.918	1	0.983	0.981	0.985	0.980	1	0.993	0.976	0.993	0.990	1
	0.50	0.898	0.954	0.954	0.944	1	0.982	0.989	0.991	0.990	1	0.996	0.984	0.997	0.995	1
	0.75	0.879	0.947	0.942	0.929	1	0.971	0.981	0.989	0.995	1	0.996	0.975	0.995	0.996	1
(b)	0.25	0.678	0.879	0.820	0.783	1	0.896	0.963	0.945	0.929	1	0.951	0.975	0.975	0.968	1
	0.50	0.750	0.913	0.867	0.842	1	0.938	0.978	0.974	0.972	1	0.990	0.977	0.989	0.990	1
	0.75	0.784	0.915	0.871	0.855	0.998	0.925	0.968	0.967	0.972	1	0.985	0.972	0.986	0.988	1
(c)	0.25	0.770	0.915	0.871	0.833	1	0.942	0.969	0.984	0.968	1	0.993	0.978	0.995	0.992	1
	0.50	0.821	0.922	0.892	0.865	1	0.952	0.977	0.980	0.977	1	0.988	0.982	0.995	0.993	1
	0.75	0.808	0.914	0.868	0.865	1	0.948	0.978	0.969	0.969	1	0.985	0.969	0.987	0.987	1
(d)	0.25	0.541	0.810	0.703	0.657	1	0.755	0.926	0.881	0.848	1	0.896	0.952	0.962	0.954	1
	0.50	0.621	0.855	0.780	0.740	1	0.834	0.946	0.930	0.914	0.999	0.950	0.954	0.972	0.971	1
	0.75	0.638	0.815	0.764	0.735	1	0.852	0.932	0.924	0.914	0.999	0.959	0.966	0.970	0.970	1

estimated standard error (**SE**) of each component of $\boldsymbol{\theta}$ based on the plug-in estimator of $(\boldsymbol{T}^{\top}\boldsymbol{\Xi}^{-1}\boldsymbol{T})^{-1}$; (3) the averaged 95% coverage probability, i.e., $\mathbf{CP} = (3+2K_0)^{-1}\sum_{j=1}^{3+2K_0}\mathcal{CP}(\{\hat{\theta}_{tj}\}_{t=1}^{T})$, based on 1000 replications, where $\mathcal{CP}(\{\hat{\theta}_{tj}\}_{t=1}^{T})$ is the coverage probability of $\hat{\theta}_j$. To save space, Table 3 only reports the results under $K=2$ and $n=1000$ and other simulation results are given in Tables S1–S5 in the Supplementary Material.

There are some conclusions that can be drawn from these simulation results. (i) **AB**. For both dropout settings (M1) and (M2), our proposed MNAR estimators have negligible biases, while the CC estimator has large ABs especially under (M2). Further, it can be seen that our proposed estimators are robust since they are less sensitive to the four different error distributions. In addition, in Tables 3 and S4, the ABs become larger when τ increases, while in Table S5 in the Supplementary Material, the ABs under (M2) are larger when $\tau = 0.25$, which is consistent with our discussion above. (ii) **SD** and **SE**. The SDs of the MNAR_{AR1} and MNAR_{CS} estimators are smaller than the values of the MNAR_{Ind} estimator. This finding shows that the estimation efficiency can be improved when the informative working correlation structures are considered, compared with using the independent structure. On the other hand,

Table 3. The simulated AB, SD, SE, MSE and CP under $K = 2$ and $n = 1000$.

		$\tau = 0.25$					$\tau = 0.5$					$\tau = 0.75$				
π	method	AB	SD	SE	MSE	CP	AB	SD	SE	MSE	CP	AB	SD	SE	MSE	CP
							Homoscedastic normal errors									
M1	CC	0.273	0.314	0.320	0.046	0.882	0.298	0.307	0.303	0.053	0.855	0.336	0.349	0.327	0.068	0.832
	MNAR_{Ind}	0.009	0.415	0.419	0.032	0.949	0.010	0.428	0.439	0.034	0.955	0.016	0.503	0.500	0.047	0.948
	MNAR_{CS}	0.013	0.348	0.353	0.022	0.948	0.020	0.357	0.362	0.024	0.949	0.055	0.435	0.406	0.037	0.931
	MNAR_{AR}	0.022	0.345	0.349	0.022	0.948	0.037	0.357	0.354	0.024	0.941	0.084	0.421	0.390	0.036	0.926
	Full	0.012	0.226	0.226	0.009	0.946	0.009	0.215	0.210	0.009	0.946	0.009	0.228	0.226	0.009	0.947
M2	CC	0.240	0.244	0.241	0.026	0.837	0.268	0.227	0.225	0.029	0.785	0.329	0.249	0.242	0.041	0.742
	MNAR_{Ind}	0.012	0.298	0.309	0.016	0.954	0.006	0.299	0.309	0.016	0.955	0.014	0.336	0.348	0.021	0.953
	MNAR_{CS}	0.018	0.248	0.251	0.011	0.949	0.018	0.239	0.247	0.010	0.956	0.043	0.272	0.279	0.014	0.948
	MNAR_{AR}	0.023	0.259	0.260	0.012	0.950	0.031	0.250	0.254	0.012	0.946	0.065	0.284	0.281	0.016	0.936
	Full	0.009	0.232	0.226	0.010	0.943	0.013	0.211	0.210	0.008	0.944	0.007	0.229	0.226	0.009	0.947
							Homoscedastic t errors									
M1	CC	0.330	0.365	0.367	0.066	0.875	0.356	0.339	0.337	0.072	0.843	0.415	0.392	0.364	0.099	0.813
	MNAR_{Ind}	0.012	0.494	0.493	0.046	0.958	0.006	0.510	0.509	0.049	0.955	0.008	0.679	0.623	0.089	0.951
	MNAR_{CS}	0.012	0.401	0.403	0.030	0.951	0.032	0.407	0.408	0.031	0.950	0.076	0.494	0.468	0.049	0.931
	MNAR_{AR}	0.024	0.390	0.396	0.028	0.956	0.058	0.400	0.395	0.031	0.942	0.119	0.489	0.442	0.051	0.917
	Full	0.016	0.261	0.257	0.012	0.945	0.005	0.234	0.234	0.010	0.951	0.011	0.261	0.258	0.012	0.946
M2	CC	0.292	0.282	0.276	0.037	0.815	0.354	0.250	0.250	0.046	0.724	0.443	0.282	0.271	0.068	0.663
	MNAR_{Ind}	0.019	0.368	0.368	0.024	0.954	0.024	0.363	0.378	0.024	0.958	0.027	0.455	0.452	0.037	0.955
	MNAR_{CS}	0.026	0.289	0.293	0.015	0.948	0.053	0.275	0.287	0.015	0.954	0.093	0.329	0.335	0.022	0.941
	MNAR_{AR}	0.036	0.302	0.300	0.017	0.941	0.070	0.282	0.290	0.016	0.948	0.124	0.338	0.330	0.025	0.924
	Full	0.014	0.264	0.258	0.013	0.941	0.004	0.232	0.234	0.010	0.949	0.010	0.265	0.258	0.013	0.945
							Heteroscedastic normal errors									
M1	CC	0.331	0.347	0.350	0.059	0.859	0.364	0.334	0.332	0.066	0.828	0.395	0.374	0.358	0.082	0.817
	MNAR_{Ind}	0.006	0.460	0.466	0.039	0.952	0.014	0.494	0.493	0.046	0.956	0.009	0.573	0.574	0.061	0.957
	MNAR_{CS}	0.015	0.384	0.392	0.027	0.952	0.034	0.399	0.403	0.030	0.949	0.055	0.477	0.457	0.043	0.935
	MNAR_{AR}	0.028	0.389	0.387	0.028	0.942	0.061	0.392	0.393	0.030	0.943	0.088	0.470	0.437	0.044	0.926
	Full	0.012	0.247	0.248	0.011	0.953	0.005	0.230	0.231	0.009	0.950	0.011	0.248	0.249	0.011	0.947
M2	CC	0.298	0.272	0.265	0.035	0.799	0.354	0.250	0.247	0.042	0.724	0.433	0.274	0.265	0.061	0.654
	MNAR_{Ind}	0.019	0.336	0.346	0.020	0.954	0.018	0.337	0.354	0.021	0.961	0.023	0.394	0.405	0.029	0.956
	MNAR_{CS}	0.024	0.282	0.280	0.014	0.944	0.036	0.267	0.280	0.013	0.957	0.074	0.313	0.319	0.019	0.946
	MNAR_{AR}	0.030	0.293	0.289	0.016	0.941	0.053	0.280	0.286	0.015	0.949	0.100	0.321	0.319	0.021	0.935
	Full	0.009	0.262	0.249	0.012	0.937	0.006	0.230	0.231	0.009	0.946	0.010	0.255	0.248	0.012	0.944
							Heteroscedastic t errors									
M1	CC	0.405	0.400	0.402	0.086	0.845	0.439	0.374	0.369	0.093	0.810	0.503	0.431	0.398	0.124	0.775
	MNAR_{Ind}	0.014	0.567	0.553	0.060	0.954	0.028	0.624	0.602	0.074	0.957	0.033	0.780	0.713	0.113	0.944
	MNAR_{CS}	0.025	0.444	0.451	0.037	0.948	0.061	0.458	0.460	0.040	0.947	0.110	0.567	0.527	0.064	0.922
	MNAR_{AR}	0.039	0.440	0.441	0.036	0.948	0.085	0.447	0.442	0.039	0.938	0.156	0.545	0.493	0.063	0.909
	Full	0.022	0.284	0.283	0.015	0.944	0.003	0.265	0.258	0.013	0.941	0.013	0.287	0.283	0.015	0.946
M2	CC	0.359	0.306	0.303	0.049	0.785	0.443	0.274	0.275	0.063	0.660	0.559	0.305	0.296	0.096	0.575
	MNAR_{Ind}	0.016	0.393	0.417	0.028	0.961	0.030	0.422	0.437	0.032	0.958	0.065	0.511	0.517	0.049	0.954
	MNAR_{CS}	0.029	0.319	0.330	0.019	0.954	0.069	0.311	0.327	0.018	0.950	0.132	0.371	0.379	0.029	0.933
	MNAR_{AR}	0.042	0.328	0.335	0.020	0.950	0.091	0.319	0.327	0.020	0.944	0.166	0.375	0.369	0.032	0.916
	Full	0.018	0.287	0.283	0.015	0.947	0.008	0.257	0.258	0.012	0.948	0.008	0.288	0.283	0.015	0.944

it can be seen that the SDs of the CC estimator are smaller than the proposed estimators since the CC method does not need to estimate $\boldsymbol{\phi}_0$. In addition, all SEs are quite close to the SDs, which shows our plug-in estimator works well. (iii) **MSE**. The MSEs of the proposed MNAR_{AR1} and MNAR_{CS} are always smaller than the MSEs of the CC and MNAR_{Ind} estimators. Although the CS structure is the true structure, our proposed MNAR_{AR1} estimator still has comparable performance. This indicates that we can obtain a more efficient estimator by assuming an informative working correlation structure that may not be correct, rather than ignoring the unknown correlation structure. (iv) **CP**. The coverage rates of 95% Wald confidence intervals of our proposed three MNAR estimators are close to the nominal level 95%, while the CPs are slightly low when $\tau = 0.75$. Moreover, it can be seen that our CPs increase slightly when n increases. However, for the CC estimator, it can be seen that its CPs decrease when n increases, since some components of the CC estimator have large biases and thus do not obey the asymptotic distribution in D. Li et al. (2022). Moreover, the CC estimator always has smaller CPs compared with the proposed estimators, especially under (M2).

4.3. Power analysis

We compare the performance of our proposed testing procedure in Algorithm 2 with the CC testing procedure for the existence of the kink points based on 600 repetitions. To be specific, we consider one single kink point $K = 1$, $\tau = 0.25$, 0.5 and 0.75, $n = 500$ and 1000, $B = 200$; the missing indicators r_{ij} are generated by

Table 4. The 95% CPs under $K = 2$ and $n = 1000$.

		$\tau = 0.25$					$\tau = 0.5$					$\tau = 0.75$				
			MNAR					MNAR					MNAR			
π		CC	Ind	CS	AR	Full	CC	Ind	CS	AR	Full	CC	Ind	CS	AR	Full
							Homoscedastic normal errors									
M1	α_0	0.657	0.954	0.954	0.959	0.958	0.548	0.964	0.947	0.936	0.943	0.497	0.955	0.925	0.910	0.950
	γ	0.919	0.949	0.941	0.934	0.933	0.919	0.939	0.947	0.948	0.949	0.911	0.952	0.923	0.916	0.948
	α_1	0.968	0.946	0.957	0.953	0.958	0.946	0.964	0.941	0.944	0.942	0.935	0.955	0.934	0.931	0.952
	β_1	0.886	0.955	0.944	0.954	0.949	0.846	0.962	0.950	0.939	0.944	0.831	0.948	0.939	0.936	0.945
	β_2	0.864	0.943	0.952	0.947	0.936	0.846	0.953	0.958	0.941	0.944	0.824	0.938	0.926	0.924	0.944
	δ_1	0.924	0.947	0.943	0.945	0.953	0.925	0.949	0.947	0.929	0.950	0.902	0.943	0.930	0.925	0.949
	δ_2	0.953	0.949	0.943	0.944	0.934	0.953	0.957	0.955	0.952	0.947	0.922	0.946	0.941	0.937	0.939
M2	α_0	0.735	0.960	0.960	0.951	0.945	0.624	0.962	0.965	0.947	0.954	0.534	0.954	0.932	0.921	0.941
	γ	0.903	0.941	0.954	0.945	0.956	0.868	0.958	0.953	0.941	0.946	0.852	0.940	0.950	0.935	0.935
	α_1	0.810	0.954	0.943	0.939	0.933	0.742	0.960	0.967	0.956	0.954	0.714	0.953	0.949	0.934	0.947
	β_1	0.756	0.962	0.953	0.959	0.938	0.666	0.953	0.955	0.956	0.938	0.596	0.973	0.958	0.954	0.958
	β_2	0.773	0.968	0.946	0.945	0.946	0.719	0.957	0.954	0.943	0.945	0.634	0.954	0.951	0.939	0.952
	δ_1	0.932	0.953	0.945	0.956	0.938	0.933	0.947	0.953	0.942	0.946	0.921	0.948	0.952	0.941	0.954
	δ_2	0.947	0.943	0.944	0.956	0.945	0.943	0.945	0.945	0.938	0.925	0.944	0.949	0.944	0.929	0.945
							Homoscedastic *t* errors									
M1	α_0	0.630	0.964	0.954	0.963	0.947	0.507	0.959	0.955	0.946	0.957	0.432	0.951	0.909	0.868	0.941
	γ	0.910	0.961	0.953	0.949	0.942	0.924	0.949	0.951	0.944	0.960	0.904	0.955	0.940	0.925	0.966
	α_1	0.957	0.964	0.954	0.963	0.949	0.950	0.949	0.944	0.942	0.947	0.919	0.951	0.926	0.915	0.932
	β_1	0.886	0.962	0.955	0.959	0.948	0.836	0.962	0.947	0.932	0.945	0.807	0.949	0.929	0.916	0.944
	β_2	0.876	0.948	0.939	0.948	0.947	0.838	0.966	0.962	0.945	0.944	0.803	0.951	0.937	0.931	0.962
	δ_1	0.930	0.956	0.952	0.955	0.939	0.905	0.952	0.942	0.941	0.955	0.901	0.958	0.934	0.927	0.942
	δ_2	0.939	0.951	0.951	0.956	0.940	0.939	0.947	0.949	0.947	0.949	0.928	0.943	0.945	0.935	0.933
M2	α_0	0.684	0.951	0.955	0.948	0.944	0.463	0.962	0.947	0.942	0.952	0.356	0.945	0.919	0.886	0.945
	γ	0.881	0.950	0.947	0.939	0.948	0.885	0.953	0.952	0.934	0.934	0.844	0.958	0.948	0.950	0.953
	α_1	0.783	0.963	0.946	0.950	0.931	0.657	0.963	0.961	0.956	0.955	0.587	0.957	0.946	0.922	0.934
	β_1	0.707	0.955	0.950	0.943	0.936	0.579	0.957	0.957	0.948	0.957	0.458	0.960	0.943	0.919	0.945
	β_2	0.770	0.965	0.957	0.936	0.948	0.603	0.960	0.949	0.940	0.955	0.534	0.959	0.942	0.920	0.959
	δ_1	0.928	0.944	0.932	0.931	0.934	0.927	0.964	0.961	0.954	0.937	0.921	0.953	0.937	0.936	0.935
	δ_2	0.955	0.951	0.952	0.941	0.945	0.953	0.950	0.949	0.964	0.950	0.939	0.950	0.949	0.936	0.941
							Heteroscedastic normal errors									
M1	α_0	0.638	0.966	0.957	0.942	0.953	0.519	0.953	0.950	0.941	0.961	0.487	0.955	0.924	0.903	0.951
	γ	0.845	0.959	0.953	0.936	0.968	0.830	0.952	0.931	0.931	0.935	0.808	0.952	0.925	0.906	0.955
	α_1	0.946	0.947	0.951	0.948	0.957	0.950	0.956	0.948	0.951	0.948	0.938	0.957	0.941	0.938	0.943
	β_1	0.867	0.950	0.957	0.951	0.947	0.826	0.969	0.959	0.950	0.956	0.832	0.962	0.937	0.935	0.943
	β_2	0.860	0.951	0.946	0.947	0.948	0.810	0.959	0.961	0.932	0.944	0.815	0.948	0.933	0.932	0.949
	δ_1	0.915	0.949	0.950	0.935	0.947	0.917	0.946	0.946	0.943	0.953	0.903	0.966	0.936	0.930	0.943
	δ_2	0.945	0.941	0.947	0.935	0.948	0.942	0.955	0.951	0.952	0.951	0.937	0.959	0.946	0.941	0.943
M2	α_0	0.685	0.952	0.943	0.937	0.933	0.540	0.966	0.958	0.948	0.945	0.413	0.950	0.934	0.926	0.944
	γ	0.826	0.954	0.942	0.937	0.929	0.751	0.962	0.948	0.943	0.946	0.709	0.959	0.941	0.923	0.941
	α_1	0.762	0.951	0.938	0.939	0.928	0.665	0.960	0.959	0.943	0.937	0.587	0.955	0.943	0.940	0.945
	β_1	0.717	0.963	0.947	0.945	0.943	0.575	0.970	0.962	0.959	0.950	0.460	0.961	0.962	0.947	0.957
	β_2	0.723	0.959	0.950	0.948	0.951	0.641	0.957	0.954	0.949	0.942	0.526	0.951	0.943	0.929	0.935
	δ_1	0.938	0.954	0.951	0.939	0.936	0.940	0.958	0.960	0.943	0.944	0.929	0.955	0.944	0.939	0.941
	δ_2	0.945	0.945	0.939	0.945	0.939	0.955	0.957	0.959	0.957	0.960	0.952	0.961	0.954	0.943	0.947
							Heteroscedastic *t* errors									
M1	α_0	0.584	0.954	0.951	0.942	0.940	0.465	0.963	0.942	0.926	0.940	0.396	0.939	0.913	0.880	0.942
	γ	0.842	0.952	0.946	0.945	0.942	0.804	0.948	0.936	0.930	0.949	0.774	0.950	0.914	0.897	0.954
	α_1	0.951	0.949	0.945	0.947	0.940	0.949	0.962	0.950	0.947	0.943	0.922	0.941	0.914	0.905	0.939
	β_1	0.837	0.960	0.954	0.959	0.946	0.786	0.960	0.946	0.938	0.940	0.759	0.947	0.931	0.923	0.945
	β_2	0.852	0.961	0.951	0.946	0.944	0.804	0.956	0.951	0.938	0.940	0.750	0.945	0.913	0.911	0.948
	δ_1	0.910	0.943	0.946	0.947	0.943	0.929	0.962	0.955	0.944	0.937	0.892	0.941	0.924	0.907	0.953
	δ_2	0.937	0.957	0.945	0.947	0.950	0.933	0.950	0.949	0.944	0.938	0.930	0.946	0.948	0.939	0.942
M2	α_0	0.648	0.967	0.954	0.955	0.947	0.410	0.951	0.956	0.941	0.948	0.279	0.950	0.911	0.896	0.945
	γ	0.826	0.953	0.949	0.952	0.944	0.712	0.962	0.952	0.948	0.954	0.604	0.957	0.926	0.910	0.944
	α_1	0.744	0.966	0.960	0.956	0.949	0.599	0.954	0.961	0.947	0.948	0.509	0.953	0.937	0.916	0.946
	β_1	0.682	0.957	0.954	0.939	0.946	0.486	0.955	0.945	0.941	0.947	0.363	0.949	0.939	0.912	0.953
	β_2	0.701	0.965	0.951	0.954	0.949	0.522	0.969	0.944	0.940	0.938	0.420	0.958	0.933	0.914	0.943
	δ_1	0.942	0.966	0.963	0.945	0.954	0.941	0.953	0.945	0.945	0.953	0.921	0.955	0.944	0.936	0.938
	δ_2	0.952	0.951	0.950	0.948	0.942	0.952	0.959	0.948	0.948	0.946	0.929	0.958	0.940	0.925	0.942

$\boldsymbol{\phi}_{j0} = (-8.0, 0.1, 1.0)^\top$ for $j = 1, 2$, $\boldsymbol{\phi}_{j0} = (-8.0, -0.1, 1.0)^\top$ for $j = 3, 4$; $\beta_1 = 0, 0.1$ and 0.15. The Type I error and local power results under $K = 2$ and $n = 1000$ are shown in Tables 4 and 5 and other simulation results are given in Tables S6–S10 in the Supplementary Material.

We have the following findings. (i) When $\beta_1 = 0$, i.e., under the null hypothesis, our proposed MNAR test has satisfactory Type I errors close to the nominal significance level 5%; the CC test has much larger errors, especially when $n = 1000$. Moreover, when n increases, it can be seen that the Type I errors of the CC test become larger while

Table 5. Power results with vary β_1 under four different errors with $n = 500, 1000$ and $\tau = 0.25, 0.5, 0.75$.

		$n = 500$						$n = 1000$					
		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
Errors	β_1	CC	MNAR	CC	MNAR	CC	MNAR	CC	MNAR	CC	MNAR	CC	MNAR
(a)	0	0.207	0.040	0.255	0.048	0.292	0.030	0.330	0.033	0.442	0.027	0.560	0.037
	0.1	0.242	0.478	0.245	0.413	0.168	0.300	0.532	0.860	0.450	0.835	0.317	0.693
	0.15	0.698	0.788	0.698	0.778	0.567	0.628	0.938	0.967	0.923	0.955	0.868	0.910
(b)	0	0.188	0.028	0.312	0.053	0.352	0.050	0.345	0.030	0.562	0.030	0.577	0.030
	0.1	0.175	0.313	0.150	0.242	0.098	0.168	0.283	0.688	0.225	0.602	0.165	0.417
	0.15	0.497	0.633	0.520	0.583	0.388	0.432	0.795	0.902	0.773	0.835	0.625	0.747
(c)	0	0.207	0.027	0.300	0.037	0.328	0.050	0.418	0.027	0.547	0.032	0.623	0.025
	0.1	0.162	0.343	0.135	0.273	0.090	0.190	0.327	0.730	0.218	0.665	0.153	0.503
	0.15	0.548	0.682	0.528	0.653	0.402	0.465	0.842	0.923	0.815	0.910	0.608	0.818
(d)	0	0.205	0.027	0.332	0.038	0.385	0.052	0.405	0.030	0.610	0.025	0.640	0.043
	0.1	0.080	0.222	0.078	0.198	0.057	0.100	0.133	0.540	0.122	0.457	0.073	0.248
	0.15	0.363	0.542	0.318	0.460	0.225	0.303	0.608	0.860	0.562	0.778	0.378	0.620

Table 6. Parameter estimates and test results for the NGHS data.

Method		$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
CC	P-values	0.000	0.000	0.000
	$\widehat{K}$	1	1	1
	$\hat{\alpha}_0$	68.665(2.910)	79.884(1.976)	84.452(2.105)
	$\hat{\gamma}$	0.214(0.091)	0.266(0.093)	0.230(0.101)
	$\hat{\alpha}_1$	1.464(0.173)	0.975(0.116)	0.924(0.118)
	$\hat{\beta}_1$	−1.176(0.183)	−0.850(0.143)	−0.842(0.148)
	$\hat{\delta}_1$	20.938(0.516)	24.966(0.879)	25.782(1.081)
	Wald	[−19.927,21.949]	[23.243,26.689]	[23.663,27.901]
MNAR	P-values	0.000	0.000	0.000
	$\widehat{K}$	1	1	1
	$\hat{\alpha}_0$	73.977(2.203)	78.968(2.525)	83.979(3.367)
	$\hat{\gamma}$	0.443(0.111)	0.415(0.125)	0.302(0.156)
	$\hat{\alpha}_1$	1.039(0.124)	0.974(0.134)	0.971(0.158)
	$\hat{\beta}_1$	−1.004(0.168)	−0.999(0.176)	−1.040(0.209)
	$\hat{\delta}_1$	24.968(0.878)	26.199(1.070)	26.636(1.193)
	Wald	[23.247,26.689]	[24.102,28.300]	[24.298,28.974]

the proposed test has stable results. The main reason is that the CC estimator of $\boldsymbol{\alpha}$ is not consistent, which has been shown in D. Li and Wang (2022), such that the CC test would have big T_n and tend to reject the null hypothesis. (ii) As β_1 increases, i.e., under the local hypothesis, the proposed MNAR test has higher powers than those of the CC test. When n increases, all powers become larger. When τ increases, all powers decrease, since the larger Y_{ij} is more likely to be missing.

5. Analysis of blood pressure and body mass index

In this section, we analyse the longitudinal NGHS data from the website https://biolincc.nhlbi.nih.gov/ to evaluate our proposed method. The covariates X_{ij} and Z_{ij} are the BMI and age respectively, and the response Y_{ij} is the blood pressure (BP). The data were collected from 400 subjects whose ages range from 9 to 19 and $\max_i m_i = 10$ for $i = 1, \ldots, 400$. As we mentioned in Section 1, the nonignorable dropout rates of the follow-up times are 91.5%, 84.1%, 73.7%, 64.9%, 58.9%, 52.6%, 45.6%, 39.3%, 37.0% and 33.0%, respectively.

We compute the CC estimator and the proposed MNAR estimators using the CS structure for $\tau = 0.25, 0.5, 0.75$. The results of P-values for the kink points detection, the estimated number of kink points, the coefficients estimates, their standard errors and the confidence intervals for kink locations are summarized in Table 6. It can be seen that the CC estimator has smaller variances compared with the proposed MNAR estimator in general, which is in accord with simulation results. However, $\hat{\delta}_1$ obtained by the CC method is far from that of the proposed MNAR method when $\tau = 0.25$. Moreover, it can be seen that the proposed MNAR estimator under $\tau = 0.75$ does not perform well, compared with the MNAR estimator under $\tau = 0.25$ and 0.5, due to the missing mechanism. In addition, Figures 1–2 show scatter plots between the BP and BMI with the fitted curves at different levels τ based on the CC and MNAR methods, respectively.

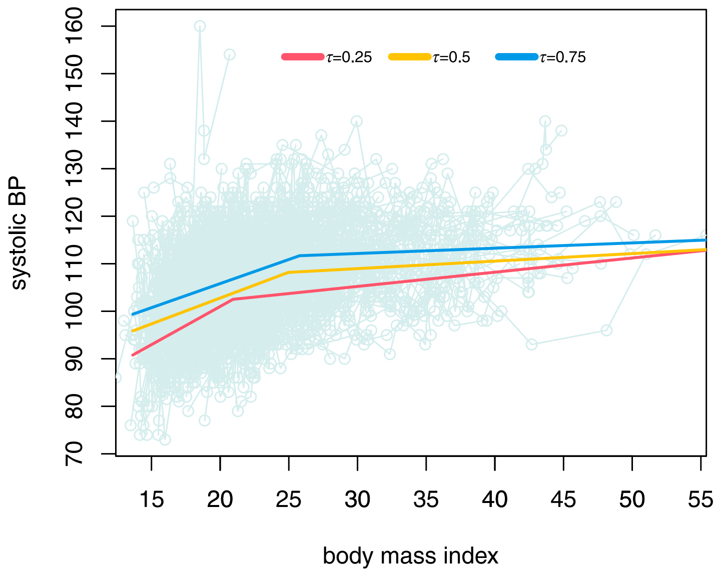

Figure 1. The fitted expectile curves using CC method at $\tau = 0.25$, 0.5 and 0.75 for BMI against systolic BP.

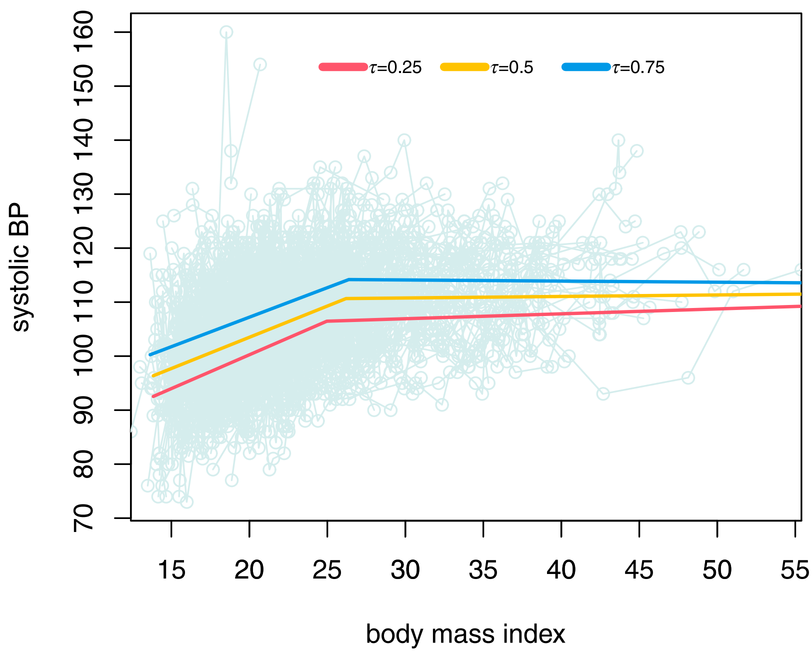

Figure 2. The fitted expectile curves using MNAR method at $\tau = 0.25$, 0.5 and 0.75 for BMI against systolic BP.

6. Conclusion

In this article, we develop a longitudinal multiple kink expectile regression model with the unknown number of kink points and nonignorable dropout. The selection consistency and the asymptotic properties of regression coefficients and kink points are derived. In order to test the existence of kink effects at a given expectile with nonignorable dropout, a weighted cumulative sum type statistic is proposed and we obtain its limiting distributions. Simulation studies and real data analysis show that our proposed estimators and proposed test have good performance.

There are some interesting topics to further study. First, we only consider the expectile regression model and this can be extended to other models, such as censored models, generalized linear models and so on. Second, in this article, we establish the theoretical properties when the number of kink points is true. Hence, extending the theoretical properties of estimators with the misspecified number of kink points is relegated to the further study. Third, it is interesting to study how to test if kink locations depend on the expectile levels. Fourth, the computation time of Algorithm 2 is long due to the bootstrap procedure and it is our further work to reduce the computational complexity of our proposed method.

Acknowledgments

The authors are grateful to the Editor, an Associate Editor and one anonymous referee for their insightful comments and suggestions on this article, which have led to significant improvements.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Our research was supported by the Fundamental Research Funds for the Central Universities and the National Natural Science Foundation of China [Grant Numbers 12271272 and 12001295].

ORCID

Lei Wang http://orcid.org/0000-0003-2530-883X

References

Aigner, D., Amemiya, T., & Poirier, D. (1976). On the estimation of production frontiers: Maximum likelihood estimation of the parameters of a discontinuous density function. *International Economic Review*, *17*(2), 377–396. https://doi.org/10.2307/2525708

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*(1), 1–22. https://doi.org/10.1002/jae.v18:1

Chen, J., Huang, Y., & Tang, N. (2022). Bayesian change-point joint models for multivariate longitudinal and time-to-event data. *Statistics in Biopharmaceutical Research*, *14*(2), 227–241. https://doi.org/10.1080/19466315.2020.1837234

Fong, Y. (2019). Fast bootstrap confidence intervals for continuous threshold linear regression. *Journal of Computational and Graphical Statistics*, *28*(2), 1–8. https://doi.org/10.1080/10618600.2018.1537927

Hansen, B. (2017). Regression kink with an unknown threshold. *Journal of Business and Economic Statistics*, *35*(2), 228–240. https://doi.org/10.1080/07350015.2015.1073595

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029–1054. https://doi.org/10.2307/1912775

Koenker, R., & Bassett, J. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50. https://doi.org/10.2307/1913643

Lee, S., Seo, M., & Shin, Y. (2011). Testing for threshold effects in regression models. *Journal of the American Statistical Association*, *106*(493), 220–231. https://doi.org/10.1198/jasa.2011.tm09800

Lee, S., Seo, M., & Shin, Y. (2016). The lasso for high dimensional regression with a possible change point. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *78*(1), 193–210. https://doi.org/10.1111/rssb.12108

Li, C., Wei, Y., Chappell, R., & He, X. (2011). Bent line quantile regression with application to an allometric study of land mammals' speed and mass. *Biometrics*, *67*(1), 242–249. https://doi.org/10.1111/j.1541-0420.2010.01436.x

Li, D., & Wang, L. (2022). Improved *k*th power expectile regression with nonignorable dropouts. *Journal of Applied Statistics*, *49*(11), 2767–2788. https://doi.org/10.1080/02664763.2021.1919606

Li, D., Wang, L., & Zhao, W. (2022). Estimation and inference for multikink expectile regression with longitudinal data. *Statistics in Medicine*, *41*(7), 1296–1313. https://doi.org/10.1002/sim.v41.7

Liang, K., & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Matteo, B., Haeran, C., & Piotr, F. (2018). Simultaneous multiple change-point and factor analysis for high dimensional time series. *Journal of Econometrics*, *206*(1), 187–225. https://doi.org/10.1016/j.jeconom.2018.05.003

Miao, W., & Tchetgen Tchetgn, E. (2018). Identification and inference with nonignorable missing covariate data. *Statistica Sinica*, *28*(4), 2049–2067.

Newey, W., & Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica*, *55*(4), 819–847. https://doi.org/10.2307/1911031

Oka, T., & Qu, Z. (2011). Estimating structural changes in regression quantiles. *Journal of Econometrics*, *162*(2), 248–267. https://doi.org/10.1016/j.jeconom.2011.01.005

Perron, P., & Qu, Z. (2006). Estimating restricted structural change models. *Journal of Econometrics*, *134*(2), 373–399. https://doi.org/10.1016/j.jeconom.2005.06.030

Qu, A., Lindsay, B., & Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika*, *87*(4), 823–836. https://doi.org/10.1093/biomet/87.4.823

Shao, J., & Wang, L. (2016). Semiparametric inverse propensity weighting for nonignorable missing data. *Biometrika*, *103*(1), 175–187. https://doi.org/10.1093/biomet/asv071

Wan, C., Zhong, W., Zhang, W., & Zou, C. (2023). Multikink quantile regression for longitudinal data with application to progesterone data analysis. *Biometrics*, *79*(2), 747–760.

Wang, L., Qi, C., & Shao, J. (2019). Model-assisted regression estimators for longitudinal data with nonignorable dropout. *International Statistical Review*, *87*(S1), S121–S138. https://doi.org/10.1111/insr.v87.S1

Wang, S., Shao, J., & Kim, J. (2014). An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statistica Sinica*, *24*(3), 1097–1116.

Zhang, F., & Li, Q. (2017). A continuous threshold expectile model. *Computational Statistics and Data Analysis*, *116*, 49–66. https://doi.org/10.1016/j.csda.2017.07.005
Zhang, L., Wang, J., & Zhu, Z. (2014). Testing for change points due to a covariate threshold in quantile regression. *Statistica Sinica*, *24*(4), 1859–1877.
Zhong, W., Wan, C., & Zhang, W. (2022). Estimation and inference for multi-kink quantile regression. *Journal of Business and Economic Statistics*, *40*(3), 1123–1139. https://doi.org/10.1080/07350015.2021.1901720
Zhou, H., & Liang, H. (2021). Change point estimation in regression model with response missing at random. *Communications in Statistics-Theory and Methods*, *51*(20), 7101–7119. https://doi.org/10.1080/03610926.2020.1871017