This article is the manuscript version of:
Journal of the American Statistical Association, Volume 101, No. 474 (June 2006), pp. 542-553 (with *erratum*).[1] DOI: 10.1198/016214505000001177

Bent-Cable Regression Theory and Applications

Grace Chiu [1]
Richard Lockhart [2]
Richard Routledge [2]

[1] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada.
[2] Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, B.C., V5A 1S6, Canada.

[1] *Erratum* for journal version: Definition of $\mathcal{X}_2$ immediately before Section 3.3 should be $\mathcal{X}_2 = \mathcal{X}_2^r \cup (r, \infty)$.

Author Footnote

Grace Chiu is Assistant Professor (Email: *gchiu@uwaterloo.ca)*, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo ON, N2L 3G1; Richard Lockhart is Professor and Graduate Studies Program Chair (Email: *lockhart@stat.sfu.ca*), and Richard Routledge is Professor and Chair (Email: *routledg@stat.sfu.ca*), Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby BC, V5A 1S6. This research has been funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) through a Postgraduate Scholarship and a Postdoctoral Fellowship to G. Chiu and Discovery Grants to R. Lockhart and R. Routledge. The authors thank the Editor, Associate Editor, and referees for their valuable suggestions; and Professor Jerry Lawless, Department of Statistics and Actuarial Science, University of Waterloo, for his suggestions of reference material.

Abstract and Keywords

We use the so-called bent-cable model to describe natural phenomena which exhibit a potentially sharp change in slope. The model comprises two linear segments, joined smoothly by a quadratic bend. The class of bent cables includes, as a limiting case, the popular piecewise-linear model (with a sharp kink), otherwise known as the broken stick. Associated with bent-cable regression is the estimation of the bend-width parameter, through which the abruptness of the underlying transition may be assessed. We present worked examples and simulations to demonstrate the regularity and irregularity of bent-cable regression encountered in finite-sample settings. We also extend existing bent-cable asymptotics which previously were limited to the basic model with known linear slopes of 0 and 1, respectively. Practical conditions on the design are given to ensure regularity of the full bent-cable estimation problem, if the underlying bend segment has non-zero width. Under such conditions, the least-squares estimators are shown (i) to be consistent, and (ii) to asymptotically follow a multivariate normal distribution. Furthermore, the deviance statistic (or the likelihood ratio statistic, if the random errors are normally distributed) is shown to have an asymptotic chi-squared distribution.

Keywords: Asymptotic theory; Change points; Least squares; Maximum likelihood; Segmented regression

1. INTRODUCTION

In regression analysis, some natural phenomena call for models which exhibit a structural change, sometimes in the form of a difference in slopes. In such instances, the cause and onset of the change are often of major interest.

For example, Figure 1 portrays the declining abundance of sockeye salmon (*Oncorhynchus nerka*) in Rivers Inlet, British Columbia. By the year 2000, this population had declined from being one of the largest in Canada to an endangered remnant. (The data were obtained from Fisheries and Oceans Canada, Pacific Region.) To this day, researchers remain uncertain over the timing and cause of the collapse, and the abruptness of its onset. To address the former question, a resource manager would commonly fit a piecewise-linear model, the so-called broken stick, to these data for estimating the unknown change point. The estimated date of onset would sometimes be used to assist in identifying the source of the decline.

Applications of the broken stick in biological studies for estimating the onset of change also appear in, for instance, Naylor and Su (1998), Barrowman and Myers (2000), and Neuman, Witting and Able (2001). This sharply kinked line is particularly appealing in its structural simplicity. However, Chiu, Lockhart and Routledge (2005) and others (e.g. Wigglesworth 1972; Brown 1987; Jones and Handcock 1991; Routledge 1991) have pointed out that researchers are often tempted to conclude an abrupt onset from a broken-stick fit, even when there is little solid theory to justify the abruptness. For example, the interpretation of an abrupt onset of decline in species abundance made this way could lead to inappropriate conservation measures.

In this instance, it is important to assess the abruptness of change. Chiu et al. (2002, 2005) propose using the bent-cable model to relax the *a priori* assumption of abruptness associated with the broken stick. The bent cable generalizes the broken stick while retaining its simple structure; therefore, it is a more flexible model for describing natural phenomena that exhibit a change. This previously unnamed linear-quadratic-linear model was invented by Tishler and Zang (1981) as a numerical device to handle the broken stick's non-differentiable kink. Chiu et al. (2002, 2005) have since

provided large-sample maximum likelihood (ML) and least-squares (LS) estimation theory (assuming normal errors with known, constant variance for ML but otherwise for LS) for the *basic case* whose linear phases have known, fixed slopes of 0 and 1, respectively. However, practical settings call for the *full* bent cable with free slopes, intercept, and transition parameters. Our article provides this extension and applies it to illustrate the often unfounded abruptness assumption for real-life phenomena.

Bent-cable regression theory is complex due to non-differentiability of the model's (hence, likelihood's) first partial derivatives. Some of the earliest authors to address non-differentiability difficulties in regression were Hinkley (1969, 1971) and Feder (1975). Hinkley acknowledged the presence of unidentifiable model parameters in testing the null one-phase linear model against the broken stick, and suggested empirical evidence of an asymptotic distribution for the classic F-statistic without formal proof. Feder considered, somewhat unorthodoxly, a vast class of continuous models in which the asymptotics are radically different depending on an odd or even order of smoothness (number of continuous derivatives plus one) for the underlying function.

Recent articles on theory for segmented models include Bhattacharya (1990) and Hušková (1998) for the broken stick; and Gallant (1974, 1975), Hušková and Steinebach (2000), Ivanov (1997), Jarušková (1998a,b, 2001), and Rukhin and Vajda (1997) for multiphase non-linear models. All but Bhattacharya (1990) assume, among other regularity conditions, a bounded or compact parameter space, and/or evenly-spaced regressors. Ivanov (1997) and Rukhin and Vajda (1997) further assume a twice-differentiable model. In contrast, we argue in Sections 2 and 3 that, with slight modifications, the unbounded parameter space and the set of general and structurally simple design conditions of Chiu et al. (2005) suffice to establish regularity for the once-differentiable full bent cable. Under such conditions, a directional Hessian (adapted from Chiu et al. 2005) is shown to overcome non-differentiability of the score function in proving standard asymptotic results for the parameter estimator and deviance statistic. This directional Hessian is used repeatedly and meticulously throughout the proofs, making the mathematics non-standard. For more concise proofs, the

idea of differentiability in quadratic mean by Le Cam (1970) might be adapted to the current context; see Pollard (1997) for a recent exposition of the technique in the context of i.i.d. observations. Rather than seeking a general formulation of this idea in a regression context, we instead pursue the direct approach of Chiu et al. (2005).

Chiu et al. (2002) have shown that the case of a missing bend segment in the underlying cable defines an irregular boundary problem, with impractically complex asymptotics and a convergence rate of no better than $n^{-1/3}$. Therefore, our article focusses on full bent-cable asymptotics assuming a non-zero underlying bend width. Examples from Section 4 illustrate an alternative technique to formal hypothesis testing for statistically distinguishing between a broken stick and a bent cable.

2. THE BENT-CABLE MODEL

We denote the bent-cable model by f, the covariate by x, and the vector of regression parameters by $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \tau, \gamma)$. To construct f, first consider the basic bent cable, q, from Chiu et al. (2005):

$$q(x;\, \tau, \gamma) = \frac{(x - \tau + \gamma)^2}{4\gamma}\, \mathbf{1}\big\{|x - \tau| \le \gamma\big\} + (x - \tau)\, \mathbf{1}\big\{x > \tau + \gamma\big\}\,.$$

We write the full bent cable as

$$f_{\boldsymbol{\theta}}(x) \equiv f(x;\, \beta_0, \beta_1, \beta_2, \tau, \gamma) = \beta_0 + \beta_1\, x + \beta_2\, q(x;\, \tau, \gamma)\,. \tag{1}$$

Note that the parameterization (1) is linear in the β_j's, but non-linear in the transition or bend parameters, τ and γ (center and half-width of bend, respectively). Given a sequence of covariate values, $\{x_i\}_{i=1}^n$, our regression model is

$$Y_i = f_{\boldsymbol{\theta}}(x_i) + \varepsilon_i\,, \qquad i = 1, \ldots, n \tag{2}$$

where the ε_i's are i.i.d. random errors with mean 0 and variance σ^2.

2.1 Making Inference

We estimate the underlying model parameter, $\boldsymbol{\theta}_0$=$(b_0, b_1, b_2, \tau_0, \gamma_0)$, by the LS estimator (LSE), $\widehat{\boldsymbol{\theta}}_n$, which minimizes over a domain Ω the error sum-of-squares (ESS) function,

$$S_n(\boldsymbol{\theta}) = \sum_i \left|Y_i - f_{\boldsymbol{\theta}}(x_i)\right|^2 .$$

In the Appendix, we prove the results of Section 3.3 for a bounded Ω=$\prod_{j=0}^{3}[-M_j, M_j]\times[0, M_4]$. In practice, the unbounded Ω=$\mathbb{R}^2\times\prod_{j=2}^{3}[-M_j, M_j]\times[0, M_4]$ or Ω=$[-M_0, M_0]\times\mathbb{R}\times[-M_2, -\epsilon_2]\cup[\epsilon_2, M_2]\times[-M_3, M_3]\times[0, \infty)$, where $\epsilon_2(>0)$ is tiny, may be considered without affecting the asymptotic properties (see Chiu 2002, pp. 99–105). In the case of multiple minimizers of S_n, one can take the approach of Chiu et al. (2005) for defining a unique $\widehat{\boldsymbol{\theta}}_n$. For σ^2 unknown, we estimate it by the minimized error mean-square,

$$\widehat{\sigma}_n^2 = \overline{S}_n(\widehat{\boldsymbol{\theta}}_n) = \frac{1}{n} S_n(\widehat{\boldsymbol{\theta}}_n) . \tag{3}$$

We first consider normally distributed ε_i's, so that LS estimation of $\boldsymbol{\theta}_0$ and variance estimation via (3) are equivalent to ML estimation. We can also then study the behavior of the log-likelihood function,

$$\ell_{n,\sigma^2}(\boldsymbol{\theta}) \equiv \ell_n(\boldsymbol{\theta}; \sigma^2) = -\frac{1}{2}\left\{\frac{1}{\sigma^2} S_n(\boldsymbol{\theta}) + n \ln \sigma^2 + \ln(2\pi)\right\} . \tag{4}$$

As the nature of the transition is our main focus, we employ the method of profiling (see McCullagh and Nelder 1989), and examine the so-called profile deviance surface over the (τ, γ)-plane. For a given dataset, the profile likelihood is $\ell_n^P(\tau, \gamma) \equiv \max_{\beta_0,\beta_1,\beta_2,\sigma} \ell_{n,\sigma^2}(\boldsymbol{\theta})$, and the profile deviance surface is $2[\ell_n^P(\tau, \gamma) - \ell_n^P(\widehat{\tau}_n, \widehat{\gamma}_n)]$. The height at any point on this surface is a *deviance drop* (negative). When evaluated at the true but unknown (τ_0, γ_0), the absolute deviance drop (i.e. profile deviance statistic) is asymptotically χ_2^2-distributed under some conditions (see Section 3.3, Theorem 3). Truncate the surface along the vertical axis at, for instance, -$\chi_2^2(0.05)$=-5.99, and an approximate 95% confidence region (CR) for (τ_0, γ_0) is formed by all those (τ, γ)-pairs enveloped under the truncation. This approximation is valid when the truncated surface is quadratic (paraboloidal), although empirical evidence from Section 5 suggests that an $F_{2,n-2}$-based critical value may improve the coverage probability in such instances. Note that if $W \sim F_{p,q}$ for q large, then pW is approximately χ_p^2-distributed. This F-based adjustment is further justified for non-linear regression in various articles cited by Cook and Weisberg (1990).

The normality assumption can be removed without affecting the validity of this method, provided that the sample is sufficiently large. The details are stated in Theorems 2 and 3 of Section 3.3 below.

3. LEAST-SQUARES ASYMPTOTICS

Due to the impractical asymptotics in the case of $\gamma_0 = 0$ (Chiu et al. 2002), we only consider the case of a strictly positive γ_0.

3.1 Parameter Space and Design Conditions

We consider the very practical open regression domain, $\mathcal{X} = \mathbb{R}$, and the parameter space, Ω, of Section 2.1. Conditions [A] to [F] in Appendix A.1 are placed on the covariate design to ensure regularity for full bent-cable regression. In essence, these conditions require the following: (1) *Five detached regions containing non-trivial fractions of data: one strictly in the bend, and two strictly in each linear phase (Conditions [A] and [B]).* This "$\{2, 1, 2\}$-configuration" is required for consistency. (2) *No observations exactly at the join points (Condition [C]) or accumulation of data in any immediate vicinity of a join point (Condition [D]).* This condition guarantees an asymptotically well-behaved second derivative, or Hessian, for the ESS function. (3) *Reasonably small average absolute and squared covariate values (Condition [E]).* This condition prevents the inclusion of extraordinarily influential covariate values, thereby ensuring that the ESS gradient and its covariance matrix are asymptotically well-behaved. (4) *A strengthened version of (3) (Condition [F]), required if normality of the random errors is not assumed:* the largest x in absolute value must grow more slowly than the square root of the sample size. (In general, asymptotic normality of the LSE comes from an asymptotically normal ESS gradient function. However, the latter fails if the furthest covariate value puts too much weight on an ε_i that is non-normally distributed.) In practice, (3) and (4) are satisfied if the x's are, say, restricted within a compact set, or generated from any probability distribution with a finite variance.

3.2 Notation

In addition to those defined in Section 2.1, the crucial quantities involved in the

main theorems of this articles are: $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}) = \boldsymbol{\nabla}\ell_{n,\sigma^2}(\boldsymbol{\theta})$, where $\boldsymbol{\nabla}$ is taken with respect to $\boldsymbol{\theta}$; $\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}) = \mathrm{Cov}_{\boldsymbol{\theta}_0}\big[\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta})\big]$; $R_{i+} = \big\{\boldsymbol{\theta} \in \Omega : \gamma = \tau - x_i\big\}$; and $R_{i-} = \big\{\boldsymbol{\theta} \in \Omega : \gamma = x_i - \tau\big\}$. If the ε_i's are non-normally distributed, then ℓ_{n,σ^2} is not the log-likelihood function, but merely a label for the anti-derivative of $\boldsymbol{U}_{n,\sigma^2}$. Except for a proportionality constant, $\boldsymbol{U}_{n,\sigma^2}$ is essentially the gradient of S_n. Thus, we examine the properties of $\boldsymbol{U}_{n,\sigma^2}$, whether we consider ML with normal ε_i's or LS with i.i.d. non-normal errors. Complexity arises from the non-differentiability of $\boldsymbol{U}_{n,\sigma^2}$ along the hyper-rays, $R_{i\pm}$'s. Similar to Chiu et al. (2005), we replace $\mathbb{V}_{n,\sigma^2} = \boldsymbol{\nabla}\boldsymbol{U}_{n,\sigma^2}$ by a directional Hessian, $\mathbb{V}^+_{n,\sigma^2}$ which is well-defined everywhere on Ω:

Suppress the dependence on σ^2 in the notation and let

$$U_{nk}(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_k}\ell_{n,\sigma^2}(\boldsymbol{\theta})\,, \qquad V^+_{n,jk}(\boldsymbol{\theta}) = \lim_{h\downarrow 0}\frac{\partial}{\partial\theta_j}U_{nk}(\theta_1,\ldots,\theta_{j-1},\theta_j+h,\theta_{j+1},\ldots,\theta_5)$$

where $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_5)=(\beta_0,\beta_1,\beta_2,\tau,\gamma)$. Then $\mathbb{V}^+_{n,\sigma^2}(\boldsymbol{\theta})$ is the matrix whose (j,k)-th element is $V^+_{n,jk}(\boldsymbol{\theta})$.

Also needed in relevant lemmas are $\boldsymbol{\theta}_0=(\theta_{01},\ldots,\theta_{05})=(b_0,b_1,b_2,\tau_0,\gamma_0)$, $\Theta_r=\big\{\boldsymbol{\theta} : |\boldsymbol{\theta}-\boldsymbol{\theta}_0| \le r\big\}$, $\mathcal{X}_0=[\tau_0-\gamma_0+\delta_{10},\tau_0+\gamma_0-\delta_{10}]$, $\mathcal{X}_{-1}=[\tau_0-\gamma_0-\delta_{12},\tau_0-\gamma_0-\delta_{11}]$, $\mathcal{X}_1=[\tau_0+\gamma_0+\delta_{11},\tau_0+\gamma_0+\delta_{12}]$, $\mathcal{X}^r_{-2}=[-r,\tau_0-\gamma_0-\delta_{13}]$, $\mathcal{X}_{-2}=(-\infty,-r)\cup\mathcal{X}^r_{-2}$, $\mathcal{X}^r_2=[\tau_0+\gamma_0+\delta_{13},r]$, and $\mathcal{X}_2=\mathcal{X}^r_2\cup(r,\infty)$, where δ_{1j}'s are tiny constants from conditions [A] and [B] of Appendix A.1, and $r>0$ is arbitrary.

3.3 Formal Statements of Results

Theorems 1 to 3 below and their proofs (see Appendix A.3) are generalizations of Theorems 1 and 2 in Chiu et al. (2005). Formal proofs of relevant lemmas appear in Chiu (2002).

The first result is consistency, of which an essential ingredient is Lemma 1 which implies that the bent-cable model is identifiable under design condition (1).

Lemma 1 (Identifiability). *Given are $w_i \in \mathcal{X}_i$ for all $i = 0, \pm1, \pm2$. Then, $f_{\boldsymbol{\theta}}(w_i) = f_{\boldsymbol{\theta}_0}(w_i)$ for all $i = 0, \pm1, \pm2$ implies $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.*

To prove that $f_{\boldsymbol{\theta}_0}$ is identifiable by the w_i's in a $\{2,1,2\}$-configuration, consider all

twenty-one five-point configurations for the candidate cable, $f_{\boldsymbol{\theta}}$. Convexity and smoothness constraints of a bent-cable function (i) prohibit $f_{\boldsymbol{\theta}}$ to go through the given $\big(w_i, f_{\boldsymbol{\theta}_0}(w_i)\big)$-pairs in any non-$\{2,1,2\}$-configuration, and (ii) together with (i), force $f_{\boldsymbol{\theta}}$ and $f_{\boldsymbol{\theta}_0}$ to coincide everywhere.

Theorem 1 (Consistency). *Under design conditions (1) and (3), $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\sigma}^2_n$ are consistent estimators of $\boldsymbol{\theta}_0$ and σ^2, respectively.*

The next result is asymptotic normality, which relies on the following lemmas.

Lemma 2 (Taylor-type expansion). *For all $\boldsymbol{\theta} \in \Omega$, we have*

$$\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}) = \boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0) + \left(\int_0^1 \big[\mathbb{V}^*_{n,\sigma^2}(\boldsymbol{\theta}, t)\big]^T \, dt \right) (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

where the matrix components of $\mathbb{V}^*_{n,\sigma^2}(\boldsymbol{\theta}, t)$ are those of $\mathbb{V}^+_{n,\sigma^2}$, each evaluated at a point of the form $(\theta_1, \ldots, \theta_{j-1}, \theta_{0j} + t(\theta_j - \theta_{0j}), \theta_{0,j+1}, \ldots, \theta_{05})$ for different values of $j = 1, 2, \ldots, 5$. The exact form of $\mathbb{V}^*_{n,\sigma^2}$ is given by (6) in Appendix A.2.

This one-term expansion handles the non-differentiability of $\boldsymbol{U}_{n,\sigma^2}$ along the $R_{i\pm}$'s. The use of $\mathbb{V}^*_{n,\sigma^2}$ (a variant of $\mathbb{V}^+_{n,\sigma^2}$) replaces $\mathbb{V}_{n,\sigma^2}$ and its gradient which are often required to exist in standard proofs of asymptotic normality for smoother models.

Lemma 3. *Given are design conditions (1) to (3), and a sequence $\delta_n \downarrow 0$. Then,*

$$\forall\ j,k = 1, \ldots, 5, \qquad \sup_{\boldsymbol{\theta}\ \in\ \Theta_{\delta_n}} \frac{1}{n} \Big| I_{n,jk}(\boldsymbol{\theta}_0) + V^+_{n,jk}(\boldsymbol{\theta}) \Big| \xrightarrow{\mathrm{P}} 0 \qquad \text{as } n \longrightarrow \infty$$

where $I_{n,jk}$ denotes the (j,k)-th component of $\mathbb{I}_{n,\sigma^2}$.

Lemma 4. *Assume that $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. zero-mean random errors with constant finite variance, σ^2. Under design conditions (1), (3), and (4), the Lindeberg-Feller Central Limit Theorem applies. That is, for all fixed non-zero $\boldsymbol{w} \in \mathbb{R}^5$,*

$$\frac{\boldsymbol{w}^T \boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)}{\sqrt{\boldsymbol{w}^T \left[\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)\right] \boldsymbol{w}}} \xrightarrow{\mathcal{L}} N(0,1) \qquad \text{as } n \longrightarrow \infty\,.$$

The key ingredient of the lemma is that under condition (4), the summands of $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)$

satisfy the Lindeberg Condition in the multivariate sense. The other conditions appear in the lemma merely for a technical purpose: $n^{-1}\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)$ must be positive definite for all sufficiently large n (see Theorem 2, Assertion 1 below).

Theorem 2 (Asymptotic Normality). *Under design conditions (1) to (3),*

1. *the matrix $n^{-1}\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)$ is positive definite for all sufficiently large n, and similarly, $P_{\boldsymbol{\theta}_0}\Big\{n^{-1}\mathbb{I}_{n,\sigma^2}\big(\widehat{\boldsymbol{\theta}}_n\big)$ is positive definite$\Big\} \longrightarrow 1$;*

2. *if $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2)$, then both $\sqrt{n}\Big[n^{-1}\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)\Big]^{1/2}\big(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\big)$ and $\sqrt{n}\Big[n^{-1}\mathbb{I}_{n,\sigma^2}\big(\widehat{\boldsymbol{\theta}}_n\big)\Big]^{1/2}\big(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\big)$ converge in distribution to a standard five-variate normal random variable;*

3. *design condition (4) can replace the normality assumption in Assertion 2;*

4. *Assertions 1 to 3 hold true when $\widehat{\sigma}_n^2$ replaces σ^2 in the expression of $\mathbb{I}_{n,\sigma^2}$.*

Assertions 1 and 2 here are essentially Parts 1 to 4 of Theorem 1 in Chiu et al. (2005), except for what is now a five-dimensional problem (assuming σ^2 known). In Assertion 3, normality of the ε_i's is removed. However, the LSE, $\widehat{\boldsymbol{\theta}}_n$, remains to be a solution of $\boldsymbol{U}_{n,\sigma^2}$, and the Taylor-type expansion of $\boldsymbol{U}_{n,\sigma^2}\big(\widehat{\boldsymbol{\theta}}_n\big)$ via Lemma 2 is unaffected. As $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)$ is asymptotically normal by Lemma 4, Lemmas 2 and 3 (uniform closeness of $-\mathbb{V}^{+}_{n,\sigma^2}$ to $\mathbb{I}_{n,\sigma^2}$) imply an asymptotically normal $\widehat{\boldsymbol{\theta}}_n$. Finally, the value of $\widehat{\boldsymbol{\theta}}_n$ does not depend on σ^2. Hence, Assertions 1 to 3 are affected by an estimated σ^2 only through the formula of $\mathbb{I}_{n,\sigma^2}$. Then, Assertion 4 is true due to the consistency of $\widehat{\sigma}_n^2$.

Theorem 3 (χ^2-limit). *Under design conditions (1) to (4), each deviance statistic below has a limiting χ^2 distribution, with its degrees-of-freedom in parentheses:*

1. *$G_{n,\sigma^2} = 2\big[\ell_n(\widehat{\boldsymbol{\theta}}_n;\sigma^2) - \ell_n(\boldsymbol{\theta}_0;\sigma^2)\big]$ (df=5), in the case of a known σ^2;*

2. *$G_n = 2\big[\ell_n(\widehat{\boldsymbol{\theta}}_n;\widehat{\sigma}_n^2) - \ell_n(\boldsymbol{\theta}_0;\widehat{\sigma}_n^2)\big]$ (df=5), in the case of an unknown σ^2 estimated by $\widehat{\sigma}_n^2$; and*

3. *$D_n = n\big[\ln\big(\widehat{\sigma}_n^{2^*}\big) - \ln\big(\widehat{\sigma}_n^{2'}\big)\big]$ under H^* (df=$p-q$, $0 \le q < p \le 5$) for testing H^* vs. H' in the case of an unknown σ^2, where p components of $\boldsymbol{\theta}_0$ are estimated*

under H', and q, under H^*; and $\widehat{\sigma}_n^{2*} = \overline{S}_n(\widehat{\boldsymbol{\theta}}_n^*)$, $\widehat{\sigma}_n^{2\prime} = \overline{S}_n(\widehat{\boldsymbol{\theta}}_n')$.

If we assume normal ε_i's here, then the three deviance statistics of Theorem 3 are likelihood-based; hence, Assertion 1 is merely an extension of Theorem 2, Part 5, in Chiu et al. (2005). That is, the χ^2-limit of G_{n,σ^2} results from an approximately quadratic deviance surface over a neighborhood of $\widehat{\boldsymbol{\theta}}_n$. (The key is the substitution of Lemma 2 into the usual one-term Taylor expansion of $\ell_{n,\sigma^2}(\widehat{\boldsymbol{\theta}}_n)$.) Again, condition (4) here can replace the normality assumption for the ε_i's without affecting the results of Lemma 2. Further generalizations, based on a consistent $\widehat{\sigma}_n^2$, to the case of an unknown σ^2 and the testing of full bent-cable hypotheses yield the latter two assertions.

Note that the overall tactic for proving Theorems 2 and 3 is largely standard; for example, see Serfling (1980). However, the repeated application of Lemmas 2 and 3 here to overcome non-differentiability in the majority of steps makes the proofs non-standard.

4. APPLICATIONS

To illustrate how bent-cable regression helps to assess the abruptness of change, we apply the method to four typical sets of observations: the Rivers Inlet Sockeye data previously shown in Figure 1, Sir Francis Galton's famous family stature data, exercise physiology data, and data from the physical sciences that provide a valuable contrast. All are examples of the sorts of change-point problems in which researchers have traditionally applied the broken-stick model.

Recall from Section 3.3 that large-sample ML theory is also valid for general non-linear LS estimation without the normality assumption in (2). Without loss of generality, below we give details based on the ML method only.

4.1 Abruptly Declining Salmon... Or Not?

Figure 1 depicts a relatively stable abundance of sockeye from 1980 until around 1993. The population has declined drastically since then. Did the decline begin abruptly around 1993? Or was it more gradual, possibly starting earlier?

Time series for abundance of Pacific salmon populations with fixed life spans often have strong autocorrelation structure. However, the population being considered has

a mixed life span of four to five years with large, unpredictable fluctuations in the age distribution. Hence, the autocorrelations here are relatively weak. We thus proceed to illustrate the method described in Section 2.1 and obtain the profile deviance surface in Figure 2. This surface peaks at the ML estimates $\widehat{\tau}$=1992 and $\widehat{\gamma}$=6. Hence, the estimated bend ranges from 1986 to 1998. Note that the surface is truncated at -$\chi^2_2(0.05)$=-5.99. Due to the surface's irregularity, this nominal 95% confidence level may not be trustworthy. However, the deviance values of the surface's upper triangular plateau roughly defined by the intersection of the regions $\tau - \gamma \leq$ 1986 (i.e. a smooth transition beginning by 1986) and $\tau + \gamma \geq$ 1998 (i.e. a transition ending no earlier than 1998) are so close to 0 that the corresponding (τ, γ)-pairs are almost certainly consistent with the data. For example, (τ, γ)=$(1990, 10)$ gives a purely quadratic fit, i.e. a cable whose bend stretches over the entire range of the data. The plateau yields many other models, including any cable whose bend begins at around 1986. Furthermore, the surface along the τ-axis (i.e. γ=0) has a peak at $\tau \approx$93. This local peak is not far below the height of the triangular plateau, and hence, (τ, γ)=$(1993, 0)$ — a sharp change in 1993 — is also highly consistent with the data. Thus, the decline in sockeye abundance could have been accelerating steadily over most of the time range shown. Or it could equally well have begun abruptly around 1993.

4.2 Could Galton's Bend be Smooth?

Wachsmuth, Wilkinson and Dallal (2003) cite the discovery by Hinkley (1971) of a kink in the parent-to-child relationship exhibited by Sir Francis Galton's family stature data from the 19th century. They subsequently apply *loess* and broken-stick fits to show non-linearity in similar data collected by Galton's disciple Karl Pearson. The authors argue that this "Galton's bend" in both Galton's and Pearson's datasets was due to the pooling of gender blocks. However, Hanley (2004) applies linear, quadratic, and cubic regressions, respectively, to the reverse relationship (child-to-parent) using Galton's data, and finds no statistical distinction among the fits.

How does bent-cable regression compare when applied to these data? Figure 3 shows Galton's original data, as reproduced by Hanley (2004). Here, we have adopted Han-

ley's practice of (1) omitting Galton's non-numerical entries, (2) multiplying female heights by 1.08, and (3) averaging within each family the resulting parents' heights. Overlaid on the scatterplot are two broken-stick fits where $\widehat{\tau}$=70.20 (solid lines) and $\widehat{\tau}$=71.50 (dotted lines), respectively; and a bent-cable fit where $\widehat{\tau}$=70.97 and $\widehat{\gamma}$=1.13 (dashed lines). Judging by eye, all three fits virtually coincide. And statistically, too, as indicated by the profile deviance surface in Figure 4(a). The deviance values for these three fits differ by less than 0.5. In fact, the overall best fit for these data is the solid-line broken stick (corresponding to the peak of the surface at (τ, γ)= $(70.2, 0)$); the next best fit is the dashed-line bent cable (rounded peak); and the third-best is the dotted-line broken stick (peak at (τ, γ)=$(71.5, 0)$). Many purely quadratic fits (upper triangular plateau, deviance around -2) are also consistent with the data. Equally good are two-phase quadratic-linear fits (upper left ridge) whose bends end at around 73", and two-phase linear-quadratic fits (upper right ridge) whose bends begin at around 69". However, purely linear fits (lower plateaus, deviance around -7) are significantly different from any of the above fits at an approximate 5% level. This agrees with findings by Hinkley and Wachsmuth et al. in pointing towards a bend. For the reverse (child-to-parent) regression, Figure 4(b) shows that deviance values are within 0.8 for the best broken-stick fit (also best overall), the best bent-cable fit (rounded peak on surface), and many purely quadratic fits (upper plateau) and linear-quadratic fits (upper right ridge). Here, purely linear fits (lower plateaus, deviance above -3) are also consistent with the data, thus agreeing with Hanley's findings. Altogether, our results are in line with the other authors', although the so-called "Galton's bend" could well have been smooth instead of kinked.

4.3 An Anaerobic Threshold?

The notion of an abrupt change also appears in the physiological sciences. For example, consider the relationship between blood lactate concentration versus oxygen uptake for an athlete engaged in a progressively demanding physical activity. At lower work intensities, one would expect a linear increase in lactate with increasing oxygen

uptake. However, when the work intensity increases to the point where metabolic homeostasis is disturbed, the slope of the lactate–oxygen relationship increases. The point when this change occurs has been called the lactate threshold, and is a key focus of training regimes (Antonutto and Di Prampero 1995; Weltman 1995). One currently used method for estimating the lactate threshold involves visual inspection for the point at which a plot of blood lactate concentrations versus some workload measure begins to curve upwards (Weltman et al. 1994; Vachon, Bassett and Clarke 1999; Moquin and Mazzeo 2000; Schneider, McLellan and Gass 2000). A more systematic method is to fit a broken stick to a graph of blood lactate versus workload (Beaver, Wasserman and Whip 1985; Kline 1997; Moquin and Mazzeo 2000), sometimes plotted on logarithmic scales (Beaver et al. 1985; Moquin and Mazzeo 2000).

Until recently, it was widely believed that carbon dioxide output could also be used to monitor what was then thought to be a similar "anaerobic threshold." Associated with this concept is a long-standing controversy over the abruptness of the anaerobic threshold (Jones and Handcock 1991; Routledge 1991), particularly when a broken stick is fitted to the data. The bent-cable model permits a direct evaluation of this controversial abruptness. We present a worked example.

An athlete's carbon dioxide output and oxygen uptake (mL/s) were monitored while he ran on a treadmill whose incline was regularly increased $\big($Figure 5(a)$\big)$. (The experiment was conducted according to a ramped workload protocol on a treadmill at the Science North Science Centre in Sudbury, Ontario. The data are available from the current authors upon request.) The general increasing trend in the CO_2–O_2 relationship obscures any subtle features. To accentuate these, we present detrended data, i.e. residuals from a linear fit $\big($Figure 5(b)$\big)$. This graph points to a change in the relationship somewhere in the vicinity of an oxygen uptake of 3,200. However, as in the previous two examples, the deviance surface (not shown here; see Chiu 2002) is irregular, and has an upper diagonal ridge roughly along $\tau - \gamma = 3,350$ for $\tau \geq 3,500$, again suggesting that numerous models are extremely good fits. One such fit has a bend that begins at an O_2-value of about 3,350 and continues through the remainder

of the values. The values between 3,450 and 3,600 along the τ-axis — a broken stick with a corner anywhere between such O_2-values — are also consistent with the data. The overall best fit is a cable whose bend ranges over O_2-values from 3,353 to 3,721. In this instance, the data do not favor a sharp break that would indicate an abrupt anaerobic threshold for the athlete.

4.4 Convincing Evidence of a Smooth Transition

For comparison, we present data from a physics experiment in Figure 6, first published in R. A. Cook's Ph. D. thesis at Queen's University, as cited by Seber and Wild (1989). (The data have also been analyzed by Bacon and Watts (1971) with a hyperbolic transition model and with other multiphase regression models discussed by Seber and Wild (1989)). Cook's experiment examines the behavior of stagnant-band-height of water as it flows down an incline at different rates. The relationship between band height and flow rate is known to exhibit a change, although the underlying nature of the change is unknown.

The small amount of chance scatter in this graph (Figure 6) is common only in the physical sciences. However, even with this improved resolution, it is hard to detect the nature of the transition from a mere visual inspection of the graph. When a bent cable is fitted to these data, we obtain overwhelming evidence for a smooth bend. The profile deviance surface $\big($Figure 7(a)$\big)$ provides a vivid contrast in its regularity to the previous examples, and it rules out any broken-stick fit (γ=0) at any reasonable confidence level such as 95%. In fact, the surface remains highly paraboloidal and well excludes γ=0 even for a cutoff value of -10 (not shown), which corresponds to a χ^2- or F-based confidence level of more than 98%. Hence, the evidence for a smooth transition is overwhelming. The actual best-fitting cable has a bend ranging between log-flow rates of -0.373 and 0.484 $\big($Figure 7(b)$\big)$.

4.5 Implications

The first three datasets are of a biological nature, and have at worst moderate sample sizes (n=21, 934, 28, respectively) with typical amounts of chance scatter; yet,

their deviance surfaces are highly irregular. In contrast, the physics data demonstrate the applicability of precise asymptotics in the case of a sample whose size (n=29) is large given the exceptionally small response errors. Thus, the asymptotic approximations appear not to be reliable for many datasets with sample sizes and residual terms typical of biological applications. Nonetheless, the deviance surface may well show in these instances that the data are thoroughly consistent with a broad range of behavior around the hypothetical change point. The technique can therefore be highly useful in assessing claims of an abrupt onset of change.

5. SIMULATIONS

We conducted two groups of simulations. The first group (5,000 experiments per set) examined the empirical coverage of nominal 95% CRs for the transition parameters, (τ_0, γ_0), when the profile deviance surface exhibited regularity (and the asymptotics were applicable). Two types of CRs were considered: one based on (A) the profile deviance statistic with an approximate χ^2-distribution deduced from Theorem 3, and another based on (B) a Wald statistic derived from Theorem 2. To assess the effect of non-normality on coverage in finite-sample settings, CRs of the two types were compared given response errors with normal, uniform, and t_5 distributions, respectively. The latter two were chosen to represent both lighter and heavier tailed error distributions. In particular, a t-distribution with df=5 has tails that are heavy enough to generate occasional outliers, but light enough to have a finite variance. The second group (100 experiments each) was used to further explore the problem's irregularity such as when observations exhibit moderate scatter. Each generated profile deviance surface, truncated at a nominal 95% confidence level, was visually scrutinized. The small number of runs was due to the painstaking nature of this visual inspection, but it suffices here for the purpose of illustrating a qualitative assessment, instead of a quantitative measure, of the theoretical notion of regularity.

5.1 Group 1

The true model parameters were b_0=b_1=τ_0=0, b_2=1, and γ_0=0.5. Similar to the

band-height data of Section 4.4, each experiment consisted of n=31 observations, where covariate values were placed between -1.5 and 1.5 at intervals of 0.1, and response errors were sampled from a normal, uniform, or t_5 distribution with mean 0 and standard deviation (SD) 0.015. Altogether, 5,000 such experiments were run for each type of error distribution. In each experiment, an *approximated* profile deviance surface was produced over a fine Euclidean grid formed by ranging τ over -0.07 to 0.07 at intervals of 0.005, and γ over 0.37 and 0.65 at intervals of 0.01. This gridding technique produced $(\tilde{\tau}, \tilde{\gamma})$, an estimate of (τ, γ) to the nearest grid point (slightly less precise but more readily obtained than $(\widehat{\tau}, \widehat{\gamma})$ produced by a Gauss-Newton algorithm); hence, the deviance surface based on $(\tilde{\tau}, \tilde{\gamma})$ was an approximation (see Chiu 2002 pp. 14–17). To produce a Type (A) CR in practice, this surface would then be truncated at -$\chi^2_2(0.05)$=-5.99 or -$2F_{2,n-2}(0.05)$=-6.655. To assess coverage, it was only necessary to determine if the (approximated) deviance, evaluated at (τ_0, γ_0), exceeded these critical values. For (B), consider $\widehat{\boldsymbol{\theta}}$ of an approximate five-variate normal distribution with mean $\boldsymbol{\theta}_0$ and covariance $\mathbb{S}_{\sigma^2}(\boldsymbol{\theta}_0) \equiv [\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)]^{-1}$. Thus, $X^2 \equiv (\widehat{\boldsymbol{\theta}}\text{-}\boldsymbol{\theta}_0)^T [\mathbb{S}_{\sigma^2}(\boldsymbol{\theta}_0)]^{-1} (\widehat{\boldsymbol{\theta}}\text{-}\boldsymbol{\theta}_0)$ has an approximate χ^2_5-distribution, and $\boldsymbol{\theta}_0$ is covered by the F-based 95% CR if $X^2 \leq 5F_{5,n-5}(0.05)$. In the simulations, we reduced the dimension from 5 to 2: $\boldsymbol{\theta}_0$ was replaced by $(\tau_0, \gamma_0)=(0, 0.5)$; $\widehat{\boldsymbol{\theta}}$ by $(\tilde{\tau}, \tilde{\gamma})$; $\mathbb{S}_{\sigma^2}(\boldsymbol{\theta}_0)$ by the lower-right 2×2 submatrix of $\mathbb{S}_{\tilde{\sigma}^2}(\tilde{\boldsymbol{\theta}})$, where $\tilde{\sigma}^2$ and $\tilde{\boldsymbol{\theta}}$ were from the five-parameter fit based on $(\tilde{\tau}, \tilde{\gamma})$; and $5F_{5,n-5}(0.05)$ by $2F_{2,n-2}(0.05)$. Also of interest was inclusion of broken sticks (γ=0) by the CRs. To this end, in each experiment, profile deviance values for (A) and X^2 values for (B) were produced for γ=0 over the above τ-range; any such deviance value exceeding -5.99 or -6.655 would indicate a broken stick covered by the Type (A) CR, and similarly by the Type (B) CR if any such X^2 value was less than 6.655.

Type (A), or likelihood-ratio-based, CRs are often deemed more reliable than Type (B), or Wald, CRs for non-linear regression parameters; see Cook and Weisberg (1990), for instance. Our simulations confirm this notion: despite exceptionally tight scatter, Table 1 shows a Type (B) coverage of 93.42% at best for a nominal 95% level, regardless of error distribution. On the other hand, Type (A) coverage was up to 93.94%

when using the F-based cutoff and/or when errors were normal. (One may expect to observe slightly different coverages for CRs obtained over a slightly different (τ, γ)-grid.) Lower Type (B) coverages indicate that the Wald method perceived more information than what the data actually contained. This was further demonstrated by the two cases in which broken sticks were covered by Type (A) CRs but excluded from Type (B) CRs, when errors were t_5-distributed. Both were borderline cases in that, given γ=0, the respective maximum deviance values were -6.46 and -5.74. Finally, we observed that all 15,000 profile deviance surfaces (not shown) were paraboloidal regardless of error distribution. These confirm the regularity notion when design conditions (1) to (4) are satisfied with n sufficiently large and/or σ sufficiently small.

5.2 Group 2

Here, the experiments were broken down into cases where the underlying model was (i) a cable whose bend ranged over the middle one-third of the dataset, and (ii) a broken stick whose kink divided the dataset equally. We ran Sets (i a), (ii a), and (ii b): n=21 and σ=0.65 in (a) (similar to the sockeye data), and n=31 and σ= 0.015 in (b) (similar to the band-height data). Note that Set (i b) already appears in Group 1 above. To assess how estimation accuracy would be affected by proximity of data to the underlying kink, Set (ii b) was further broken down into (ii b1), in which an x-value coincided with the underlying τ_0; and (ii b2), in which x-values from (b1) were translated so that τ_0 lay one-fifth of the distance between the middle two x-values. Response errors in each set were sampled from a normal distribution. All sets had 100 runs, in each of which b_0=b_1=0, b_2=1, and equidistant x-values ranged from 80 to 100 for (i a), 81 to 101 for (ii a), -1.5 to 1.5 for (ii b1), and -1.48 to 1.52 for (ii b2). In addition, (τ_0, γ_0) was $(90, 3.2)$ for (i a), $(90, 0)$ for (ii a), and $(0, 0)$ for (ii b).

All four sets of experiments were expected to exhibit irregularity: Sets (a) had too much scatter (see Section 4.5), while Sets (ii b) had an $n^{-1/3}$ convergence rate at best (see Chiu et al. 2002). Thus, we focussed on three main aspects of the profile deviance surfaces, truncated at the nominal 95% confidence cutoff of -5.99. First, we

examined the shape of the surface, which provides a qualitative assessment of the methodology's regularity in practice. Second, we assessed the qualitative accuracy of the best-fitting model (for example, $\widehat{\gamma} = 0$ would yield a qualitatively incorrect best fit for Set (i a)). This helps to identify the circumstances under which $\widehat{\gamma}$ alone is a reliable indication of the abruptness of change. Third, we examined the coverage of broken sticks by the 95% CRs for (τ_0, γ_0). This provides insight into the effectiveness of formally assessing the abruptness of change via standard statistical means.

Table 2 provides a summary. In particular, when chance scatter was more substantial (Sets (a)) — whether the underlying structure was kinked or smooth — over 60% of the surfaces exhibited irregularity in the form of plateaus and/or ridges such as appear in Figures 2 and 4, pointing to vast collections of bent-cable fits (often including broken sticks) which were highly consistent with the data. The remaining surfaces were mostly smooth "half-domes" naturally truncated along γ=0; thus, they all included broken sticks. The overall rates of (1) exclusion of broken sticks and (2) qualitatively correct best fits were low. Altogether, these results reflect the inherent difficulty in clearly distinguishing between abrupt and smooth changes from typical sets of observations. When the underlying structure was kinked (Sets (ii)), the asymptotics from Section 3 were inapplicable, and full paraboloids were not expected in this case. For data with little chance scatter (Sets (ii b)), virtually all surfaces were half-domes and thus included broken sticks. Under 70% of them yielded qualitatively correct best fits. However, no ridges or plateaus were present. When a design point lay exactly at the underlying kink (Set (ii b1)), the percentage of qualitatively correct best fits was statistically higher than that when no design point coincides with the kink (Set (ii b2)). This reflects a design condition for consistency when γ_0=0 (Chiu et al. 2002, 2005).

6. CONCLUSION

Our simulation results and the analyses from Section 4 suggest the following. When the underlying bent cable has a non-trivial bend segment, only data containing an exceptional amount of information about this smooth structure lead to a profile deviance surface that exhibits the kind of regularity that can be used as evidence against a

kinked structure. When either this information is deficient, or the transition is instantaneous, then the deviance surface will be irregular. In this case, the true structure of the transition would be confounded, and claims of an abrupt threshold, premature.

In practice, uncontrollable fluctuations are more substantial in typical biological situations; data of this sort cannot provide definitive evidence as did Cook's data. In such instances, both the broken stick and its generalization, the bent cable, tend to be consistent with the data. Conventional practice then is to adopt the slightly more parsimonious broken stick as an adequate description of the data. However, a broken-stick fit may lead to misinterpretations where the investigator attempts to attribute the estimated threshold to some source even in the absence of solid auxiliary evidence supporting an abrupt change. As the existence of a sharp threshold cannot be tested with any reliability, descriptions of distinct phases or regimes should be viewed simply as partitions of convenience that are not supportable by statistical analyses. Much more extensive data or other auxiliary evidence related to the potentially abrupt change would be needed. For example, in the case of a declining fish population, knowledge of an abrupt change in, say, the abundance of a predator known to feed on this fish, or a key aspect of habitat quality affected by logging activity, could provide such auxiliary evidence. In contrast, a bent-cable fit for these data would suggest numerous sources that took place over a period of years, or a single source whose continuous influence prompted a gradual decline. Thus, interpreting the decline onset as abrupt without any solid evidence could lead to inappropriate conservation measures.

Bent-cable regression can be applied as an alternative to classic change-point techniques which do not allow for possible smoothness in the transition between phases. We recommend its use in cases which lack the auxiliary knowledge necessary to support the intrinsic abruptness assumption of kinked models.

APPENDIX: MATHEMATICAL DETAILS

A.1 Design Conditions

For $\delta, \delta_1, \delta_2 > 0$ and $\xi_n \downarrow 0$, define

$$
\begin{aligned}
b_\ell(\delta) &= \liminf_{n\longrightarrow\infty} \frac{1}{n} \sum \mathbf{1}\big\{x_i \le \tau_0 - \gamma_0 - \delta\big\}\,, \\
b_c(\delta) &= \liminf_{n\longrightarrow\infty} \frac{1}{n} \sum \mathbf{1}\Big\{|x_i - \tau_0| \le \gamma_0 - \delta\Big\} \\
b_r(\delta) &= \liminf_{n\longrightarrow\infty} \frac{1}{n} \sum \mathbf{1}\Big\{x_i \ge \tau_0 + \gamma_0 + \delta\Big\} \\
c_\ell(\delta_1,\delta_2) &= \liminf_{n\longrightarrow\infty} \frac{1}{n} \sum \mathbf{1}\big\{x_i \in [\tau_0 - \gamma_0 - \delta_2, \tau_0 - \gamma_0 - \delta_1]\big\}\,, \qquad \delta_2 > \delta_1 > 0\,, \\
c_r(\delta_1,\delta_2) &= \liminf_{n\longrightarrow\infty} \frac{1}{n} \sum \mathbf{1}\big\{x_i \in [\tau_0 + \gamma_0 + \delta_1, \tau_0 + \gamma_0 + \delta_2]\big\}\,, \qquad \delta_2 > \delta_1\,, \\
\nu(\delta,p) &= \limsup_{n\longrightarrow\infty} \frac{1}{n} \sum |x_i|^p\, \mathbf{1}\big\{|x_i| > \delta\big\}\,, \qquad p = 0,1,2,\ldots\,, \\
\zeta_n(\xi_n) &= \frac{1}{n} \sum \mathbf{1}\left\{\Big||x_i - \tau_0| - \gamma_0\Big| \le \xi_n\right\}\,.
\end{aligned}
$$

Then, the design conditions below provide regularity for full bent-cable regression. Conditions [A], [C], and [D] are taken from Chiu et al. (2005) for the basic case. ([C] could be eliminated at the cost of some notational complexity.) We have modified their condition [B] for the full model here. The additional conditions for the full bent cable are [E] and [F].

[A] $\exists\ \delta_{10} > 0$ such that $c_0 \equiv b_c(\delta_{10}) > 0$.

[B] $\exists\ \delta_{13} > \delta_{12} > \delta_{11} > 0$ such that

$$
\begin{aligned}
&c_{-1} \equiv c_\ell(\delta_{11},\delta_{12}) > 0, \quad c_1 \equiv c_r(\delta_{11},\delta_{12}) > 0, \\
&c_{-2} \equiv b_\ell(\delta_{13}) > 0, \quad \text{and} \quad c_2 \equiv b_r(\delta_{13}) > 0.
\end{aligned}
$$

[C] $x_i \ne \tau_0 \pm \gamma_0\ \forall\ i = 1,\ldots,n$.

[D] $\forall\ \xi_n \downarrow 0, \quad \zeta_n(\xi_n) \longrightarrow 0$.

[E] $\exists\ a > 0$ (finite) such that $\kappa_2 \equiv \nu(a,2)\ <\ \infty$.

[F] $\max_{1\le i\le n}\left\{\frac{1}{\sqrt{n}}|x_i|\right\} \quad\longrightarrow\quad 0$.

A.2 Notation for Section A.3

In addition to the definitions from the text of the article, let $f_i(\boldsymbol{\theta}) = f_{\boldsymbol{\theta}}(x_i)$ and $d_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(x) = f_{\boldsymbol{\theta}_1}(x) - f_{\boldsymbol{\theta}_2}(x)$. Also let the vector-valued function $\boldsymbol{\psi}(\centerdot,\centerdot,\centerdot)$ be

$$\boldsymbol{\psi}(\boldsymbol{\theta}, t, j) = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_{j-1} \\ \theta_{0j} + t(\theta_j - \theta_{0j}) \\ \theta_{0,j+1} \\ \vdots \\ \theta_{05} \end{bmatrix} \qquad \forall\ t \in [0,1], \quad j = 1, \ldots, 5\,, \tag{5}$$

and take $\boldsymbol{\psi}(\boldsymbol{\theta}, t, 0) = \boldsymbol{\theta}_0$. That is, given $t \in [0,1]$, the first $j-1$ elements of $\boldsymbol{\psi}$ are from $\boldsymbol{\theta}$, the last $5-j$ from $\boldsymbol{\theta}_0$, and the j-th element taken to be a value between θ_{0j} and θ_j defined by t. For example, $\boldsymbol{\psi}(\boldsymbol{\theta}, t, 3) = \big[\beta_0,\, \beta_1,\, b_2 + t(\beta_2 - b_2),\, \tau_0,\, \gamma_0\big]$.

Now, let $\boldsymbol{V}^+_{n,j\bullet}$ be the j-th row of the directional Hessian, $\mathbb{V}^+_{n,\sigma^2}$. That is,

$$\boldsymbol{V}^+_{n,j\bullet}(\boldsymbol{\theta}) = \big[V^+_{n,\,j1}(\boldsymbol{\theta}),\ V^+_{n,\,j2}(\boldsymbol{\theta}),\ \ldots, V^+_{n,\,j5}(\boldsymbol{\theta})\big]\ .$$

Using (5), we stack the row vectors, $\boldsymbol{V}^+_{n,j\bullet}\big(\boldsymbol{\psi}(\boldsymbol{\theta}, t, j)\big)$'s, to form the $\mathbb{V}^*_{n,\sigma^2}$ matrix in the Taylor-type expansion of Lemma 2. That is,

$$\mathbb{V}^*_{n,\sigma^2}(\boldsymbol{\theta}, t) = \begin{bmatrix} \boldsymbol{V}^+_{n,1\bullet}\big(\boldsymbol{\psi}(\boldsymbol{\theta}, t, 1)\big) \\ \boldsymbol{V}^+_{n,2\bullet}\big(\boldsymbol{\psi}(\boldsymbol{\theta}, t, 2)\big) \\ \vdots \\ \boldsymbol{V}^+_{n,5\bullet}\big(\boldsymbol{\psi}(\boldsymbol{\theta}, t, 5)\big) \end{bmatrix}. \tag{6}$$

A.3 Theorem Proofs

Proof of Theorem 1

We prove the consistency of $\widehat{\boldsymbol{\theta}}_n$ only. Proving consistency for $\widehat{\sigma}^2_n$ is similar.

Step 1. Fix $\delta > 0$. Let $T^*(\boldsymbol{\theta}, w_{-2}, w_{-1}, w_0, w_1, w_2) = \sum_{i=-2}^{2} \big|d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x_i)\big|^2$ where $\boldsymbol{\theta} \in D_\delta = \big\{\boldsymbol{\theta} \in \Omega : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \geq \delta\big\}$, $w_i \in \mathcal{X}_i$ for $i = 0, \pm 1, \pm 2$. As D_δ is compact, one can show that there exists $r > 0$ so large that for any $\boldsymbol{\theta} \in D_\delta$, T^* is non-decreasing as w_2 increases over $[r, \infty)$ given the other w_i's fixed, or as w_{-2} decreases over $(-\infty, -r]$ given the other w_i's fixed. Hence, $\inf T^*$ over $D_\delta \times \prod_{i=-2}^{2} \mathcal{X}_i$ is no less than $\inf T^*$ over $D_\delta \times \prod_{i=0,\pm 1} \mathcal{X}_i \prod_{i=-2,2} \mathcal{X}_i^r$. Denote the latter infimum by η, which is strictly positive by continuity, compactness, and Lemma 1. Re-label the x_i's so that $x_{i1}, \ldots, x_{i,n_i} \in$

$\mathcal{X}_i$ for $i = 0, \pm 1, \pm 2$ and disregard the rest. By [A] and [B], there exists N such that for all $n > N$, $\min_i n_i \geq nc^*$ where $c^* = \min\{c_0, c_{\pm 1}, c_{\pm 2}\} > 0$. Let $T_n(\boldsymbol{\theta}) = n^{-1} \sum_1^n \big| d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x_i) \big|^2$ and $H_n(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}_0} \big[\overline{S}_n(\boldsymbol{\theta}) \big]$. Note $H_n(\boldsymbol{\theta}) = \sigma^2 + T_n(\boldsymbol{\theta})$. Then, for all $n > N$, $\inf_{D_\delta} H_n(\boldsymbol{\theta}) \geq \sigma^2 + n^{-1} \sum_{j=1}^{nc^*} \inf_{\boldsymbol{\theta} \in D_\delta} T^*(\boldsymbol{\theta}, x_{-2,j}, x_{-1,j}, x_{0j}, x_{1j}, x_{2j}) \geq \sigma^2 + \eta c^* > \sigma^2$. This proves that H_n is not asymptotically flat, and is uniquely minimized at $\boldsymbol{\theta}_0$.

Step 2. Fix $p > 0$ and $\boldsymbol{\theta}' \in \Theta_p$. Write $d_i(\boldsymbol{\theta}) = d_{\boldsymbol{\theta}', \boldsymbol{\theta}}(x_i)$. Note $\boldsymbol{\nabla} d_i(\boldsymbol{\theta}) = -\boldsymbol{\nabla} f_i(\boldsymbol{\theta})$. For all i, a one-term Taylor expansion about $\boldsymbol{\theta}'$ and the Cauchy-Schwarz inequality give $|d_i(\boldsymbol{\theta})| = |d_i(\boldsymbol{\theta}) - d_i(\boldsymbol{\theta}')| \leq |\boldsymbol{\theta} - \boldsymbol{\theta}'| \int_0^1 \big| \boldsymbol{\nabla} f_i(\boldsymbol{\theta}' + t(\boldsymbol{\theta} - \boldsymbol{\theta}')) \big| dt$. Recall a from [E]. One can show that there exists $r \geq a$ so large that for all $\boldsymbol{\theta} \in \Omega$, $|\boldsymbol{\nabla} f_i(\boldsymbol{\theta})| \leq 3\big[r \mathbf{1}\{|x_i| \leq r\} + |x_i| \mathbf{1}\{|x_i| > r\} \big]$. Hence,

$$\sup_{\boldsymbol{\theta} \in \Theta_p \cap \Omega} |d_{\boldsymbol{\theta}', \boldsymbol{\theta}}(x_i)| \leq 3p \big[r \mathbf{1}\{|x_i| \leq r\} + |x_i| \mathbf{1}\{|x_i| > r\} \big] \quad \forall\, p > 0,\ \boldsymbol{\theta}' \in \Theta_p \,. \tag{7}$$

Step 3. Recall κ_2 from [E] and M_j's from Section 2.1. Let $M = \max_{j=0}^4 M_j$. Fix $\boldsymbol{\theta} \in \Omega$. By [E] and Chebyshev's inequality, one can apply (7) to show that for all $\epsilon > 0$, $P_{\boldsymbol{\theta}_0} \big\{ n^{-1} \big| \sum \varepsilon_i d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x_i) \big| > \epsilon \big\} \leq (\sigma/\epsilon)^2 \big\{ n^{-1} \sum_{|x_i| \leq r} |d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x_i)|^2 + \sum_{|x_i| > r} |d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x_i)|^2 \big\} n^{-1} \leq 9(M\sigma/\epsilon)^2 (r^2 + \kappa_2) n^{-1} \to 0$ as $n \to \infty$. Note $\overline{S}_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta}) = n^{-1} \sum \varepsilon_i^2 + 2n^{-1} \sum \varepsilon_i d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x_i) - \sigma^2 \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$.

Step 4. For any small $\delta > 0$, we can define a finite cover of Ω by picking distinct values $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_R \in \Omega$, such that $B_k = \{\boldsymbol{\theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}_k| \leq \delta\}$ is the k-th subcover, $k = 1, \ldots, R$. By (7), the Cauchy-Schwarz inequality, and [E], there exists N such that $n > N$ implies that for all k

$$\begin{aligned}
\sup_{B_k \cap \Omega} |H_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta}_k)| &\leq \frac{1}{n} \sum_i \big(|d_{\boldsymbol{\theta}_k, \boldsymbol{\theta}}(x_i)| + |d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x_i)| \big) |d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x_i)| \\
&< 9\delta(\delta + M)(r^2 + \kappa_2) \,, \\
\sup_{B_k \cap \Omega} \left| \frac{1}{n} \sum_1^n \varepsilon_i d_{\boldsymbol{\theta}_k, \boldsymbol{\theta}}(x_i) \right| &\leq \frac{1}{n} \sum_{|x_i| \leq r} |\varepsilon_i| \big| d_{\boldsymbol{\theta}_k, \boldsymbol{\theta}}(x_i) \big| + \left| \sum_{|x_i| > r} \frac{\varepsilon_i d_{\boldsymbol{\theta}_k, \boldsymbol{\theta}}(x_i)}{n} \right| \\
&\leq \frac{3\delta r}{n} \sum_1^n |\varepsilon_i| + 3\delta \sqrt{\kappa_2} \sqrt{\frac{1}{n} \sum_1^n \varepsilon_i^2} \xrightarrow{\mathrm{P}} 3\delta(r\mu^* + \sigma\sqrt{\kappa_2}) \,,
\end{aligned}$$

where $\mu^* = \mathrm{E}(|\varepsilon_i|)$ for all i. Note $\big|\overline{S}_n(\boldsymbol{\theta}) - \overline{S}_n(\boldsymbol{\theta}_k)\big| \le 2\big|n^{-1}\sum \varepsilon_i d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)\big| + |H_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta}_k)|$. Recall Step 3. Then for $n > N$, $\sup_{B_k\cap\Omega}\big|\overline{S}_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta})\big| \le \sup_{B_k\cap\Omega}\big\{\big|\overline{S}_n(\boldsymbol{\theta}) - S_n(\boldsymbol{\theta}_k)\big| + \big|\overline{S}_n(\boldsymbol{\theta}_k) - H_n(\boldsymbol{\theta}_k)\big| + \big|\overline{H}_n(\boldsymbol{\theta}_k) - H_n(\boldsymbol{\theta})\big|\big\} \le W$ where $W \xrightarrow{\mathrm{P}} 6\delta(r\mu^* + \sigma\sqrt{\kappa_2}) + 18\delta(\delta + M)(r^2 + \kappa_2)$. Since $\sup_\Omega |\overline{S}_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta})| = \max_k \sup_{B_k\cap\Omega}\big|\overline{S}_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta})\big|$ and δ is arbitrary, then $P_{\boldsymbol{\theta}_0}\big\{\sup_\Omega |\overline{S}_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta})| \le \epsilon\big\} \to 1$ for all $\epsilon > 0$. This uniform convergence in probability of $\overline{S}_n$ to H_n and asymptotic curvature of H_n around $\boldsymbol{\theta}_0$ from Step 1 imply consistency of $\widehat{\boldsymbol{\theta}}_n$.

Proof of Theorem 2

Assertion 1. A compactness argument similar to that in Step 1 above, together with [A], [B], and [E], ensures that $n^{-1}\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0) = n^{-1}\sigma^{-2}\sum_i \big[\boldsymbol{\nabla} f_i(\boldsymbol{\theta}_0)\big]\big[\boldsymbol{\nabla} f_i(\boldsymbol{\theta}_0)\big]^T$ has asymptotically finite and strictly positive eigenvalues. The assertion follows from basic laws of linear algebra and the consistency of $\widehat{\boldsymbol{\theta}}_n$.

Assertion 2. By Theorem 1, we may focus on an ever-decreasing neighborhood of $\boldsymbol{\theta}_0$, namely, Θ_{ξ_n}, where $\xi_n \downarrow 0$ is some sequence such that $P_{\boldsymbol{\theta}_0}\big\{\big|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big| \le \xi_n\big\} \to 1$.

Step 1. Fix n. Let $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta_{\xi_n}$. Define $h_{n,\sigma^2}(t) = \ell_{n,\sigma^2}(\boldsymbol{\theta}_1 + t(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1))$ for $t \in [0,1]$. With tedious algebra, one can show $h''_{n,\sigma^2}(t) = (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)^T\big[\mathbb{V}^+_{n,\sigma^2}(\boldsymbol{\theta})\big](\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)$ for all but isolated values of t which correspond to the intersections of the hyper-rays, $R_{i\pm}$'s. Denote by $\mathcal{A}_n$ the event that $n^{-1}\mathbb{V}^+_{n,\sigma^2}(\boldsymbol{\theta})$ is negative definite uniformly over Θ_{ξ_n}. Given $\mathcal{A}_n$, $h''_{n,\sigma^2}(t) < 0$ for all but isolated t's. As $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are arbitrary, the convexity lemma in Chiu et al. (2005, Lemma 4) then asserts that given $\mathcal{A}_n$, ℓ_{n,σ^2} is strictly concave over Θ_{ξ_n}. By Lemma 3 and Assertion 1, $P_{\boldsymbol{\theta}_0}\big\{\mathcal{A}_n\big\} \to 1$ as $n \to \infty$. Hence, $P_{\boldsymbol{\theta}_0}\big\{\widehat{\boldsymbol{\theta}}_n$ is the unique root of $\boldsymbol{U}_{n,\sigma^2}$ over $\Theta_{\xi_n}\big\} \to 1$ as $n \to \infty$.

Step 2. Let $\big[\mathbb{M}_{n,\sigma^2}(\boldsymbol{\theta})\big]^T = \int_0^1 \big[\mathbb{V}^*_{n,\sigma^2}(\boldsymbol{\theta}, t) + \mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)\big]dt$. By Lemma 3 and (6), each component of $\mathbb{M}_{n,\sigma^2}$ is $o_p(n)$ uniformly over Θ_{ξ_n}. By Lemma 2,

$$\mathbf{0} = \boldsymbol{U}_{n,\sigma^2}(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0) + n^{-1}\big[\mathbb{M}_{n,\sigma^2}(\widehat{\boldsymbol{\theta}}_n) - \mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)\big]n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\,. \tag{8}$$

Apply the conclusion of Step 1, the exact normality of $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)$, and Theorem 1, and the subsequent algebra is standard for proving the normality assertion.

Assertion 3. Lemma 4 replaces the exact normality in the proof of Assertion 2.

Assertion 4. This is straight-forward due to a consistent $\widehat{\sigma}_n^2$.

Proof of Theorem 3

Assertion 1. By a one-term Taylor expansion, $\ell_{n,\sigma^2}(\boldsymbol{\theta})=\ell_{n,\sigma^2}(\boldsymbol{\theta}_0)+(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\int_0^1 \boldsymbol{U}_{n,\sigma^2}\big(\boldsymbol{\theta}_0+t(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\big)\,dt$. Let $\mathbb{Z}_{n,\sigma^2}(s,t,\boldsymbol{\theta}) = \mathbb{V}^*_{n,\sigma^2}(\boldsymbol{\theta}_0+t(\boldsymbol{\theta}-\boldsymbol{\theta}_0),s)+\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)$. By Lemma 2, the integral is $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)+\big\{\int_0^1\int_0^1 t\big[\mathbb{Z}_{n,\sigma^2}(s,t,\boldsymbol{\theta})\big]^T ds\,dt-(1/2)\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)\big\}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)$. Note that each component of $\mathbb{Z}_{n,\sigma^2}$ is $o_p(n)$ uniformly over $[0,1]^2\times\Theta_{\xi_n}$ by Lemma 3 and (6). Replace $\boldsymbol{\theta}$ by $\widehat{\boldsymbol{\theta}}_n$. Apply (8) and assemble. Then

$$\begin{aligned} G_{n,\sigma^2} &= 2\big[\ell_{n,\sigma^2}(\widehat{\boldsymbol{\theta}}_n)-\ell_{n,\sigma^2}(\boldsymbol{\theta}_0)\big] \\ &= \big[n^{-1/2}\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)\big]^T\big[n^{-1}\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)\big]^{-1}\big[n^{-1/2}\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)\big]+W_n\,, \end{aligned} \tag{9}$$

where $W_n=o_p(1)$ by the properties of $\mathbb{Z}_{n,\sigma^2}$ and $\mathbb{M}_{n,\sigma^2}$, Theorem 1, and Theorem 2, Assertion 2. Then $G_{n,\sigma^2}\xrightarrow{\text{P}}\chi^2_5$ by Lemma 4 and Theorem 2, Assertion 2.

Assertion 2. This is straight-forward due to a consistent $\widehat{\sigma}_n^2$.

Assertion 3. This proof is mostly standard, except for the use of (9), whose properties hinge on the non-standard Lemmas 2 and 3. In particular, let $\boldsymbol{\theta}^*$ be the parameter value under the null hypothesis, H^*; and let $G'_n=2\big[\ell_n(\widehat{\boldsymbol{\theta}}'_n;\widehat{\sigma}^2_n)-\ell_n(\boldsymbol{\theta}^*;\widehat{\sigma}^2_n)\big]$ and $G^*_n=2\big[\ell_n(\widehat{\boldsymbol{\theta}}^*_n;\widehat{\sigma}^2_n)-\ell_n(\boldsymbol{\theta}^*;\widehat{\sigma}^2_n)\big]$. Differentiate ℓ_{n,σ^2} with respect to the q unknown parameters under H^* to obtain $\boldsymbol{U}^q_{n,\sigma^2}$. Then, by (9) and consistency of $\widehat{\sigma}^2_n$, $G^*_n=\big[n^{-1/2}\boldsymbol{U}^q_{n,\sigma^2}(\boldsymbol{\theta}^*)\big]^T\big[n^{-1}\mathbb{I}^q_{n,\sigma^2}(\boldsymbol{\theta}^*)\big]^{-1}\big[n^{-1/2}\boldsymbol{U}^q_{n,\sigma^2}(\boldsymbol{\theta}^*)\big]+o_p(1)$, where $\mathbb{I}^q_{n,\sigma^2}=\text{Cov}\big[\boldsymbol{U}^q_{n,\sigma^2}\big]$. Similarly, for the p-dimensional estimation problem of H', $G'_n=\big[n^{-1/2}\times\boldsymbol{U}^p_{n,\sigma^2}(\boldsymbol{\theta}')\big]^T\big[n^{-1}\mathbb{I}^p_{n,\sigma^2}(\boldsymbol{\theta}')\big]^{-1}\big[n^{-1/2}\boldsymbol{U}^p_{n,\sigma^2}(\boldsymbol{\theta}')\big]+o_p(1)$. With some standard algebraic manipulation, one can show that under H^*, $D_n=G'_n-G^*_n=\boldsymbol{Z}^T_{n,\sigma^2}\mathbb{Q}_{n,\sigma^2}\boldsymbol{Z}_{n,\sigma^2}+o_p(1)$, where $\boldsymbol{Z}_{n,\sigma^2}=\big[n^{-1}\mathbb{I}^p_{n,\sigma^2}(\boldsymbol{\theta}^*)\big]^{-1/2}\big[n^{-1/2}\boldsymbol{U}^p_{n,\sigma^2}(\boldsymbol{\theta}^*)\big]$, and $\mathbb{Q}_{n,\sigma^2}$ is idempotent with rank $p-q$. The limiting χ^2_{p-q} distribution for D_n follows by applying Theorem 2, Lemma 4, and the Fisher-Cochran Theorem.

REFERENCES

Antonutto, G., and Di Prampero, P. E. (1995), "The Concept of Lactate Threshold. A Short Review," *Journal of Sports, Medicine and Physical Fitness*, 35, 6–12.

Barrowman, N. J., and Myers, R. A. (2000), "Still More Spawner-Recruitment Curves: the Hockey Stick and its Generalizations," *Canadian Journal of Fisheries and Aquatic Sciences*, 57, 665–676.

Beaver, W. L., Wasserman, K., and Whipp, B. J. (1985), "Improved Detection of Lactate Threshold During Exercise Using a Log-Log Transformation," *Journal of Applied Physiology*, 59, 1936–1940.

Brown, C. C. (1987), "Approaches to Intraspecies Dose Extrapolation," in *Toxic Substances and Human Risk: Principles of Data Interpretation*, eds. R. G. Tardiff and J. V. Rodrick, New York: Plenum, pp. 237–268.

Cook, R. D., and Weisberg, S. (1990), "Confidence curves in nonlinear regression," *Journal of the American Statistical Association*, 85, 544–551.

Chiu, G. S. (2002), "Bent-Cable Regression for Assessing Abruptness of Change," unpublished Ph.D. dissertation, Simon Fraser University, Department of Statistics and Actuarial Science.

Chiu, G., Lockhart, R., and Routledge, R. (2002), "Bent-Cable Asymptotics when the Bend is Missing," *Statistics and Probability Letters*, 59, 9–16.

—— (2005), "Asymptotic Theory for Bent-Cable Regression — the Basic Case," *Journal of Statistical Planning and Inference*, 127, 143–156.

Feder, P. I. (1975), "On asymptotic distribution theory in segmented regression problems — identified case," *Annals of Statistics*, 3, 49–83.

Gallant, A. R. (1974), "The Theory of nonlinear regression as it relates to segmented polynomial regressions with estimated join points," Mimeograph Series No. 925, North Carolina State University, Institute of Statistics.

—— (1975), "Inference for nonlinear models," Mimeograph Series No. 875, North Carolina State University, Institute of Statistics.

Hanley, J. A. (2004), " 'Transmuting' women into men: Galton's family data on human stature," *The American Statistician*, 58, 237–243.

Hinkley, D. V. (1969), "Inference about the intersection in two-phase regression," *Biometrika*, 56, 495–504.

—— (1971), "Inference in two-phase regression," *Journal of the American Statistical Association*, 66, 736–743.

Hušková, M. (1998), "Estimators in the location model with gradual changes," *Commentationes Mathematicae Universitatis Carolinae*, 39, 147–157.

Hušková, M. and Steinebach, J. (2000), "Limit theorems for a class of tests of gradual changes," *Journal of Statistical Planning and Inference*, 89, 57–77.

Jarušková, D. (1998a), "Testing appearance of linear trend," *Journal of Statistical Planning and Inference*, 70, 263–276.

—— (1998b), "Change-point estimator in gradually changing sequences," *Commentationes Mathematicae Universitatis Carolinae*, 39, 551–561.

—— (2001), "Change-point estimator in continuous quadratic regression," *Commentationes Mathematicae Universitatis Carolinae*, 42, 741–752.

Jones, M. C., and Handcock, M. S. (1991), "Determination of Anaerobic Threshold: What Anaerobic Threshold?" *Canadian Journal of Statistics*, 19, 236–239.

Kline, K. A. (1997), "Metabolic Effects of Incremental Exercise on Arabian Horses Fed Diets containing Corn Oil and Soy Lecithin," unpublished M.S. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Department of Animal and Poultry Sciences.

Le Cam, L. (1970), "On the assumptions used to prove asymptotic normality of maximum likelihood estimators," *Annals of Mathematical Statistics*, 41, 802–828.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman and Hall.

Moquin, A., and Mazzeo, R. S. (2000), "Effect of Mild Dehydration on the Lactate Threshold in Women," *Medicine and Science in Sports and Exercise*, 32, 396–402.

Naylor, R. E. L., and Su, J. (1998), "Plant Development of Triticale Cv. Lasko at Different Sowing Dates," *Journal of Agricultural Science*, 130, 297–306.

Neuman, M. J., Witting, D. A., and Able, K. W. (2001), "Relationships Between Otolith Microstructure, Otolith Growth, Somatic Growth and Ontogenetic Transitions in Two Cohorts of Windowpane," *Journal of Fish Biology*, 58, 967–984.

Pollard, D. (1997), "Another Look at Differentiability in Quadratic Mean," in *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, eds. D. Pollard, E. N. Torgersen, and G. L. Yang, New York: Springer, pp. 305–314.

Routledge, R. D. (1991), "Using Time Lags in Estimating Anaerobic Threshold," *Canadian Journal of Statistics*, 19, 233–236.

Rukhin, A. L. and Vajda, I. (1997), "Change-point estimation as a nonlinear regression problem," *Statistics*, 30, 181–200.

Schneider, D. A., McLellan, T. M., and Gass, G. C. (2000), "Plasma Catecholamine and Blood Lactate Responses to Incremental Arm and Leg Exercise," *Medicine and Science in Sports and Exercise*, 32, 608–613.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley and Sons.

Tishler, A., and Zang, I. (1981), "A New Maximum Likelihood Algorithm for Piecewise Regression," *Journal of the American Statistical Association*, 76, 980–987.

Vachon, J. A., Bassett Jr., D. R., and Clarke, S. (1999), "Validity of the Heart Rate Deflection Point as a Predictor of Lactate Threshold During Running," *Journal of Applied Physiology*, 87, 452–459.

Wachsmuth, A., Wilkinson, L., and Dallal, G. E. (2003), "Galton's bend: a previously

undiscovered nonlinearity in Galton's family stature regression data," *The American Statistician*, 57, 190–192.

Weltman, A. (1995), *The Blood Lactate Response to Exercise*, Champaign, Illinois: Human Kinetics.

Weltman, A., Wood, C. M., Womack, C. J., Davis, S. E., Blumer, J. L., Alvarez, J., Sauer, K., and Gaesser, G. A. (1994), "Catecholamine and Blood Lactate Responses to Incremental Rowing and Running Exercise," *Journal of Applied Physiology*, 76, 1144–1149.

Wigglesworth, V. B. (1972), *The Principles of Insect Physiology* (7th ed.), London: Chapman and Hall.

Table 1. Empirical Coverage of Nominal 95% CRs from Group 1 Simulations.

CR type	(τ_0, γ_0) normal	(τ_0, γ_0) uniform	(τ_0, γ_0) t_5	γ=0 normal / uniform	γ=0 t_5
(A): χ^2	92.02%	91.78%	91.70%	0%	0.02%
(A): F	93.94%	93.34%	93.62%	0%	0.04%
(B): F	93.42%	93.00%	93.18%	0%	0 %

NOTE: Coverage is based on 5,000 simulated bent-cable datasets with response errors from a normal, uniform, or t_5 distribution, scaled to have SD=0.015. Type (A) CRs are based on the deviance (likelihood ratio) statistic using $\chi^2_2(0.05)$- and $2F_{2,n-2}(0.05)$-cutoffs. Type (B) CRs are based on the Wald statistic using a $2F_{2,n-2}(0.05)$-cutoff. Entries under "(τ_0, γ_0)" are rates for a CR covering the true transition parameters, and those under "γ=0" are rates covering a broken stick.

Table 2. No. of Simulated Deviance Surfaces (out of 100) Truncated at -5.99 Exhibiting Various Characteristics.

Set	Shape *ridges/ plateaus*	Shape *half-dome*	Shape *paraboloidal*	Best Fit *qualit. correct*	Best Fit *qualit. incorrect*	Broken-Stick *in*	Broken-Stick *out*
(i a)	83	17	0	76	24	90	10
(ii a)	65	35	0	47	53	97	3
(ii b1)	0	99	1	67	33	99	1
(ii b2)	0	100	0	52	48	100	0

NOTE: Data were generated from (i) a smooth bent cable (γ_0>0) and (ii) a broken stick (γ_0=0) with chance scatter that is typical in (a) biological studies and (b) experiments in the physical sciences. One x-value coincided with the underlying kink for (ii b1), and none for (ii b2). Set (i b) appears in Table 1.

Figure Titles and Legends

Figure 1: Sockeye Data. When does the decline begin for Rivers Inlet sockeye salmon (*Oncorhynchus nerka*), and how abrupt is the onset? The bent-cable model can be used to address these questions.

Figure 2: Sockeye Deviance Surface. The profile log-likelihood deviance surface vs. τ and γ (center and half-width of bend, respectively) for the data in Figure 1. All values of τ and γ with deviance values in the upper plateau are consistent with the data. For instance, these include a cable bend which ranges over the entire dataset, and a broken stick with a corner at 1993.

Figure 3: Galton's Data. What is the nature of "Galton's bend"? Overlaid on Galton's famous family stature data (reproduced by Hanley 2004) are two broken-stick fits (solid and dotted lines, respectively) and a cable fit with a bend of half-width 1.13 inches (dashed lines). Here, kinked and smooth bends seem equally plausible.

Figure 4: Deviance Surfaces for Galton's Data. In his analysis, Galton assumed a parent-to-child linear relationship. Here, we provide profile deviance surfaces for bent-cable regression of mid-parent height on child's height (panel **(a)**) and vice versa (panel **(b)**). For **(a)**, the two peaks close to $(\tau, \gamma)=(71, 0)$ and the rounded peak nearby correspond to the fits shown in Figure 3, which (along with many others, including purely quadratic fits corresponding to the upper plateau of the surface) are equally consistent with the data. Similar conclusions can be made to the surface in **(b)** for the reverse regression. In addition, single-phase linear fits corresponding to the lower flat regions of this surface are also consistent with the data here.

Figure 5: Anaerobic Data. (a) Carbon dioxide (CO_2) output vs. oxygen (O_2) uptake in mL/s for an athlete on a treadmill with a continually increasing incline. **(b)** The same dataset with the best linear least-squares fit removed. These data are consistent with either an abrupt threshold or a wide smooth bend.

Figure 6: Band Height Data. A plot of log(stagnant-band-height) vs. log(water flow rate), as cited by Seber and Wild (1989). Does the graph exhibit an abrupt break or a smooth transition?

Figure 7: (a) Deviance Surface. The log-likelihood deviance surface vs. τ and γ for the data in Figure 6. All deviance values above -5.99 (χ^2-based) are consistent with the data at an approximate 95% confidence level. All these values lead to cables with a long smooth bend. The same is also true when the nominal 95% cutoff is taken to be -6.71 (F-based). **(b) Best-Fitting Bent Cable.** The true bend is estimated to range over log-flow rates of -0.373 to 0.484 (dotted lines).

Figure 1

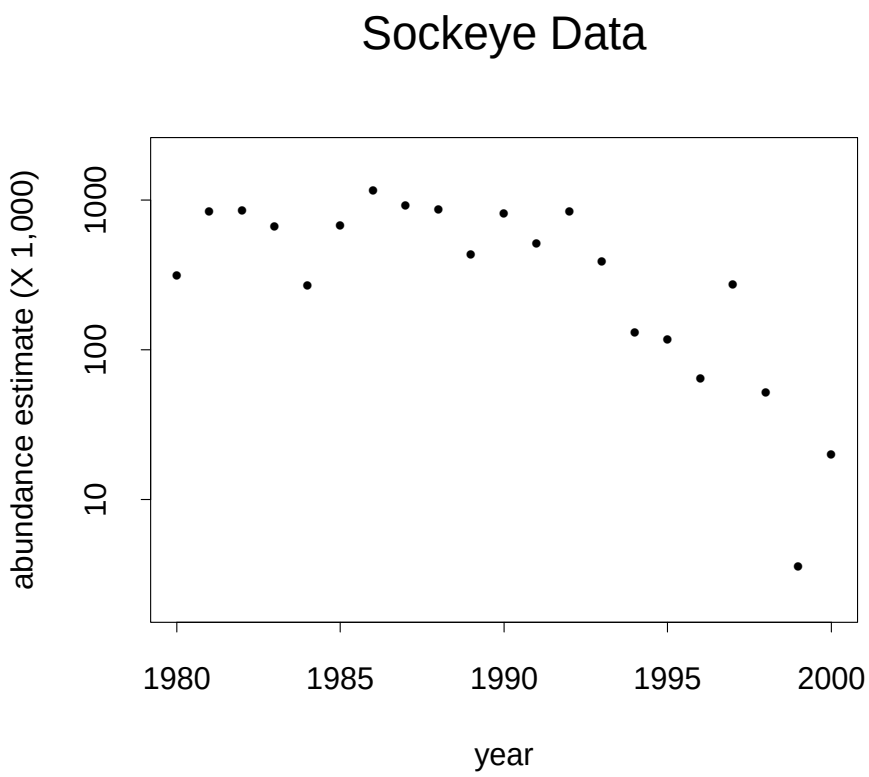

Sockeye Data
abundance estimate (X 1,000)
1000
100
10
1980
1985
1990
1995
2000
year

Figure 2

Sockeye Deviance Surface

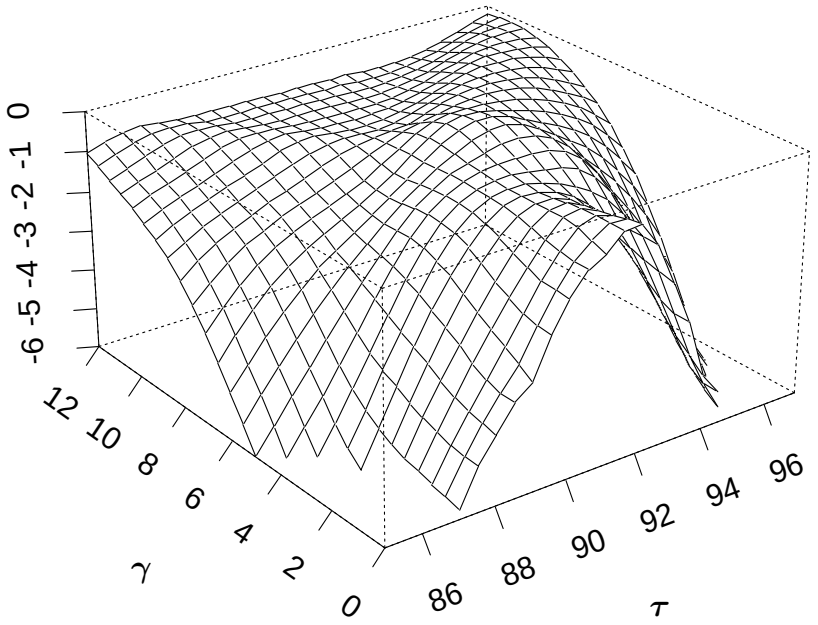


Figure 3

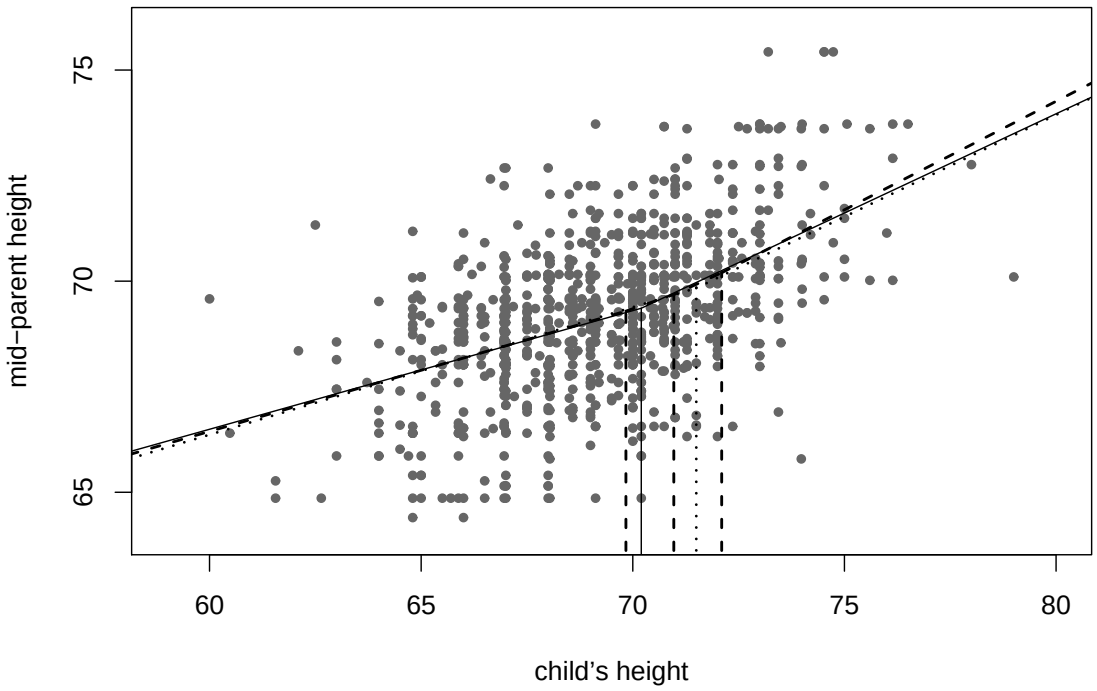
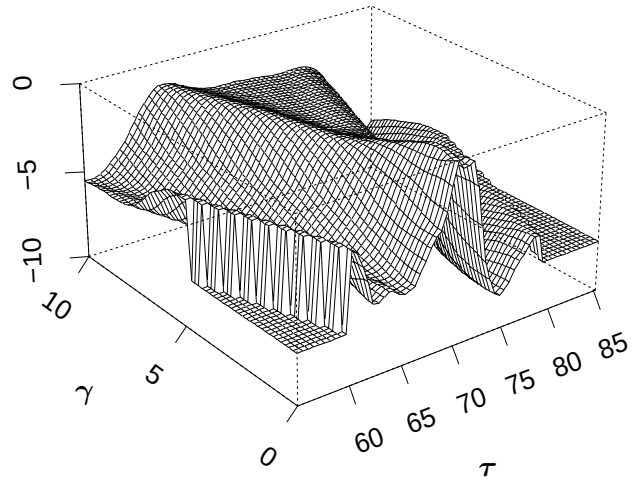

Galton's Data
mid−parent height
child's height
60
65
70
75
80
65
70
75

Figure 4

(a) Deviance Surface for Y = Mid−parent Height



(b) Deviance Surface for X = Mid−parent Height

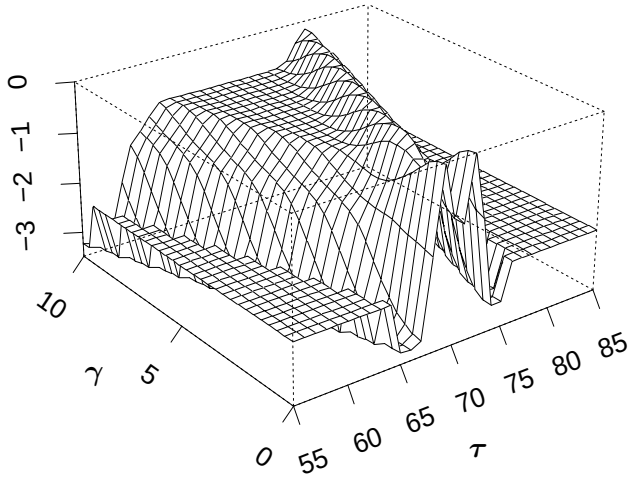
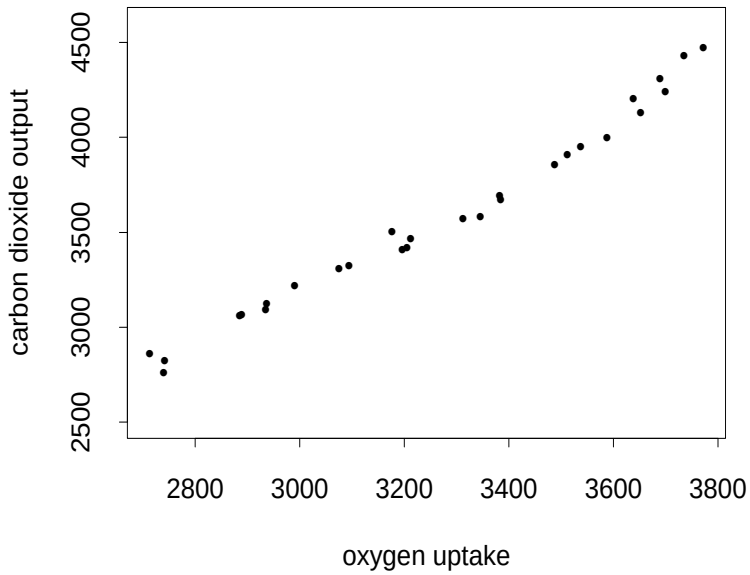


Figure 5

(a) Anaerobic Data



(b) (detrended)

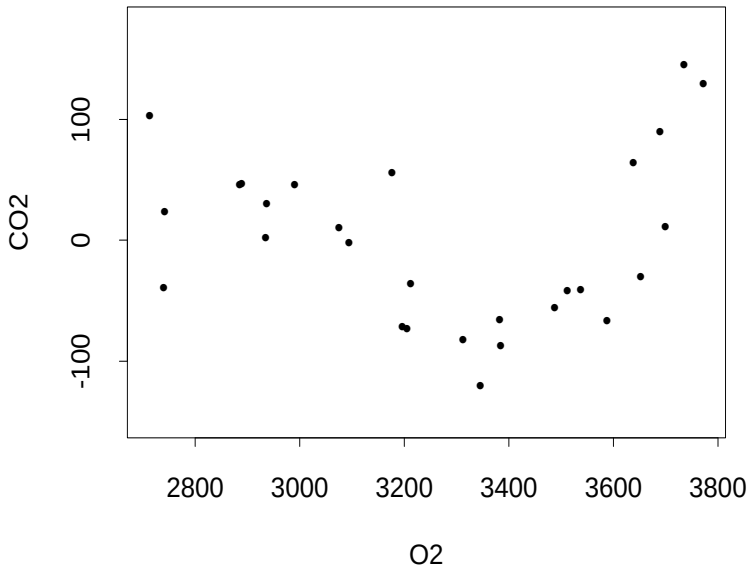


Figure 6

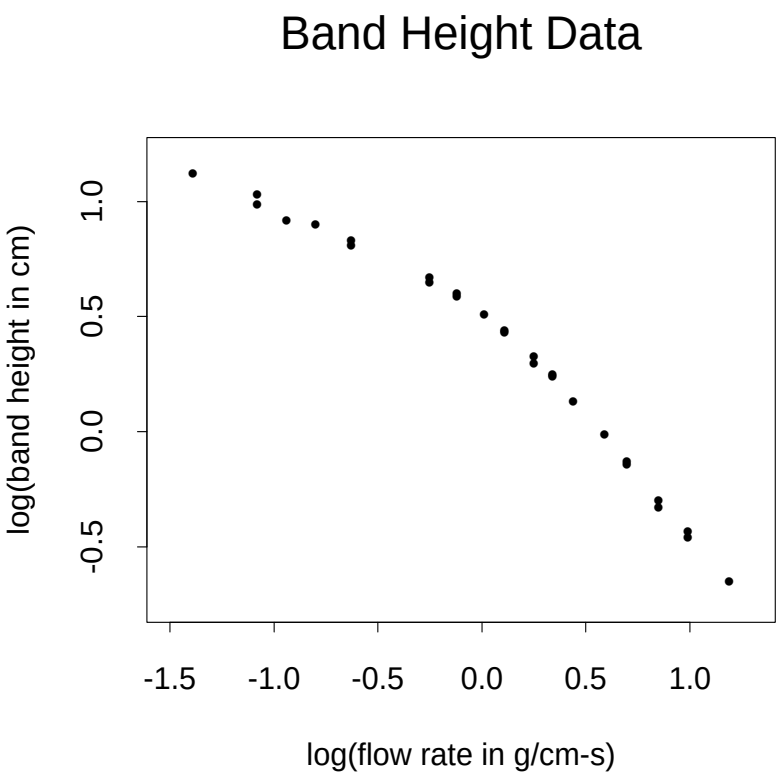
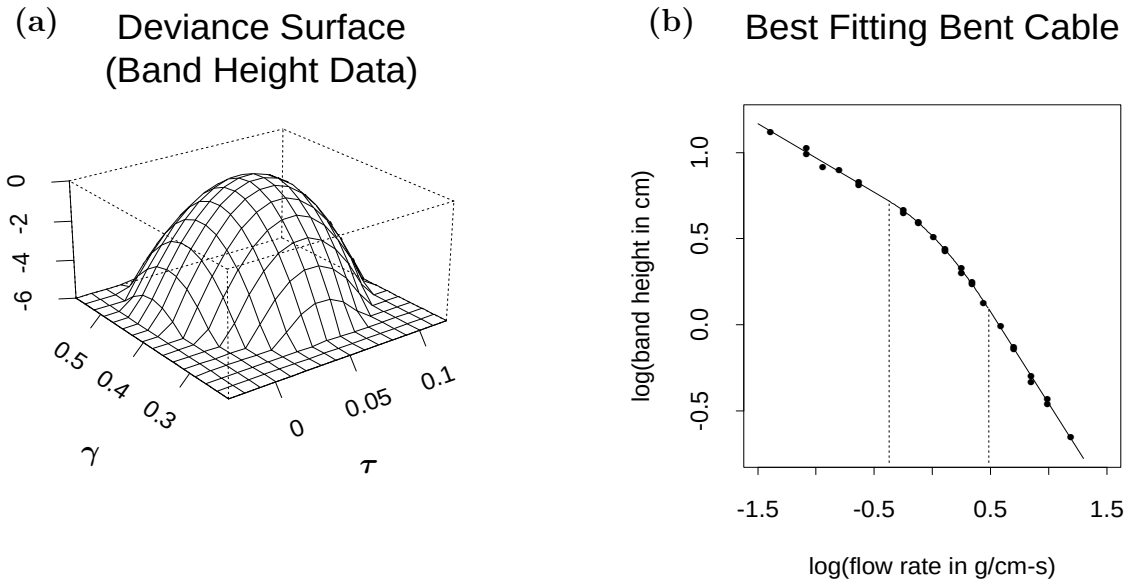

Band Height Data
log(band height in cm)
1.0
0.5
0.0
-0.5
-1.5
-1.0
-0.5
0.0
0.5
1.0
log(flow rate in g/cm-s)

Figure 7


(a)
Deviance Surface
(Band Height Data)
0
-2
-4
-6
0.5
0.4
0.3
γ
0
0.05
0.1
τ
(b)
Best Fitting Bent Cable
log(band height in cm)
1.0
0.5
0.0
-0.5
-1.5
-0.5
0.5
1.5
log(flow rate in g/cm-s)