Published in final edited form as:

Biometrics. 2017 June ; 73(2): 452–462. doi:10.1111/biom.12623.

Model-robust Inference for Continuous Threshold Regression Models

Youyi Fong*, Chongzhi Di, Ying Huang, and Peter B. Gilbert

Fred Hutchinson Cancer Research Center, Seattle WA 98109, USA

Summary

We study threshold regression models that allow the relationship between the outcome and a covariate of interest to change across a threshold value in the covariate. In particular we focus on continuous threshold models, which experience no jump at the threshold. Continuous threshold regression functions can provide a useful summary of the association between outcome and the covariate of interest, because they offer a balance between flexibility and simplicity. Motivated by collaborative works in studying immune response biomarkers of transmission of infectious diseases, we study estimation of continuous threshold models in this paper with particular attention to inference under model misspecification. We derive the limiting distribution of the maximum likelihood estimator, and propose both Wald and test-inversion confidence intervals. We evaluate finite sample performance of our methods, compare them with bootstrap confidence intervals, and provide guidelines for practitioners to choose the most appropriate method in real data analysis. We illustrate the application of our methods with examples from the HIV-1 immune correlates studies.

Keywords

profile likelihood ratio under model misspecification; regression kink; mother-to-child transmission of HIV-1; RV144 immune correlates studies

1. Introduction

Change point problems have arisen in many substantive areas, e.g. Alzheimer's disease (Hall et al., 2000), heart diseases (e.g. Goetghebuer and Pocock, 1995; Pastor and Guallar, 1998; Carmody et al., 2014) and diabetes (e.g. Pastor-Barriuso et al., 2003). Our interests in change point analysis stem from collaborative works in infectious disease vaccine development. An important aspect of modern vaccine development is to identify human immune response biomarkers that are associated with the risk of infection/transmission or vaccine efficacy (Gilbert et al., 2008). The relationship between the risk of infection/transmission and immune biomarkers is often thought to be nonlinear (Tomaras and Haynes,

* yfong@fhcrc.org.

Supplementary Materials

Web Appendices, Tables, and Figures referenced in Sections 2–6 are available with this paper at the *Biometrics* web site on Wiley Online Library. Also available is a zip file containing the R code for the numerical studies. A R package implementing the proposed methods, *chngpt*, is available from the Comprehensive R Archive Network.

2014). In particular, it has been hypothesized that only an immune response above a certain quality and quantity threshold can result in protection from HIV-1 infection or transmission.

Change point problems cover a large number of interesting situations. A useful distinction (Hansen, 2000) can be made between two subclasses of change point models: change-point models and threshold models. In a change-point model (e.g. Muller, 1992; Loader et al., 1996; Csörgö and Horváth, 1997; Antoch et al., 2007; Gijbels, 2008; Zeileis et al., 2002) the samples have a natural order based on, for example, time or position on a chromosome. Change-point models are often used to detect structural changes along the natural axis. In a threshold model (e.g. Chan and Tsay, 1998; Hansen, 2000; Banerjee and McKeague, 2007; Fong et al., 2016) there is typically not a natural ordering of samples or the ordering is not essential to the problem, e.g. when there is only a single covariate. The goal for using threshold models is often to approximate a nonlinear relationship between the outcome and a covariate of interest, which we will refer to as change point covariate. Compared to nonparametric smoothing methods of modeling nonlinearity, threshold models have the advantage of being simple and easy to interpret. Our focus in this paper is on the estimation of threshold models; and because we view threshold models as a useful summary of the true relationship between outcome and change point covariate, particular emphasis is given to making valid inference when the true data-generating model differs from the assumed threshold models.

Threshold regression models come in many different flavors (Fong et al., 2016). We focus on two continuous threshold logistic regression models:

$$\text{segmented: } \text{logit } \{\Pr(Y=1|x, \mathbf{z})\} = \beta_1 + \boldsymbol{\beta}_z^T \mathbf{z} + \beta_x x + \beta_{(x-e)_+} (x - e)_+ \quad (1)$$

$$\text{hinge: } \text{logit } \{\Pr(Y=1|x, \mathbf{z})\} = \beta_1 + \boldsymbol{\beta}_z^T \mathbf{z} + \beta_{(x-e)_+} (x - e)_+. \quad (2)$$

Here x is a single covariate, e is the threshold parameter, $(x - e)_+$ denotes $(x - e) I(x > e)$, where $I(\cdot)$ is an indicator function, and $\mathbf{z}$ is a vector of additional covariates. Both models are continuous at the change point, and the hinge model is a special case of the segmented model by assuming zero slope before the threshold.

Much work has been done regarding hypothesis testing in threshold models (e.g. Davies, 1987; Ulm, 1991; Koziol and Wu, 1996; Xu and Adak, 2002; Mazumdar et al., 2003; Pastor-Barriuso et al., 2003; Antoch et al., 2004; Zheng and Chen, 2005; Vexler and Gurevich, 2006, 2009; Lee et al., 2011). Our goal in this paper is to develop model-robust methods for constructing confidence intervals for coefficients in continuous threshold models, which should provide proper coverage even when the true underlying model differs from the assumed model. In our view, this topic has not received enough attention in the literature with two notable exceptions. Banerjee and McKeague (2007) studied model-robust inference for discontinuous threshold models with cube root n convergence rate; and, more recently, Hansen (2015) studied model-robust inference for discontinuous threshold models. Our

work builds upon the work of Hansen and makes the following unique contributions: first, we improve the coverage of Wald-type confidence intervals by deriving a correct asymptotic covariance matrix for the maximum likelihood estimator; and second, we propose a test-inversion confidence interval for the threshold parameter based on a profile likelihood ratio test and derive the asymptotic distribution of the test statistic. Finally, we evaluate the performance of the Wald, test-inversion and bootstrap confidence intervals in Monte Carlo studies both under true threshold models and misspecified models.

The balance of the paper is organized as follows. Sections 2 and 3 develop the asymptotic theories for estimation. Section 4 conducts simulation studies, and Section 5 uses real data examples from HIV-1 immune correlates studies to illustrate the application of threshold models. In Section 6, we discuss the choice of methods for constructing confidence intervals in practice.

2. Limiting distributions of maximum likelihood estimators

We first treat the more general segmented model. Let $\beta = (\beta_1, \beta_z^T, \beta_x, \beta_{(x-e)_+})^T$ and $\theta = (\beta^T, e)^T$ denote all parameters in the model. Let ℓ_θ denote the log likelihood under the working model (1). For a single observation $\ell_\theta(y, x, z) = y \log p_\theta(x, z) + (1 - y) \log \{1 - p_\theta(x, z)\}$, where $p_\theta(x, z) = \text{expit}\{\eta_\theta(x, z)\}$ with $\text{expit}(x) = 1 / (1 + \exp(-x))$ and $\eta_\theta(x, z) = \beta_1 + \beta_z^T z + \beta_x x + \beta_{(x-e)_+}(x - e)_+$. Let $\hat{\theta}_n \equiv \arg\max_{\theta \in \Theta} P_n \ell_\theta$ be the maximum likelihood estimator (MLE) of θ over domain $\Theta \subset R^{D(\theta)}$, where $P_n(\cdot) = n^{-1} \sum_{i=1}^n (\cdot)$ is the empirical average operator, and $D(\theta)$ is the dimension of θ . An effective algorithm (Fong et al., 2016) for finding the MLE is to alternate between updating $\{e, \beta_{(x-e)_+}, \beta_x\}$ and $\{\beta_1, \beta_z^T\}$. When updating $\{e, \beta_{(x-e)_+}, \beta_x\}$, we use a smooth function, logistic function, to approximate the step function $I(x - e)$ (Pastor-Barriuso et al., 2003), which allows the use of gradient-based methods to explore the parameter space.

We first show that under mild conditions, $\hat{\theta}_n$ converges in probability to $\theta_0 \equiv \arg\max_{\theta \in \Theta} E_{XYZ} \ell_\theta$, where the expectation is taken with regard to the joint distribution function of (Y, X, Z) , $F(y, x, z)$. If the model is correctly specified, i.e. the data-generating model and the working model match, θ_0 is the true parameter value of the data-generating model; otherwise, we say that the model is misspecified and ℓ_{θ_0} can be viewed as the projection of the data-generating model on the working model space using the Kullback-Leibler divergence metric. Whether or not the model is correctly specified, we always assume that θ_0 is unique, meaning that if the distributions of data generated from two models with parameters θ_1 and θ_2 are indistinguishable, it is necessarily true that $\theta_1 = \theta_2$.

Theorem 1

Assume we have i.i.d. observations (Y_i, X_i, Z_i) for $i = 1, \dots, n$. Let $E X^2 < \infty$, $E \|Z\|^2 < \infty$, and the distribution of X be absolutely continuous. Assume Θ is compact and θ_0 is unique. Under these assumptions, $\hat{\theta}_n \rightarrow_p \theta_0$.

Theorem 1 follows from Theorem 5.14 of van der Vaart (2000), which does not require a high degree of smoothness in the likelihood function. The key regularity condition to check is $E \sup_{\theta \in U} \ell_{\theta} < \infty$ for every sufficiently small ball U , which holds trivially in our model because ℓ_{θ} is continuous in θ and nonpositive. Adding the fact that $\hat{\theta}_n$ maximizes the criterion function ℓ_{θ} , we have satisfied all the conditions of Theorem 5.14 of van der Vaart (2000).

We next study the asymptotic distribution of $\hat{\theta}_n$. Define $\tilde{x}_{\theta} \equiv \{1, z, x, (x - e)_+, -\beta_{(x-e)_+} I(x > e)\}^T$. Let $\dot{\ell}_{\theta}$ denote the first derivative of ℓ_{θ} with respect to θ . For any $x \neq e$, $\dot{\ell}_{\theta} = (y - p_{\theta}) \tilde{x}_{\theta}$.

Theorem 2

In addition to the assumptions of Theorem 1, assume the distribution of X has density function $f_X(\cdot)$. Denote the true mean of Y conditional on x and z by $m_0(x, z)$, i.e. $m_0(x, z) = E(Y | x, z)$. Then the second derivative matrix of $E_{XYZ} \ell_{\theta}$ is

$$V_{\theta} = -E_{XZ} p_{\theta} (1 - p_{\theta}) \tilde{x}_{\theta} \tilde{x}_{\theta}^T + \begin{bmatrix} 0 & \underline{0}^T & 0 & 0 & 0 \\ \underline{0} & [0] & \underline{0} & \underline{0} & \underline{0} \\ 0 & \underline{0}^T & 0 & 0 & 0 \\ 0 & \underline{0}^T & 0 & 0 & -E_{XZ} (m_0 - p_{\theta}) I(x > e) \\ 0 & \underline{0}^T & 0 & -E_{XZ} (m_0 - p_{\theta}) I(x > e) & \beta_{(x-e)_+} f_X(e) E_Z \{m_0(e, z) - p_{\theta}(e, z)\} \end{bmatrix}, \quad (3)$$

where $\underline{0}$ is a vector of 0's of length $D(\beta_z)$, $D(\beta_z)$ being the dimension of β_z , and $[0]$ is a $D(\beta_z) \times D(\beta_z)$ square matrix of 0's. Further assume V_{θ_0} is non-singular. Denote

$M_{\theta} = E_{XYZ} \dot{\ell}_{\theta} \dot{\ell}_{\theta}^T$. Then the limit distribution of $\hat{\theta}_n$ is

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \rightarrow_d N(0, V_{\theta_0}^{-1} M_{\theta_0} V_{\theta_0}^{-1}).$$

The derivation of (3) is detailed in Section A of the Supplementary Materials. The fact that V_{θ} is continuous in θ implies that $E_{XYZ} \ell_{\theta}$ admits a second order Taylor expansion in a neighborhood of θ_0 :

$$E_{XYZ} \ell_{\theta} = E_{XYZ} \ell_{\theta_0} + \frac{1}{2} (\theta - \theta_0)^T V_{\theta_0} (\theta - \theta_0) + o(\|\theta - \theta_0\|^2). \quad (4)$$

Adding the fact that ℓ_{θ} is Lipschitz continuous in θ , the asymptotic normality follows from Theorem 5.23 of van der Vaart (2000).

Comparing our Theorem 2 with Hansen's (2015) Theorem 2, we see that the last diagonal term of (3) is non-zero. This element vanishes along with the rest of the second term of V_{θ}

when the model is correctly specified, but is often positive when the model is misspecified. Its presence suggests that the likelihood surface may have a smaller curvature along the change point parameter at θ_0 when the model is misspecified.

When the model is correctly specified, $V_{\theta_0} = -M_{\theta_0}$ and the asymptotic variance is simply the inverse of the Fisher information. The asymptotic variance can be consistently estimated

by $\left\{ \mathbb{P}_n p_{\hat{\theta}} (1 - p_{\hat{\theta}}) \tilde{x}_{\hat{\theta}} \tilde{x}_{\hat{\theta}}^T \right\}^{-1}$, which is always positive definite; Wald-type confidence intervals for $\hat{\beta}$ can then be constructed.

When the model is misspecified, we can estimate M_{θ_0} by $\mathbb{P}_n (y - p_{\hat{\theta}})^2 \tilde{x}_{\hat{\theta}} \tilde{x}_{\hat{\theta}}^T$ and the first term of V_{θ_0} by $-\mathbb{P}_n p_{\hat{\theta}} (1 - p_{\hat{\theta}}) \tilde{x}_{\hat{\theta}} \tilde{x}_{\hat{\theta}}^T$. The off-diagonal element in the second term of V_{θ_0} can be estimated by $-\mathbb{P}_n (y - p_{\hat{\theta}}) I(x > \hat{e})$. The last diagonal element in the second term of V_{θ_0} depends on four terms: $\beta_{(x-e)_+}$ can be estimated by $\hat{\beta}_{(x-e)_+}$; $f_X(e_0)$ can be estimated by a kernel density estimate at $\hat{e}$; both $E_Z m_0(e_0, z)$ and $E_Z p_{\theta_0}(e_0, z)$ are harder to estimate because it is (x, z, y) that we observe and not (e_0, z, y) . Assuming the true mean function m_0 under model misspecification is a smooth function of x and z , we may approximate $E_Z m_0(e_0, z)$ by fitting a sufficiently flexible model to obtain an estimate of m_0 and use the plug-in estimate $\mathbb{P}_n \hat{m}_0(\hat{e}, z)$.

The last term we need to estimate the variance is $E_Z p_{\theta_0}(e_0, z)$. A natural estimator is $\mathbb{P}_n \hat{p}_{\hat{\theta}}(\hat{e}, z)$, and it presents an interesting twist. Hansen (2015) discovered that in a segmented threshold linear regression model, the asymptotic distribution of $\eta_{\hat{\theta}}(x, z)$ does not center at $\eta_{\theta_0}(x, z)$ for $x = e_0$. The same phenomenon exists here, as the following theorem shows.

Theorem 3

Under the assumptions of Theorem 2, at every x and z ,

$$\sqrt{n} \{p_{\hat{\theta}}(x, z) - p_{\theta_0}(x, z)\} \rightarrow_d g_{\theta_0}(N(0, V_{\theta_0}^{-1} M_{\theta_0} V_{\theta_0}^{-1})),$$

where, for $h = (h_{\beta}^T, h_e)^T$

$$g_{\theta}(h) = p_{\theta}(1 - p_{\theta}) \left[h_{\beta}^T \begin{pmatrix} 1 \\ z \\ x \\ (x - e)_+ \end{pmatrix} + \begin{cases} 0 & \text{if } x < e \\ (-h_e)_+ \beta_{(x-e)_+} & \text{if } x = e \\ -h_e \beta_{(x-e)_+} & \text{if } x > e \end{cases} \right].$$

The proof of the theorem can be found in Section B of the Supplementary Materials. At $x = e_0$, the limit distribution is the sum of a mean 0 normal random variable and

$p_{\theta_0}(1 - p_{\theta_0}) \beta_{(x-e)_+,0} \{-\sqrt{n}(\hat{e} - e_0)\}_+$. At first look, this makes estimating $E_Z p_{\theta_0}(e_0, z)$ difficult; however, upon closer inspection we realize that

$$p_{\theta_0}(e_0, z) = \text{expit}\{\beta_{1,0} + \beta_{z,0}^T z + \beta_{x,0} e_0 + \beta_{(x-e)_+,0}(e_0 - e_0)_+\} = \text{expit}(\beta_{1,0} + \beta_{z,0}^T z + \beta_{x,0} e_0),$$

which can be well approximated by the plug-in estimator $\text{expit}(\hat{\beta}_1 + \hat{\beta}_z^T z + \hat{\beta}_x \hat{e})$.

3. Test-inversion confidence sets

The Wald-type confidence intervals of threshold model parameters can be constructed from the asymptotic distribution of the maximum likelihood estimator derived in the last section. Under correct model specification, it has long been recognized that the finite-sample performance of Wald-type confidence intervals may be misleading for some models due to parameter-effects curvature (Seber and Wild, 1989; Pastor-Barriuso et al., 2003). An alternative approach to constructing confidence sets is by inverting hypothesis tests. A likelihood ratio test-inversion confidence set is a collection of θ that is close to the MLE $\hat{\theta}$. More precisely, it is of the form

$$\left\{ \theta : 2\ell_n(\hat{\theta}) - 2\ell_n(\theta) \leq C_{D(\theta)}(\alpha) \right\}, \quad (5)$$

where ℓ_n is the sum of the log likelihood, $\hat{\theta}$ is the maximum likelihood estimator, and $C_{D(\theta)}(1 - \alpha)$ is the $(1 - \alpha)^{th}$ quantile of the limit distribution of $2\ell_n(\hat{\theta}) - 2\ell_n(\theta_0)$. Thus, the set (5) comprises parameter values that are consistent with the data as determined by the likelihood ratio test. The following theorem gives the asymptotic distribution of the likelihood ratio statistic in segmented threshold models.

Theorem 4

Under the assumptions of Theorem 2,

$$2\ell_n(\hat{\theta}) - 2\ell_n(\theta_0) \rightarrow_d \sum_{j=1}^{D(\theta)} \lambda_j \chi_1^2,$$

where $\{\lambda_j\}_{j=1}^{D(\theta)}$ are the eigenvalues of $M_{\theta_0}^{1/2} V_{\theta_0}^{-1} M_{\theta_0}^{1/2}$.

Theorem 4 follows from the second order expansion (4) and the fact that the random quadratic form $n(\hat{\theta} - \theta_0)^T V_{\theta_0}(\hat{\theta} - \theta_0)$ can be written as a sum of weighted chi-squared distributions.

When the model is correctly specified, the limit distribution of this likelihood ratio statistic is simply a chi-squared distribution with $D(\theta)$ degrees of freedom, which is free of any parameters. When the model is misspecified, the limit distribution is a weighted sum of chi-square distributions with parameters that depend on the asymptotic distribution of the MLE.

Constructing a likelihood ratio test-inversion confidence set is typically not quite feasible unless $D(\theta)$ is rather small as the computational cost grows exponentially with $D(\theta)$. Fortunately, test-inversion confidence sets can also be based on profile likelihood (e.g. Davison, 2003, Section 4.5). Suppose $\theta = (\beta^T, \gamma^T)^T$. Let $p\ell_n(\gamma) = \max_{\beta} \ell(\beta, \gamma)$ denote the profile likelihood, also known as the concentrated likelihood of γ . A confidence set for γ can be constructed if we know the distribution of the profile likelihood ratio statistic. In threshold models, it is natural to let γ be ϵ because the model can be easily fit for any given threshold value, while the same cannot be said for the slope parameters. We now state the

limit distribution of the profile likelihood ratio statistic for the change point parameter, the proof of which is given in Section C of the Supplementary Materials.

Theorem 5

Under the assumptions of Theorem 2,

$$2p\ell_n(\hat{e}) - 2p\ell_n(e_0) \rightarrow_d \lambda \chi_1^2,$$

where $\lambda = -[V_{\theta_0}^{-1} M_{\theta_0} V_{\theta_0}^{-1}]_{ee} / [V_{\theta_0}^{-1}]_{ee}$. Here $[\cdot]_{ee}$ denote the last diagonal element of a square matrix.

When the model is correctly specified, λ is 1. When the model is misspecified, the limit distribution of the likelihood ratio statistic is not any simpler than the limit distribution of the parameter estimate. We note that Hansen (2015) had proposed the use of a test-inversion confidence set for the threshold parameter under model misspecification, but assumed $\lambda = 1$.

To construct test-inversion confidence intervals for e after finding the MLE by approximating the step function in the likelihood with a smooth function, we perform a grid search locally. Specifically, to find the upper bound, we test increasingly larger e in small increments starting from $\hat{e}$ until the difference between $2p\ell_n(\hat{e})$ and $2p\ell_n(e)$ drops below the critical value, and we find the lower bound similarly. If the MLE is found not through smooth approximation but by comparing profile likelihoods $p\ell_n(e)$ on a global grid of candidate change points, the test-inversion confidence interval would come as a by-product and incur no additional computational burden. Overall it is still faster to compute the MLE by smooth approximation and find the test-inversion confidence interval by local grid search.

4. Monte Carlo studies

In this section we present two simulation scenarios. In scenario I the data are generated from threshold models; in scenario II the data are generated from a quadratic model. Results from an additional model misspecification scenario are shown in the Supplementary Materials.

4.1 Scenario I

The first batch of Monte Carlo experiments is conducted under correct model specification. We simulate data from the segmented and hinge logistic regression models (1) and (2) with $\beta_z = \log(1.4) = 0.34$, $\beta_x = -\log(0.7) = 0.36$ and $\beta_{(x-e)_+} = \log(0.4) = -0.92$ and $e = 2.2$. The intercepts are chosen so that the prevalence of Y is around 1/4 under each model. The covariate z is simulated from a standard normal distribution. We simulate x from a scaled and shifted gamma distribution with median 2.06 and interquartile range 1.76, as the asymmetry and heavy-tailedness of a gamma random variable better models the distribution of immune biomarker measurements. We experiment with sample sizes: 25, 50, 100, 250, 500, 2000 and 10,000.

Numerical summaries of the simulation results from 10,000 replicates are shown in Table 1, and the histograms of the sampling distributions are shown in the Supplementary Materials Figure D.1 and D.2. To compute relative bias of the slope estimates, we divide bias by the true value; to compute relative bias of the threshold estimate, we divide bias by the interquartile range of the x covariate. In addition to point estimates and relative bias, we show the width of 95% confidence intervals and actual coverage from three methods: ‘approx’, ‘model’ and ‘test inv’. The first two are symmetric Wald-type intervals, and the last is the asymmetric profile likelihood ratio test-inversion confidence interval for the threshold parameter only. The ‘approx’ method implements an approach from Pastor-Barriuso et al. (2003) that approximates the indicator function $I(x > e)$ in the threshold models with a logistic function, and uses the observed information of the smoothed model to estimate the covariance matrix. The ‘model’ method is based on the asymptotic variance formula in Theorem 2 under correct model specification. The ‘test inv’ method is based on Theorem 5 under correct model specification.

Under the segmented model, we see that bias decreases as the sample size increases for all parameters. The degree of bias differs between parameters. For example, at $n = 500$ the relative bias is 47%, 37%, 5% and 1% for $\hat{\beta}_x$, $\hat{\beta}_{(x-e)_+}$, $\hat{e}$ and $\hat{\beta}_z$, respectively. This fits our intuition that non-smoothness of threshold models makes estimation more difficult for parameters associated with the change point covariate, especially in small to moderate samples. Among the methods for confidence intervals construction, ‘approx’ works reasonably well for β_z and $\beta_{(x-e)_+}$ but undercovers β_x and e . ‘Model’ has improved coverage for β_x and e , yet the coverage for e is still unsatisfactory. Finally, the coverage for e is substantially improved in ‘test inv’. In summary, ‘model’+‘test inv’ provide reasonable finite sample coverage for both the slope and threshold parameters.

Under the hinge model, we see similar results with some notable differences: $\hat{\beta}_{(x-e)_+}$ is much less biased in the hinge model than in the segmented model under this simulation setting. For example, at $n = 500$, the relative bias is 37% and 12% under the segmented and hinge model, respectively. However, the relative bias of $\hat{e}$ under the segmented model remains small at 2% even when n drops down to 50 and 25, but the relative bias of $\hat{e}$ in a hinge model increases to 15% and 24%, respectively, at those sample sizes. In addition, while the precision of $\hat{\beta}_z$ as measured by confidence interval width is unchanged between the two models, the precision of $\hat{\beta}_{(x-e)_+}$ and $\hat{e}$ is much higher in the hinge model than in the segmented model. For example, the confidence interval for $\beta_{(x-e)_+}$ at $n = 500$ has a median width of 1.38 and 1.01 under the segmented and hinge model, respectively, which corresponds to a relative efficiency of 1.87.

To study the impact of threshold location on model estimation, we compared three hinge models with $e \in \{2.2, 1.5, 1\}$ and three segmented models with the same threshold values. The results, summarized and discussed in more detail in Section D.4 of the Supplementary Materials, show that for the hinge model the threshold estimate is least biased when the true threshold is at 1.5, which is ~ 30 percentile of the distribution of x ; while for the segmented model the threshold estimate is least biased when the true threshold is 2.2, which is near the median of X .

4.2 Scenario II

The second batch of experiments is conducted under the scenario that the model is misspecified. We simulate data from a quadratic model:

$$\text{logit} \{ \Pr(Y=1|x, z) \} = \alpha + \log(1.4)z - x + 0.3x^2.$$

The covariates are simulated as in scenario I, and the intercept is also chosen such that there are about a quarter of cases. Figure 1 shows a simulated dataset with $n = 500$ along with the fitted models. To determine θ_0 under the change point working models, we simulate datasets of sample size 100,000 and take the median of 2,000 replicates. In the segmented model, $\beta_{z,0} = 0.34$, $\beta_{x,0} = -0.040$, $\beta_{(x-e)_+,0} = 1.49$, and $e_0 = 2.82$; in the hinge model, $\beta_{z,0} = 0.34$, $\beta_{(x-e)_+,0} = 1.46$, and $e_0 = 2.87$. These values approximate the asymptotic limit of the threshold model parameter estimates and are used to help define finite sample biases of the MLE with respect to the population limit.

Numerical summaries of the simulation results for sample sizes ranging from 25 to 10,000 from 10,000 replicates are shown in Table 2 and 3, and the histograms of the sampling distributions are shown in the Supplementary Materials Figure D.3 and D.4. In addition to the estimate and relative bias, we compare five methods for constructing confidence intervals: ‘model’, ‘robust’, ‘test inv’, ‘boot basic’, and ‘boot bca*’. The first three are analytical confidence intervals, and the last two are bootstrap confidence intervals. ‘Model’ and ‘robust’ are based on the asymptotic variance formula in Theorem 2 under correct model specification and model misspecification, respectively. Under the ‘robust’ method, to estimate $E_Z m_0(e_0, z)$, we model the effect of x through a natural cubic spline with two degrees of freedom. The ‘test inv’ method is based on Theorem 5 under model misspecification. ‘Boot basic’ and ‘boot bca*’ are implemented with the help of the R package boot. Results from 2000 bootstrap replicates are reported. The basic bootstrap method is a member of the pivotal family of bootstrap methods (Carpenter and Bithell, 2000). Related to the basic bootstrap method is the studentized bootstrap method, whose performance is unstable in our models and not included in the table. Among the non-pivotal family of bootstrap methods, we find that the bias corrected (bc) percentile method (Efron, 1981) and the bias corrected and accelerated method (bca) (Efron, 1987) have similar performance when the sample size is 250 or 500. In ‘bca*’, we report results from bca when the sample size is less than or equal to 500; and results from bc when the sample size is larger because bca requires more bootstrap replicates when the sample size is larger. Another member of the non-pivotal family is the percentile method, whose overall performance is not as good as bc and bca, hence not included in the table.

Under the segmented model, bias decreases with sample size. The relative bias of $\hat{\beta}_x$ is artificially high because $\beta_{x,0} = -0.040$ is very close to 0. As in scenario I, the biases of $\hat{\beta}_x$ and $\hat{\beta}_{(x-e)_+}$ are higher than those of $\hat{e}$ and $\hat{\beta}_z$. Among the analytical methods for confidence intervals construction, ‘model’ works reasonably well for β_z and $\beta_{(x-e)_+}$ but undercovers β_x and e , substantially so for e . The undercoverage is not helped by increasing sample size. ‘Robust’ improves the coverage for both β_x and e , bringing it close to the nominal level, but it shows some degree of overcoverage, particularly when the sample size is 2000 or above.

‘Test inv’ provides further improvement in the coverage for e by improving coverage and reducing the confidence interval width.

Among the bootstrap methods for constructing confidence intervals, ‘basic’ provides good coverage for β_z and $\beta_{(x-e)_+}$ but undercovers β_x and e . ‘Bca*’ improves the coverage for β_x and e but sees some pullback in the coverage for $\beta_{(x-e)_+}$, e.g. 88% at $n = 500$. Overall, ‘bca*’ is preferable over ‘basic’. Comparing the ‘bca*’ bootstrap interval with the ‘robust’ + ‘test inv’ analytical interval, we see that the bootstrap intervals are typically shorter and have lower coverage for β_x , $\beta_{(x-e)_+}$, and e when sample sizes are small. This difference decreases when the sample size reaches 2000 and disappears at $n = 10,000$.

Under the hinge model, we see similar results with some interesting differences, some of which are also seen in scenario 1. First, as in scenario 1, $\hat{\beta}_{(x-e)_+}$ is much less biased in the hinge model fit than in the segmented model fit. For example, at $n = 500$, the relative bias is 33% and 10% under the segmented and hinge model, respectively. Second, there is less of an overcoverage problem in the hinge model fit than in the segmented model fit. Third, the width of the ‘robust’, ‘test inv’ and ‘bca*’ intervals are similar to each other. In some cases, the ‘robust’ method has the shorter intervals and better coverage. For example, for $\beta_{(x-e)_+}$ the ratios between the width of the ‘robust’ and ‘bca*’ intervals are 0.39, 0.82 and 0.94 for $n = 25, 250$ and 2000, respectively; and the coverage probabilities are 92%, 95% and 94% for the ‘robust’ intervals and 84%, 85% and 96% for the ‘bca*’ intervals.

5. Application in HIV-1 immune correlates of risk studies

Developing an effective HIV-1 vaccine is currently one of the greatest public health challenges. One strategy to accelerate development is to study so-called immune correlates of risk (Gilbert et al., 2008; Plotkin and Gilbert, 2012), which are biomarkers measuring immune responses to vaccination or natural infection that are associated with the risk of HIV-1 infection. As part of this strategy, Permar et al. (2015) measured antibody-based immune responses in blood samples from a cohort of HIV-1 infected pregnant women in the era before the advent of antiretroviral therapy (Rich et al., 2000) to study their association with the risk of mother-to-child transmission (MTCT) of HIV-1. An array of immunological assays was conducted to measure binding antibodies, neutralizing antibodies, antibody-dependent cell-mediated cytotoxicity, and antibody avidity, the last of which refers to the combined effect of multiple antibody binding affinities. Permar et al. showed that the risk of transmission of HIV-1 from mothers to children was associated with binding antibodies against the V3 region of the HIV-1 envelope protein as well as neutralization activities against the so-called tier 1 HIV-1 strains, which are relatively easy to neutralize compared to other circulating strains of HIV-1. The latter finding came as a surprise because the weakly neutralizing antibodies had not been associated with protection against HIV-1 infection in vaccine studies. A potential explanation for the association of weakly neutralizing antibodies with MTCT risk was the finding that maternal antibodies could indeed neutralize concomitant autologous virus strains (Moody et al., 2015).

Permar et al. used logistic regression models with linear terms to select immune responses that are associated with MTCT risk. Our goal here is to further refine our understanding of

the association between MTCT risk and the covariate $NAb_SF162LS(x)$, which is a measure of the neutralizing activities against a tier 1 HIV-1 strain SF162LS. The dataset comprises 236 samples, each corresponding to one infected pregnant woman. There are 79 transmitters and 157 non-transmitters. In addition to $NAb_SF162LS$, all models adjust for $birth(z)$, which indicates whether the birth was C-section or vaginal.

We first seek to detect nonlinearity by testing the null hypothesis $\beta_{(NAb_SF162LS-e)+} = 0$ against the alternative $\beta_{(NAb_SF162LS-e)+} \neq 0$ in a segmented model using a maximum of scores test statistic (Fong et al., 2015). The null is rejected with p-value 0.003. We then fit four logistic regression models to the data. Two of them model $NAb_SF162LS$ through natural cubic splines with two and three degrees of freedom, respectively; and the other two models $NAb_SF162LS$ through the hinge and segmented threshold effects, respectively. The fitted risks are shown as functions of $NAb_SF162LS$ in Figure 2. The nonlinearity of the relationship between MTCT risk and $NAb_SF162LS$ is apparent in the two cubic spline fits. Table 4 reports the estimated parameters and 95% confidence intervals for the two threshold models. The two models have similar estimates for the slope of the $birth$ covariate. In the segmented model fit, both $\hat{\beta}_{NAb_SF162LS}$ and $\hat{\beta}_{(NAb_SF162LS-e)+}$ are not significant according to either model-robust Wald or BCa bootstrap confidence intervals. We thus focus on interpreting the hinge model fit.

In the hinge model fit, $\hat{\beta}_{(NAb_SF162LS-e)+}$ is significant by both model-robust Wald and BCa confidence intervals. The model also suggests that $NAb_SF162LS$ is not associated with MTCT risk when it is less than 7.37, although the estimated confidence intervals for the threshold are fairly wide by either the model-robust test-inversion method or BCa. Above the threshold, one unit increase in $NAb_SF162LS$ is associated with an odds ratio of 0.67 (95% robust confidence interval: 0.51, 0.88). In contrast, in a logistic regression model fit with linear $NAb_SF162LS$, the estimated odds ratio for one unit increase in $NAb_SF162LS$ is attenuated to be 0.82 (95% robust confidence interval: 0.71, 0.95).

As a second example we apply threshold models to the immune correlates dataset from the RV144 HIV-1 vaccine trial (Haynes et al., 2012). The results are shown in the Supplementary Materials Section F. Results from both of these examples suggest that, when appropriate, hinge threshold models are particularly attractive compared with generalized linear models and segmented threshold models. Our applications deal with covariates that are measured by sensitive immunoassays. Assay readouts in the lower range are likely to be dominated by experimental noises rather than biological signals. By using hinge threshold models we can get a more accurate and less attenuated picture of the relationship between immune responses and the outcome of interest than generalized linear models.

6. Conclusions

In this paper we investigate the asymptotic behavior of continuous threshold model parameter estimates and use the knowledge to construct two types of model-robust analytical confidence intervals. To provide guidance for choosing the most appropriate inference procedures in real data analyses, we conduct simulation studies to evaluate these methods as well as bootstrap confidence intervals. Three main conclusions can be drawn: (1) When

using model-robust intervals, profile likelihood ratio test-inversion intervals should replace Wald intervals for the threshold parameter. (2) Among the commonly used bootstrap confidence interval methods, bias corrected (accelerated) intervals have the best performance. (3) Analytical confidence intervals assuming correct model specification have low coverage when the model is in fact misspecified, hence overly optimistic. Both model-robust and bootstrap confidence intervals provide a remedy to the problem, but neither is perfect at moderate sample sizes in the segmented model: model-robust intervals have overcoverage and bootstrap intervals have undercoverage of some parameters. Thus, the choice of method will depend on which error is more tolerable in a given application. In the hinge model model-robust intervals work better overall.

When the sampling distributions are skewed and heavy-tailed, the bootstrap distributions intervals are asymmetric around the point estimates by construction. The model-robust analytical intervals, on the other hand, are always symmetric. In the hinge model, the sampling distributions are close to being normal, and the length of the model-robust and bootstrap intervals are close to each other as well. In the segmented model, the sampling distributions are more skewed and heavy-tailed. As the model-robust intervals are constructed based on the second moment only, they tend to be longer than the bootstrap intervals. In addition, when the skewness of the sampling distributions and the asymmetry of the bootstrap intervals go in a coordinated direction, we get good coverage; but if they go in the wrong direction, we get undercoverage, and that is what happens with, e.g. the bca intervals of $\hat{\beta}_{(x-e)+}$.

Our numerical studies suggest that the hinge model may be estimated with higher precision than the segmented model. To investigate this further, we generate data from a hinge model as specified in scenario I, and fit the segmented model. Results are summarized in Table D.1 in the Supplementary Materials, where, for the ease of comparison, we include a replicate of the hinge model fit results. Based on the results at $n = 10,000$, it takes twice as many samples to fit a segmented model to achieve the same level of precision in $\hat{\beta}_{(x-e)+}$ as it does to fit a hinge model. For $\hat{e}$, the ratio is $(0.36/0.28)^2 = 1.65$. Under model misspecification, in scenario II, $\beta_{x,0}$ is -0.04 , small enough that a hinge model fit seems a sufficient summary of the relationship between the outcome and x . Indeed, $\beta_{(x-e)+,0}$ under the segmented model is close to that under the hinge model, so is e_0 . Based on results from $n = 10,000$ in Table 2 and 3, the relative efficiency of $\hat{\beta}_{(x-e)+}$ is close to 1 and the relative efficiency of $\hat{e}$ is $(0.75/0.36)^2 = 4.34$. In scenario II, the hinge model fit and the segmented model fit are very different; thus, it is not as meaningful to compare the precision of estimation.

Our theoretical results depend on the assumption that θ_0 is unique. When θ_0 is not unique, we have a nonregularity problem, which is connected with several other such problems as listed by McKeague and Qian (2015). They include post-model-selection inference, the Davies problem (of which threshold models are an example), and when a parameter is on the boundary of the parameter space. A unifying theme of these problems is that the asymptotic distributions of parameter estimates are discontinuous near the boundary between identifiability and non-identifiability in the parameter space, while their finite sample distributions are continuous. We illustrate this with a simulation study on the segmented

model in Section D.5 of the Supplementary Materials. The approach of McKeague and Qian, which builds upon local asymptotics results, is also promising for threshold models.

Threshold models are useful in practice because they are more flexible than simple linear models and more interpretable than spline or polynomial models. To make judicious use of threshold models in practice, we should compare the threshold model fit with fits from more flexible smooth models to ensure that the trends they represent match each other. It is also important to note that the meanings of the threshold model parameter estimates depend on the model, and we should be careful not to over-interpret them. To illustrate, we present an example based on a hybrid hinge-quadratic model in Section G of the Supplementary Materials.

A useful alternative to the threshold models is the transition models proposed by Pastor-Barriuso et al. (2003), which use either logistic function $[\exp \{(x - e)/\varepsilon\} - 1]/[\exp \{(x - e)/\varepsilon\} + 1]$ or hyperbolic function $\sqrt{(x - e)^2 + \varepsilon^2}/(x - e)$ in the place of the indicator function $2I(x > e) - 1$. As $\varepsilon \rightarrow 0$, the two functions approach the indicator function. The parameter ε may be specified beforehand or estimated from the data. When it is estimated from the data, the resulting model fit may look very different from their threshold model counterparts, especially around the threshold. Formal methods for guiding the choice between threshold models and transition models are worth further investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Co-Editor, Associate Editor, and two referees for their valuable comments. This work was supported by the National Institutes of Health (R01-AI122991; R21-ES022332; R01-GM106177; UM1-AI068635) and the Henry M. Jackson Foundation (W81XWH-07-2-0067).

References

- Antoch J, Gregoire G, Hušková M. Tests for continuity of regression functions. *Journal of statistical planning and inference*. 2007; 137:753–777.
- Antoch J, Gregoire G, Jarušková D. Detection of structural changes in generalized linear models. *Statistics & probability letters*. 2004; 69:315–332.
- Banerjee M, McKeague IW. Confidence sets for split points in decision trees. *The Annals of Statistics*. 2007; 35:543–574.
- Carmody JB, Seckeler MD, Ballengee CR, Conaway M, Jayakumar KA, Charlton JR. Pre-operative renal volume predicts peak creatinine after congenital heart surgery in neonates. *Cardiol Young*. 2014; 24:831–839. [PubMed: 24229479]
- Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Statistics in medicine*. 2000; 19:1141–1164. [PubMed: 10797513]
- Chan K-S, Tsay RS. Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika*. 1998; 85:413–426.
- Csörgő, M., Horváth, L. *Wiley series in probability and statistics*. New York: Wiley; 1997. Limit theorems in change-point analysis.
- Davies R. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*. 1987; 74:33–43.

- Davison, AC. Statistical models. Vol. 11. Cambridge University Press; 2003.
- Efron B. Nonparametric standard errors and confidence intervals. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*. 1981; 9:139–158.
- Efron B. Better bootstrap confidence intervals. *Journal of the American statistical Association*. 1987; 82:171–185.
- Fong Y, Di C, Permar S. Change point testing in logistic regression models with interaction term. *Statistics in medicine*. 2015; 34:1483–1494. [PubMed: 25612253]
- Fong Y, Huang Y, Gilbert P, Permar S. Estimation and hypothesis testing for four types of threshold regression models. University of Washington Department of Biostatistics Technical Report. 2016
- Gijbels I. Smoothing and preservation of irregularities using local linear fitting. *Applications of Mathematics*. 2008; 53:177–194.
- Gilbert P, Qin L, Self S. Evaluating a surrogate endpoint at three levels, with application to vaccine development. *Statistics in Medicine*. 2008; 27:4758–4778. [PubMed: 17979212]
- Goetghebeur EJ, Pocock SJ. Detection and estimation of j-shaped risk-response relationship. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. 1995; 158:107–121.
- Hall CB, Lipton RB, Sliwinski M, Stewart WF. A change point model for estimating the onset of cognitive decline in preclinical alzheimer's disease. *Statistics in Medicine*. 2000; 19:1555–1566. [PubMed: 10844718]
- Hansen BE. Sample splitting and threshold estimation. *Econometrica*. 2000; 68:575–603.
- Hansen BE. Regression kink with an unknown threshold. *Journal of Business and Economic Statistics*. 2015 in press.
- Haynes BF, Gilbert PB, McElrath MJ, Zolla-Pazner S, Tomaras GD, Alam SM, et al. Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *New England Journal of Medicine*. 2012; 366:1275–1286. [PubMed: 22475592]
- Koziol JA, Wu S-CH. Changepoint statistics for assessing a treatment-covariate interaction. *Biometrics*. 1996; 52:1147–1152. [PubMed: 8924575]
- Lee S, Seo M, Shin Y. Testing for threshold effects in regression models. *Journal of the American Statistical Association*. 2011; 106:220–231.
- Loader CR, et al. Change point estimation using nonparametric regression. *The Annals of Statistics*. 1996; 24:1667–1678.
- Mazumdar M, Smith A, Bacik J. Methods for categorizing a prognostic variable in a multivariable setting. *Statistics in Medicine*. 2003; 22:559–571. [PubMed: 12590414]
- McKeague IW, Qian M. An adaptive resampling test for detecting the presence of significant predictors. *Journal of the American Statistical Association*. 2015; 110:1422–1433. [PubMed: 27073292]
- Moody MA, Gao F, Gurley TC, Amos JD, Kumar A, Hora B, Marshall DJ, Whitesides JF, Xia S-M, Parks R, et al. Strain-specific v3 and cd4 binding site autologous hiv-1 neutralizing antibodies select neutralization-resistant viruses. *Cell host & microbe*. 2015; 18:354–362. [PubMed: 26355218]
- Muller H-G. Change-points in nonparametric regression analysis. *The Annals of Statistics*. 1992:737–761.
- Pastor R, Guallar E. Use of two-segmented logistic regression to estimate change-points in epidemiologic studies. *American journal of epidemiology*. 1998; 148:631–642. [PubMed: 9778169]
- Pastor-Barriuso R, Guallar E, Coresh J. Transition models for change-point estimation in logistic regression. *Statistics in Medicine*. 2003; 22:1141–1162. [PubMed: 12652559]
- Permar SR, Fong Y, Vandergrift N, Fouda GG, Gilbert P, Parks R, et al. Maternal hiv-1 envelope-specific antibody responses and reduced risk of perinatal transmission. *Journal of Clinical Investigation*. 2015; 125:2702–2706. [PubMed: 26053661]
- Plotkin S, Gilbert P. Nomenclature for immune correlates of protection after vaccination. *Clinical Infectious Diseases*. 2012; 54:1615–1617. [PubMed: 22437237]
- Rich KC, Fowler MG, Mofenson LM, Abboud R, Pitt J, Diaz C, Hanson IC, Cooper E, Mendez H. for the Women, and Group, I. T. S. Maternal and infant factors predicting disease progression in

- human immunodeficiency virus type 1-infected infants. *Pediatrics*. 2000; 105:e8. [PubMed: 10617745]
- Seber, G., Wild, C. *Wiley Series in Probability and Statistics*. Wiley; 1989. *Nonlinear Regression*.
- Tomaras GD, Haynes BF. Advancing toward hiv-1 vaccine efficacy through the intersections of immune correlates. *Vaccines*. 2014; 2:15–35. [PubMed: 24932411]
- Ulm K. A statistical method for assessing a threshold in epidemiological studies. *Statistics in Medicine*. 1991; 10:341–349. [PubMed: 2028118]
- van der Vaart, A. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press; 2000. *Asymptotic Statistics*.
- Vexler A, Gurevich G. Guaranteed local maximum likelihood detection of a change point in nonparametric logistic regression. *Communications in Statistics - Theory and Methods*. 2006; 35:711–726.
- Vexler A, Gurevich G. Average most powerful tests for a segmented regression. *Communications in Statistics - Theory and Methods*. 2009; 38:2214–2231.
- Xu R, Adak S. Survival analysis with time-varying regression effects using a tree-based approach. *Biometrics*. 2002; 58:305–315. [PubMed: 12071403]
- Zeileis A, Leisch F, Hornik K, Kleiber C. strucchange: An r package for testing for structural change in linear regression models. *Journal of Statistical Software*. 2002; 7:1–38.
- Zheng G, Chen Z. Comparison of maximum statistics for hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrics*. 2005; 61:254–258. [PubMed: 15737101]

Figure 1.

A sample dataset under scenario II, $n = 1000$. The lines show the estimated linear combinations $\hat{\eta}_{\hat{\theta}}(x, 0)$ (left) and estimated probabilities $\hat{p}_{\hat{\theta}}(x, 0)$ (right) based on the fitted models in the left and right panels, respectively.

Figure 2.
Estimated log odds ratio as functions of NAb_SF162LS in the MTCT example.

Table 1

Simulation results under correct model specification. $\beta_{x,0} = 0.34$, $\beta_{x,0} = 0.36$, $\beta_{(x-e)_+,0} = -0.92$ and $e_0 = 2.2$. Est: MC mean of parameter estimate; %bias: relative bias with respect to β_0 for the β' 's and interquartile range of x for e . Three confidence interval methods are compared based on MC median width of 95% confidence interval and actual coverage (in parentheses). Approx: Wald CI based on asymptotic variance derived by approximating the indicator function $I(x > e)$ in the threshold models with a smooth function; model: Wald CI based on asymptotic variance derived exactly; test inv: test inversion CI based on asymptotic distribution derived exactly.

n	Segmented			Hinge				
	Est(%bias)	Approx	Model	Test Inv	Est(%bias)	Approx	Model	Test Inv
β_x								
25	0.47(41)	2.29(97)	2.33(98)		0.48(42)	2.18(98)	2.23(98)	
50	0.39(15)	1.42(95)	1.43(95)		0.37(9)	1.41(96)	1.42(96)	
100	0.36(8)	0.95(95)	0.95(95)		0.35(5)	0.96(95)	0.96(96)	
250	0.35(4)	0.58(95)	0.58(95)		0.34(3)	0.59(95)	0.59(95)	
500	0.34(1)	0.41(95)	0.41(95)		0.34(1)	0.41(95)	0.42(95)	
2000	0.34(0)	0.20(95)	0.20(95)		0.33(-1)	0.21(95)	0.21(95)	
10000	0.34(0)	0.09(96)	0.09(96)		0.34(0)	0.09(96)	0.09(96)	
β_e								
25	0.64(61)	6.30(95)	7.17(96)					
50	1.10(174)	3.93(93)	4.57(95)					
100	1.11(178)	2.31(93)	2.74(95)					
250	0.79(98)	1.29(87)	1.52(91)					
500	0.59(47)	0.89(84)	1.04(88)					
2000	0.42(6)	0.47(87)	0.52(90)					
10000	0.40(1)	0.23(93)	0.23(93)					
$\beta_{(x-e)_+}$								
25	-1.17(28)	10.58(97)	11.91(98)		-1.27(39)	3.23(94)	4.33(96)	
50	-1.97(115)	6.61(93)	7.35(96)		-1.35(47)	2.39(97)	3.11(99)	
100	-2.04(123)	4.03(93)	4.42(95)		-1.42(55)	1.87(95)	2.41(97)	
250	-1.58(73)	2.09(94)	2.21(96)		-1.22(33)	1.20(91)	1.46(94)	
500	-1.26(37)	1.33(94)	1.38(95)		-1.02(12)	0.85(89)	1.01(93)	
2000	-0.96(5)	0.62(94)	0.63(95)		-0.93(2)	0.45(93)	0.49(95)	

n	Segmented			Hinge				
	Est(%bias)	Approx	Model	Test Inv	Est(%bias)	Approx	Model	Test Inv
10000	−0.92(1)	0.28(94)	0.28(94)		−0.92(0)	0.21(95)	0.22(96)	
ρ								
25	2.23(2)	1.09(47)	2.80(84)	3.28(96)	1.78(24)	1.36(50)	5.32(98)	2.85(89)
50	2.16(2)	0.91(41)	2.16(73)	3.32(94)	1.93(15)	1.18(55)	3.75(98)	3.04(96)
100	2.22(1)	0.84(39)	1.92(69)	3.60(91)	2.12(4)	1.04(55)	2.61(92)	2.91(96)
250	2.29(5)	0.77(45)	1.59(70)	2.58(90)	2.18(1)	0.85(60)	1.70(87)	2.11(93)
500	2.29(5)	0.69(53)	1.28(73)	1.72(89)	2.18(1)	0.72(68)	1.23(88)	1.43(93)
2000	2.23(2)	0.50(72)	0.72(84)	0.80(92)	2.19(1)	0.49(85)	0.62(92)	0.65(95)
10000	2.20(0)	0.29(89)	0.33(91)	0.34(94)	2.19(0)	0.26(93)	0.28(94)	0.28(95)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Simulation results when model is misspecified, segmented model fit, $\beta_{z0} = 0.34$, $\beta_{x0} = -0.040$, $\beta_{(x-e)_{+0}} = 1.49$ and $e_0 = 2.82$. Est: MC mean of parameter estimate; %bias: relative bias with respect to β_0 for the β' 's and interquartile range of x for e . Five confidence interval methods are compared based on MC median width of 95% confidence interval and actual coverage (in parentheses). Model: Wald CI based on asymptotic variance derived assuming correct model specification; robust and test inv: Wald and test inversion CI based on asymptotic variance derived without assuming correct model specification; boot basic: basic bootstrap CI; boot beca*: bias corrected (and accelerated when $n < 2000$) bootstrap CI.

n	Est(%bias)	Model	Robust	Test Inv	Boot basic	Boot beca*
β_z						
25	0.45(35)	2.58(99)	2.48(95)		3.66(97)	3.71(92)
50	0.41(21)	1.71(96)	1.76(95)		2.29(99)	2.21(96)
100	0.36(8)	1.12(95)	1.18(96)		1.32(99)	1.30(97)
250	0.35(3)	0.68(95)	0.71(96)		0.73(97)	0.72(96)
500	0.34(2)	0.48(94)	0.48(95)		0.49(96)	0.49(95)
2000	0.34(0)	0.23(95)	0.24(95)		0.24(96)	0.24(95)
10000	0.34(0)	0.10(95)	0.10(95)		0.11(95)	0.10(95)
β_x						
25	-1.51(3677)	6.76(99)	8.99(93)		8.25(79)	7.91(81)
50	-1.23(2979)	4.06(97)	9.21(94)		7.42(86)	7.16(86)
100	-0.85(2044)	2.23(94)	7.09(96)		6.34(93)	5.93(92)
250	-0.44(1017)	1.28(84)	4.05(96)		3.53(88)	3.24(94)
500	-0.24(491)	0.89(78)	2.43(96)		2.05(82)	1.78(94)
2000	-0.09(117)	0.45(75)	0.97(99)		0.90(86)	0.85(96)
10000	-0.05(37)	0.20(70)	0.38(99)		0.40(90)	0.39(96)
$\beta_{(x-e)_{+}}$						
25	3.60(142)	12.89(98)	11.96(91)		13.01(72)	13.27(79)
50	3.41(129)	8.26(96)	11.03(91)		10.77(79)	11.50(82)
100	3.10(108)	5.28(96)	10.23(96)		7.80(88)	8.35(81)
250	2.41(62)	2.60(95)	6.49(98)		5.86(96)	4.85(85)
500	1.97(33)	1.61(94)	2.89(98)		3.51(96)	1.79(88)
2000	1.59(7)	0.73(94)	0.82(98)		0.99(95)	0.75(94)
10000	1.51(1)	0.31(95)	0.32(97)		0.34(95)	0.32(95)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

<i>n</i>	Est(%bias)	Model	Robust	Test Inv	Boot basic	Boot bca*
25	2.43(22)	3.31(79)	3.35(64)	2.72(70)	2.80(42)	1.82(57)
50	2.51(18)	2.53(72)	6.02(77)	3.28(82)	2.93(44)	2.21(65)
100	2.69(7)	1.93(63)	6.82(80)	3.79(88)	3.02(46)	2.54(77)
250	2.80(1)	1.38(57)	6.05(87)	4.24(94)	2.91(56)	2.76(90)
500	2.83(1)	1.06(57)	4.50(92)	3.93(97)	2.52(64)	2.51(94)
2000	2.81(0)	0.58(54)	1.94(99)	1.80(99)	1.59(77)	1.57(96)
10000	2.80(1)	0.27(51)	0.77(98)	0.75(98)	0.78(83)	0.76(96)

e

Table 3

Simulation results when model is misspecified, hinge model fit, $\beta_{z0} = 0.34$, $\beta_{(x-e)_+,0} = 1.46$ and $e_0 = 2.87$. Est: MC mean of parameter estimate; %bias: relative bias with respect to β_0 for the β' 's and interquartile range of x for e . Five confidence interval methods are compared based on MC median width of 95% confidence interval and actual coverage (in parentheses). Model: Wald CI based on asymptotic variance derived assuming correct model specification; robust and test inv: Wald and test inversion CI based on asymptotic variance derived without assuming correct model specification; boot basic: basic bootstrap CI; boot bca*: bias corrected (and accelerated when $n < 2000$) bootstrap CI.

n	Est(%bias)	Model	Robust	Test Inv	Boot basic	Boot bca*
β_z						
25	0.43(29)	2.64(99)	2.31(94)		3.98(98)	3.84(94)
50	0.38(12)	1.66(96)	1.58(94)		2.19(100)	2.11(97)
100	0.36(7)	1.11(95)	1.09(95)		1.26(99)	1.24(97)
250	0.34(2)	0.68(95)	0.68(95)		0.71(97)	0.71(96)
500	0.34(0)	0.47(95)	0.47(95)		0.49(96)	0.49(95)
2000	0.34(0)	0.23(95)	0.23(95)		0.24(96)	0.24(95)
10000	0.34(0)	0.10(96)	0.10(95)		0.11(95)	0.10(95)
$\beta_{(x-e)_+}$						
25	1.95(33)	4.73(98)	3.07(92)		7.64(85)	7.89(84)
50	2.11(44)	3.65(96)	3.19(93)		6.72(81)	5.58(84)
100	2.19(50)	2.88(94)	3.15(94)		6.12(78)	3.74(85)
250	1.82(24)	1.66(93)	2.07(95)		4.39(84)	2.52(93)
500	1.61(10)	1.13(90)	1.47(94)		2.37(86)	1.73(96)
2000	1.48(1)	0.55(88)	0.73(94)		0.83(89)	0.78(96)
10000	1.46(0)	0.25(86)	0.33(94)		0.35(91)	0.34(96)
e						
25	2.25(35)	6.29(98)	2.90(65)	2.18(57)	3.20(43)	2.78(68)
50	2.63(14)	3.41(94)	3.44(83)	3.17(81)	3.20(59)	2.98(85)
100	2.87(0)	2.19(85)	3.15(90)	3.23(91)	2.84(63)	2.76(89)
250	2.89(1)	1.38(78)	2.32(96)	2.35(97)	2.18(76)	2.30(94)
500	2.87(0)	0.98(78)	1.68(98)	1.65(98)	1.67(82)	1.71(96)
2000	2.86(1)	0.50(79)	0.83(97)	0.80(97)	0.83(87)	0.83(96)
10000	2.86(1)	0.22(77)	0.37(95)	0.36(96)	0.38(90)	0.38(95)

Table 4

Threshold regression model fits in the MTCT example. OR: odds ratio; robust/BCa: model-robust and bias corrected or accelerated bootstrap 95% CI.

	segmented model		hinge model	
	OR (robust)	(BCa)	OR (robust)	(BCa)
birthVaginal	1.23 (0.63, 2.39)	(0.60, 2.39)	1.24 (0.64, 2.40)	(0.61, 2.45)
NAb_SF162LS	1.71 (0.38, 7.62)	(0.88, 4.80)		
(NAb_SF162LS-chngpt)+	0.40 (0.10, 1.56)	(0.14, 1.38)	0.67 (0.51, 0.88)	(0.27, 0.87)
chngpt	6.09 (2.75, 9.42)	(4.81, 7.33)	7.37 (5.14, 9.61)	(6.02, 10.46)