Bent-Cable Regression for Assessing Abruptness of Change

by

Grace S. Chiu

M. Sc. University of British Columbia, 1996

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the Department
of
Statistics and Actuarial Science

SIMON FRASER UNIVERSITY
June 2002

APPROVAL

Name: Grace S. Chiu

Degree: Doctor of Philosophy

Title of thesis: Bent-Cable Regression for Assessing Abruptness of Change

Examining Committee: Dr. Larry Weldon
Chair

____

Dr. Richard Routledge
Senior Supervisor

____

Dr. Richard Lockhart
Co-Supervisor

____

Dr. Tim Swartz

____

Dr. Michael Stephens

Dr. Nancy Heckman
External Examiner
Department of Statistics
University of British Columbia

Date Approved: ______________________________

Abstract

We consider fitting a *bent-cable model* to data that arise from natural phenomena viewed as containing a threshold or change point. The bent cable comprises two straight lines joined smoothly by a quadratic bend in the middle. As the domain of the bend shrinks, the model approaches the widely used "broken stick" with an abrupt threshold. While fitting a broken stick requires an *a priori* abruptness assumption, fitting a bent cable allows the observed data to reveal whether a sharp or smooth change is more convincing. Worked examples demonstrate the use of bent-cable regression for assessing the abruptness of change.

The majority of this thesis is devoted to developing asymptotic theory for least-squares estimation of the bent-cable parameters. Practical conditions on the placement of covariate values are given to ensure regularity of the estimation problem, despite non-differentiability of the model's first partial derivatives (with respect to the covariate and model parameters). Under these conditions, we have (1) consistency of the estimators, and (2) the usual normal and chi-square limiting distributions for the estimators and deviance (log-likelihood ratio) statistic, respectively, in the case of a non-trivial bend region. We also show that, when the underlying relationship does not contain the middle segment, the estimation problem is irregular and gives exceptionally peculiar asymptotics.

Acknowledgments

I am eternally indebted to my thesis co-supervisors, Professors Richard Routledge and Richard Lockhart, for providing continual guidance and support throughout this long and often rugged journey. They have helped to create the most pleasurable work environment, and made all that head-banging and frustration worthwhile and sweetly fruitful at the end. I will forever cherish those countless hours spent in their offices, where tears were shed and laughs were uttered, all amidst horrendous mathematical formulae.

I would also like to thank Prof. Nancy Heckman for her moral support and friendship since my undergraduate years at UBC.

Many thanks to the SFU Department of Statistics and Actuarial Science (and the late Department of Mathematics and Statistics) for facilitating the friendliest and most family-like atmosphere for its graduate students. The down-to-earth faculty and staff were undeniably responsible for the most enjoyable four years of my education.

Many hugs and kisses to my fellow graduate students who have been through good and hard times with me, particularly those who participated in my thirtieth-birthday party. The pleasant surprise I got there will indeed be hard to beat.

My heartfelt thanks go to *Elektra Women's Choir* for having helped to preserve my sanity while I pursued this degree.

Finally, my love and thanks go to my dear parents, who have always made me their pride.

Contents

Abstract iii

Acknowledgments iv

List of Figures viii

Notation xi

1 Introduction 1

2 Examples of Bent-Cable Regression 5

2.1 Sharply Declining Salmon... Or Not? 5

2.2 An Anaerobic Threshold? 8

2.3 Convincing Evidence of a Smooth Transition 10

2.4 Implications 11

2.5 Statistical Details 12

2.5.1 Computational Methods 13

2.5.2 Analyses 23

3 Precedent Models and Theory 26

4 The Basic Bent Cable 29

4.1 Basic Bent-Cable Regression 30

4.2 Results 30

4.3 The Maximum Likelihood Estimator 31

4.4 Conditions for the Basic Bent Cable 32

4.5 Formal Statements of Results 33

4.6 Auxiliary Lemmas for Proving Theorem 4.2 37

4.7 Theorem Proofs 39

4.7.1 Proof of Theorem 4.1 39

4.7.2 Proof of Theorem 4.2 44
4.8 A Kinked Truth 49
4.8.1 The One-Parameter Basic Bent Cable 49
4.8.2 Results for $\gamma_0 = 0$ 50
4.8.3 Notation for Estimation of the Kinked Truth 51
4.8.4 The Design 52
4.8.5 Formal Results for Kinked Truth Estimation 53
4.8.6 Proof of Theorem 4.3 54
4.8.7 Final Remarks on Estimating the Kinked Truth . . . 55
4.9 Chapter Summary 56
4.10 Chapter Appendix 57
4.10.1 Lemma Proofs 57
4.10.2 Weak Convergence of the Translated Sum-of-Squares 69
5 The Full Bent Cable 71
5.1 Results 72
5.2 Parameter Space and Design Conditions 72
5.2.1 Design Conditions for Full Bent-Cable Regression . . 74
5.3 Notation 77
5.4 Formal Statements of Results 78
5.5 Theorem Proofs 82
5.5.1 Proof of Theorem 5.1 82
5.5.2 Important Quantities in Theorems 5.2 and 5.3 86
5.5.3 Auxiliary Lemmas for Proving Theorem 5.2 89
5.5.4 Proof of Theorem 5.2 90
5.5.5 Proof of Theorem 5.3 94
5.6 Chapter Appendix 99
5.6.1 More on the Choice of Ω 99
5.6.2 Some Bounds on $d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x)$ 105
5.6.3 Some Notation for the Taylor-type Expansion 107
5.6.4 Lemma Proofs 108
6 Summary of Main Results 119

7 Future Work 122
7.1 Testing $\gamma = 0$ 122
7.2 What *Is* the Transition Region? 124
7.2.1 Dispersion of the Second Derivative Function 126
Bibliography 129

List of Figures

1.1 A bent-cable model (solid line) and a broken-stick model (dotted line). Beyond the edges of the middle quadratic segment of the bent cable, both models have the same incoming and outgoing linear phases. Thus, the broken stick is a bent cable whose smooth bend (transition region) is shrunk to a sharp break (change point). We also refer to the two-phase model as a bent cable with zero or no bend (i.e. one with a missing middle segment). 2

2.1 Estimates of spawning abundance (plotted on a logarithmic scale) vs. year for Rivers Inlet sockeye salmon (*Oncorhynchus nerka*). When does the decline begin, and how abrupt is the onset? The bent-cable model can be used to address these questions. 6

2.2 The profile log-likelihood (see [11] and Section 2.5.1) deviance surface vs. τ and γ (center and half-width of bend, respectively) for the data. All values of τ and γ with deviance values in the upper plateau are clearly consistent with the data. These include a cable bend which ranges over the entire dataset, and a broken stick with a corner at 1993. 7

2.3 **(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 detrended by removing the best linear least-squares fit. These data are consistent with either an abrupt threshold or a single quadratic curve. (See Section 2.5.2 for statistical details). 9

2.4 A plot of $\log_e$(stagnant-band-height) vs. $\log_e$(water flow rate), taken from [10]. Does the graph exhibit an abrupt break or a smooth transition? . . . 11

2.5 **(a)** The log-likelihood deviance surface vs. τ and γ. All deviance values above the horizontal plane at -5.99 are consistent with the data at an approximate 95% confidence level. (Refer to Section 2.5.2 for details.) All these values lead to cables with a long smooth bend. **(b)** The data are presented with the best-fitting bent cable. The bend ranges over log-flow rates of -0.373 to 0.484 (dotted lines). . . . 12

2.6 Profile deviance contours for **(a)** the sockeye data (Figures 2.1 and 2.2), and **(b)** the band height data (Figure 2.5). Both plots are constructed by the "gridding" technique described above. That is, they are projections of the "approximate" profile deviance surface, $-\tilde{D}_{\text{prof}}$, onto the $k \times k$ grid of preset (τ_i, γ_j)'s. The S-Plus function '`contour()`' automatically places '0' on the plot, marking $\tilde{\boldsymbol{\alpha}}$ — the grid point which gives the maximum value among the k^2 values of $-\tilde{D}_{\text{prof}}(\boldsymbol{\alpha}_{ij})$. . . . 16

2.7 The profile log-likelihood deviance surface vs. τ and γ for the anaerobic data of Figure 2.3. . . . 24

4.1 An example of the rays $R_{i\pm}$'s (on the (τ, γ)-plane) along which the score surfaces $U_{n\tau}$ and $U_{n\gamma}$ have folds. Each pair of $R_{i\pm}$ is indexed by a data point x_i. The "+" rays have slope 1 (i.e. 45 degrees), while the "$-$" rays have slope -1 (i.e. 135 degrees). . . . 38

5.1 The $\mathcal{X}_i$ regions are required to contain non-trivial fractions of data to ensure regular asymptotics for the bent-cable estimation problem. We refer to this design as the "2-1-2"-configuration for the underlying bent cable. (Refer to Section 5.3 for the definitions of the $\mathcal{X}_i^a$'s.) . . . 75

7.1 For the transition region of both quadratic and quartic models to stretch over the fixed interval $[-\gamma, \gamma]$, the sharper quartic bend must sweep underneath the more mildly curved quadratic bend. . . . 125

7.2 The first and second derivative functions of the (centered) quadratic basic bent cable, q, and the (centered) quartic basic bent cable, u. The term *basic* refers to the model constraints of (i) to (iv) on page 125. . 127

Notation

n	:	sample size
$x_i, Y_i \in \mathbb{R}$	:	i-th design point and (random) response, respectively, $i = 1, \ldots, n$
$\varepsilon_i \in \mathbb{R}$	:	random error associated with i-th response;
		error has mean 0 and constant variance σ^2
$\boldsymbol{x} = (x_1, \ldots, x_n)$	:	boldface type denotes vector
$\Omega \subset \mathbb{R}^p$	:	space of bent-cable regression parameters;
		$p = 2$ in Chapter 4, $p = 5$ in Chapter 5
$\boldsymbol{\theta} \in \Omega$	:	vector of regression parameters
$\boldsymbol{\theta}_0 \in \Omega$	:	subscript '0' denotes true parameter value
$\theta_j, \theta_{0j} \in \mathbb{R}$	:	j-th components of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$, respectively, $j = 1, \ldots, p$
$\beta_0, \beta_1, \beta_2$	:	linear bent-cable parameters
τ	:	center of cable bend, non-linear parameter
γ	:	half-width of cable bend, non-linear parameter
b_0, b_1, b_2	:	true values of linear parameters
τ_0, γ_0	:	true values of non-linear parameters
$\Theta_\delta \subset \Omega$	:	δ-neighborhood of $\boldsymbol{\theta}_0$, $\delta > 0$, i.e. $\Theta_\delta = \big\{\boldsymbol{\theta} \in \Omega : \lvert\boldsymbol{\theta} - \boldsymbol{\theta}_0\rvert \le \delta\big\}$
$q_{\boldsymbol{\theta}}(x) \equiv q(x; \boldsymbol{\theta})$	:	basic bent-cable model (Chapter 4), $\boldsymbol{\theta} = (\tau, \gamma)$;
		$q(x; \tau, \gamma) = \dfrac{(x - \tau + \gamma)^2}{4\gamma}\, \mathbf{1}\big\{\lvert x - \tau\rvert \le \gamma\big\} + (x - \tau)\, \mathbf{1}\big\{x > \tau + \gamma\big\}$
$f_{\boldsymbol{\theta}}(x) \equiv f(x; \boldsymbol{\theta})$	:	full bent-cable model (Chapter 5), $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \tau, \gamma)$;
		$f(x; \beta_0, \beta_1, \beta_2, \tau, \gamma) = \beta_0 + \beta_1\, x + \beta_2\, q(x; \tau, \gamma)$
$q_{\boldsymbol{\theta_0}}\,,\ f_{\boldsymbol{\theta}_0}$	:	underlying bent cables;
		$Y_i = q_{\boldsymbol{\theta_0}}(x_i) + \varepsilon_i$ (Chapter 4); $\quad Y_i = f_{\boldsymbol{\theta}_0}(x_i) + \varepsilon_i$ (Chapter 5)
$q_{\boldsymbol{\theta}}\,,\ f_{\boldsymbol{\theta}}$	:	"candidate" bent cables being considered for the estimation of $\boldsymbol{\theta}_0$
$d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x)$	:	distance between true and candidate cables, both evaluated at x
$S_n(\boldsymbol{\theta})$	:	error sum-of-squares function given dataset, evaluated at $\boldsymbol{\theta}$
$\overline{S}_n(\boldsymbol{\theta})$	:	error mean-square function, i.e. $\overline{S}_n(\boldsymbol{\theta}) = n^{-1} S_n(\boldsymbol{\theta})$
$H_n(\boldsymbol{\theta})$	:	expected error mean-square function, i.e. $H_n(\boldsymbol{\theta}) = E_{\boldsymbol{\theta_0}}\big[\overline{S}_n(\boldsymbol{\theta})\big]$
$T_n(\boldsymbol{\theta})$	:	"gap mean-square" function, i.e. $T_n(\boldsymbol{\theta}) = n^{-1} \sum_1^n \big\lvert d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x_i)\big\rvert^2$
$\boldsymbol{\nabla}$	:	gradient operator on functions of vector-valued arguments
∇_j	:	partial differentiation operator with respect to argument's j-th component

$\ell_n(\boldsymbol{\theta})$	:	likelihood function of $\boldsymbol{\theta}$ given dataset $\big\{(x_i, Y_i)\big\}_{i=1}^n$
$\boldsymbol{U}_n(\boldsymbol{\theta})$	:	score function (vector-valued), i.e. $\boldsymbol{U}_n(\boldsymbol{\theta}) = \boldsymbol{\nabla}\ell_n(\boldsymbol{\theta})$
$U_{nj}(\boldsymbol{\theta})$	:	j-th component of $\boldsymbol{U}_n(\boldsymbol{\theta})$, i.e. $U_{nj}(\boldsymbol{\theta}) = \nabla_j\, \ell_n(\boldsymbol{\theta}) = \dfrac{\partial\, \ell_n}{\partial\, \theta_j}$
$U_{j,i}(\boldsymbol{\theta})$	:	i-th summand of $U_{nj}(\boldsymbol{\theta})$, i.e. $U_{nj}(\boldsymbol{\theta}) = \sum_{i=1}^n U_{j,i}(\boldsymbol{\theta})$
$\mathbb{I}_n(\boldsymbol{\theta})$	:	Fisher Information matrix, as a function of $\boldsymbol{\theta}$; $\mathbb{I}_n(\boldsymbol{\theta}) = \mathrm{Cov}_{\boldsymbol{\theta_0}}\big[\boldsymbol{U}_n(\boldsymbol{\theta})\big]$
$I_{n,jk}(\boldsymbol{\theta})$	:	(j,k)-th component of $\mathbb{I}_n(\boldsymbol{\theta})$
$\mathbb{V}_n(\boldsymbol{\theta})$	:	Hessian matrix of $\ell_n(\boldsymbol{\theta})$, i.e. $\mathbb{V}_n(\boldsymbol{\theta}) = \boldsymbol{\nabla}\boldsymbol{U}_n(\boldsymbol{\theta}) = \boldsymbol{\nabla}^2\, \ell_n(\boldsymbol{\theta})$, wherever defined
$V^+_{jk,i}(\boldsymbol{\theta})$	:	directional partial derivative of $U_{k,i}(\boldsymbol{\theta})$ with respect to θ_j, i.e. $V^+_{jk,i}(\boldsymbol{\theta}) = \lim_{h\downarrow 0} \dfrac{\partial}{\partial\, \theta_j} U_{k,i}(\theta_1, \ldots, \theta_{j-1}, \theta_j + h, \theta_{j+1}, \ldots, \theta_p)$
$\mathbb{V}^+_n(\boldsymbol{\theta})$	:	"directional" Hessian; (j,k)-th component is $V^+_{n,jk}(\boldsymbol{\theta}) = \sum_{i=1}^n V^+_{jk,i}(\boldsymbol{\theta})$
$\widehat{\boldsymbol{\theta}}_n \in \Omega$, MLE	:	global maximizer of ℓ_n, i.e. maximum likelihood estimator of $\boldsymbol{\theta}_0$
$\widehat{\boldsymbol{\theta}}_n \in \Omega$, LSE	:	global minimizer of S_n (and $\overline{S}_n$), i.e. least-squares estimator of $\boldsymbol{\theta}_0$
$\widehat{\sigma^2_n} \in [0, \infty)$	:	minimized error mean-square, i.e. $\widehat{\sigma^2_n} = \overline{S}_n\big(\widehat{\boldsymbol{\theta}}_n\big)$
$\xrightarrow{\mathrm{P}}$	:	convergence in probability, as $n \longrightarrow \infty$
$\xrightarrow{\mathcal{L}}$	:	convergence in distributional law, as $n \longrightarrow \infty$

Chapter 1

Introduction

The so-called broken stick is a popular piecewise-linear model used in biological studies for estimating the onset of change ([1]–[3]). This sharply kinked line is particularly appealing in its structural simplicity. For example, consider the relationship of abundance versus time for a declining fish population. A fisheries manager will naturally try to use the date of onset as a clue to the actual cause of the decline. In this and other fields, such as physiology, researchers are often tempted to conclude an abrupt onset from a broken-stick fit, although there is seldom solid theory to justify the abruptness ([4]–[7]).

To assess the nature of change, a non-parametric approach has been suggested in [8], mostly as an exploratory tool when abruptness has not yet been verified. In this thesis, we take a more formal parametric approach, by examining the use of a flexible model developed by Tishler and Zang ([9]). We now call it the "bent cable" due to the smooth bend as opposed to a sharp break in a snapped stick. The bent cable can be regarded as an extension to the widely used broken stick with its simple structure retained. We exploit the fact that an extremely sharp bend reduces the bent cable to a broken stick (Figure 1.1). Thus, besides relaxing the sometimes unreasonable *a priori* assumption of abruptness, the bent cable also allows the data to reveal to us whether an abrupt change point or a smooth transition region is more convincing.

The bent-cable model was originally developed to handle the non-differentiability of the first derivative when fitting a broken stick ([9]). A “phoney” bend of fixed, non-trivial width replaces the kink. This was a computational tactic, as continuous differentiability provided more numerical stability in the estimation procedure. Upon numerical convergence, the phoney bend would be ignored. However, if a gradual transition is considered possible for the underlying regression function, then the bend width can be regarded as part of the parametric model.

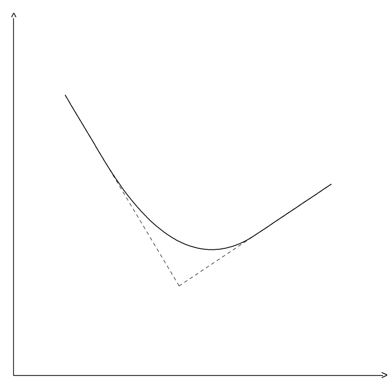

Figure 1.1: A bent-cable model (solid line) and a broken-stick model (dotted line). Beyond the edges of the middle quadratic segment of the bent cable, both models have the same incoming and outgoing linear phases. Thus, the broken stick is a bent cable whose smooth bend (transition region) is shrunk to a sharp break (change point). We also refer to the two-phase model as a bent cable with zero or no bend (i.e. one with a missing middle segment).

Three worked examples are shown in Chapter 2 to demonstrate the use of bent-cable regression for assessing the abruptness of change. Confidence regions for the bend parameters can be used to provide insight into the nature of the change in the underlying relationship. An easy computational algorithm is suggested for such estimation procedures. From these examples, we will see that data from typical biological phenomena are far too imprecise to support claims of abrupt thresholds. However, formal assessments of abruptness via confidence regions rely on regularity of the least-squares estimation problem. This is not the case for two of the three examples. In fact, we will see in Chapter 4, Section 4.8, that conventional likelihood ratio test procedures are inappropriate for a formal hypothesis test of an abrupt change versus a quadratic bend.

Fitting the bent cable as an alternative model to the broken stick has been previously suggested by Seber and Wild ([10], Chapter 9). However, neither they nor the inventors of the model provide estimation theory for it. Later chapters of this thesis are devoted to developing least-squares estimation theory necessary for the validity of those procedures employed in Chapter 2.

We precede the new theory by Chapter 3 with a review of existing change-point models and theory. There, we will see that the available theory is mostly associated with segmented regression models less specialized than the bent cable. We will preview what is different in this thesis from these other works.

Chapter 4 presents the framework for the new bent-cable regression theory. Instead of considering multiple unknown parameters all at once, we confine our attention to the *basic bent cable* — one with fixed slopes of 0 and 1, respectively, for the incoming and outgoing linear phases. The center and half-width of the bend are the only unknown parameters. Assuming normally distributed random errors with a known, constant variance, conditions on the design are given to ensure regularity of the estimation problem, despite non-differentiability of the model's first partial derivatives (with respect to the covariate and model parameters). These are practical conditions which serve as a guideline for data collection in real-life settings. Under such conditions, the maximum likelihood estimators (MLE's) are shown (1) to be consistent, whether the underlying cable has a bend region or not; and (2) asymptotically to follow a bivariate normal distribution, if the underlying cable has a non-trivial bend. In the latter case, the deviance (or likelihood ratio test) statistic is shown to have an asymptotic chi-squared distribution with two degrees of freedom.

The case of no bend in the underlying regression function is an irregular boundary problem, and must be handled separately. It is discussed in Section 4.8, where we consider a basic cable whose zero bend width is the only unknown. We show that the maximum likelihood estimation error converges either (1) at a rate of $n^{-1/3}$ (where n is the sample size), and to an impractically complex limiting distribution; or (2) at a slower rate than $n^{-1/3}$. These negative results suggest that no practical benefits can be gained by pursuing the more general case of zero bend width with other bent-cable parameters unknown, and/or non-normal errors with an unknown variance.

We move on to the so-called *full bent cable* in Chapter 5. There, we extend the theory from Chapter 4 to the case of five unknown regression parameters, including the incoming and outgoing slopes and intercepts. A set of modified design conditions are given to compensate for the extra complexity of the problem. In this chapter, the normality assumption is relaxed, and the unrealistic assumption of a known error variance, removed. With normality removed, we now consider least-squares estimation. The usual asymptotic results stated in Chapter 4 (consistency, normality, and chi-square limit) indeed apply to the least-squares estimators (LSE's) and deviance statistic of this general case, provided that the underlying regression relationship contains all three segments.

To conclude the thesis, we provide in Chapter 6 a brief discussion of the theoretical results, and, in Chapter 7, of possible extensions to this thesis. In particular, Chapter 6 summarizes the simple logic behind the complex yet practical bent-cable theory. Chapter 7 suggests the Bayesian approach and the method of bootstrap regression for formally testing a sharp kink versus a quadratic bend. It also mentions a bent-cable model with a quartic transition region.

Chapter 2

Examples of Bent-Cable Regression

In this chapter, we consider modeling by a bent cable (Chapter 1, Figure 1.1) three datasets from natural phenomena which exhibit a potential change in their underlying response-covariate structure. The theoretical aspect of the estimation problem is discussed in detail in the subsequent chapters.

The first dataset is a time series on a collapsing sockeye salmon population, the second is from exercise physiology, and the third provides a valuable contrast from the physical sciences. All three are classic examples of change-point problems in which researchers have traditionally applied the structurally simple broken-stick model. Here, we instead apply the more general bent-cable model, which illustrates the inherent difficulty in supporting the abruptness notion of a broken-stick kink from analyses of typical sets of observations.

2.1 Sharply Declining Salmon... Or Not?

The scatter plot for the sockeye data appears in Figure 2.1. It portrays the declining abundance of sockeye salmon (*Oncorhynchus nerka*) in Rivers Inlet, British Columbia.[1] (Note the logarithmic scale.) This population has declined from being one of the largest in Canada to an endangered remnant. The cause of the collapse

[1] The data were obtained from Fisheries and Oceans Canada, Pacific Region.

remains uncertain. Researchers are uncertain also over the timing and abruptness of its onset.

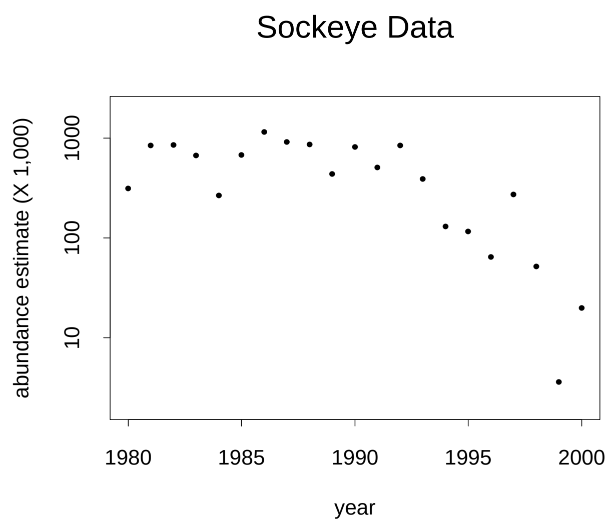

Figure 2.1: Estimates of spawning abundance (plotted on a logarithmic scale) vs. year for Rivers Inlet sockeye salmon (*Oncorhynchus nerka*). When does the decline begin, and how abrupt is the onset? The bent-cable model can be used to address these questions.

Did the decline begin abruptly around 1993? Or was it more gradual, possibly starting earlier? To address this question, a researcher would commonly fit a broken-stick model to the data for estimating the unknown change point. In such a case, an *a priori* abrupt change in the abundance-year relationship is assumed. However, the graph seems also to suggest a range of years between 1985 and 2000 over which abundance gradually dropped. Instead of constraining the model to have a break point, we allow for a smooth change by "extending" the break point to a curved "transition region" in the model — hence, the *bent cable*.

Since the kinked broken stick is merely a special case of the smooth bent cable (see Chapter 1, Figure 1.1), fitting the latter allows the investigator to disregard any notion of an *a priori* sharp corner. Instead, the bent-cable fit allows the data to reveal whether an abrupt change or a smooth transition is more convincing.

Here, we assume normally distributed random errors with a constant variance,[2] and use maximum likelihood estimation to fit a bent cable to these sockeye data. The bent-cable model, among others, recently appeared in [2] for modeling also fish spawner-recruitment relationships. However, methods for drawing inference on the abruptness of the change are not available in this article, nor in [9], the first article that introduced the bent-cable model. Formal inference based on maximum likelihood involves examining the behavior of the likelihood function. In particular, as the transition is our main focus, we examine the so-called profile deviance surface over the plane spanned by the transition parameters, as portrayed in Figure 2.2 below.

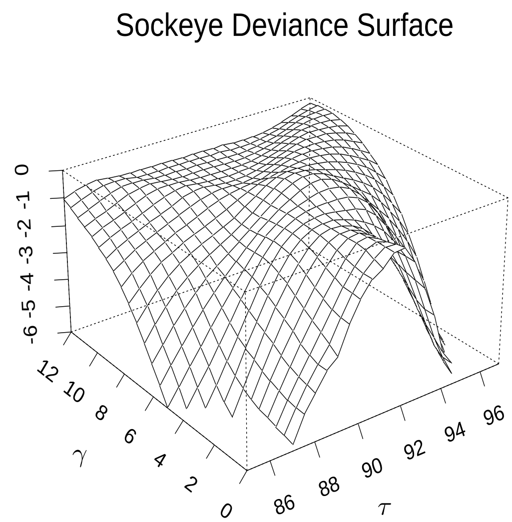

Figure 2.2: The profile log-likelihood (see [11] and Section 2.5.1) deviance surface vs. τ and γ (center and half-width of bend, respectively) for the data. All values of τ and γ with deviance values in the upper plateau are clearly consistent with the data. These include a cable bend which ranges over the entire dataset, and a broken stick with a corner at 1993.

As we will see in Chapter 5, the likelihood function involves five regression parameters. The transition or bend parameters are τ — the center of the bend, and γ — the half-width of the bend. When all other parameters in the regression model

[2] In Chapter 5, we will see that the normality assumption can be removed, so long as the random errors are i.i.d. and homoscedastic.

are profiled out ([11]), and the resulting log-likelihood surface magnified by a scale of 2, then translated to have a peak value of 0, we have obtained the profile deviance surface as shown. A computational algorithm is provided in Section 2.5.1. Formal mathematical derivations of the estimation theory appear in subsequent chapters.

For our sockeye data, this profile deviance peaks at the maximum likelihood estimates of $\widehat{\tau} = 1992$ and $\widehat{\gamma} = 6$. Hence, the estimated bend ranges from 1986 to 1998. However, the data are consistent with a broad range of other values. Any point for which the deviance is in the upper, roughly triangular plateau is clearly consistent with the data. These (τ, γ) pairs correspond to any cable whose bend begins at around 1986; a cable whose bend ranges over the entire dataset (1980 to 2000); and a broken stick (i.e. $\gamma = 0$) whose corner occurs anywhere around 1993. (See Section 2.5.2 for more details.) 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. The data shown are consistent with both interpretations.

2.2 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 relationship of lactate to oxygen uptake increases. The point when this change occurs has been called the lactate threshold, and is a key focus of training regimes ([12], [13]). 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 measure of workload begins to curve upwards ([14]–[17]). A more systematic method is to fit a broken stick to a graph of blood lactate versus workload ([14], [18], [19]), sometimes plotted on logarithmic scales ([14], [19]).

Until recently, it was widely believed that carbon dioxide output could also be used

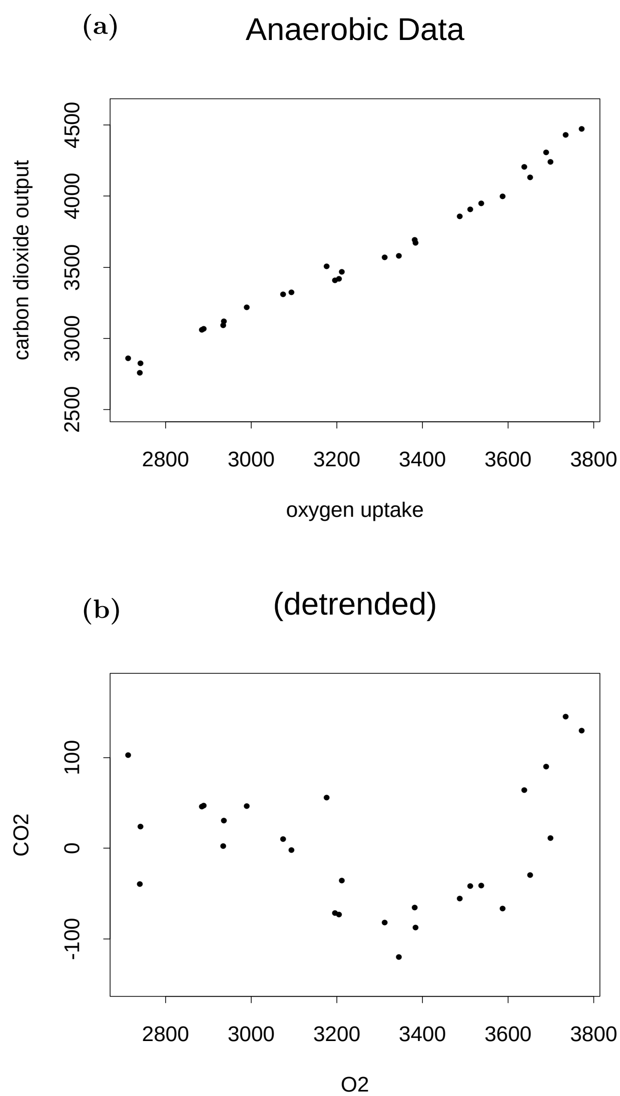

Figure 2.3: **(a)** Carbon dioxide (CO_2) output vs. oxygen (O_2) uptake in mL/s for an athlete on a treadmill with a continually increasing incline. **(b)** The same dataset detrended by removing the best linear least-squares fit. These data are consistent with either an abrupt threshold or a single quadratic curve. (See Section 2.5.2 for statistical details).

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 ([4], [5]), 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 (Figure 2.3 (a)).[3] The general increasing trend in the CO_2–O_2 relationship obscures any subtle features. To accentuate these, we present detrended data (Figure 2.3 (b)). This graph points to a change in the relationship somewhere in the vicinity of an oxygen uptake of 3,200. However, the best-fitting bent-cable model has a bend that ranges over oxygen uptake 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. (See Section 2.5.2 for statistical details.)

2.3 Convincing Evidence of a Smooth Transition

For comparison, we present data from a physics experiment in Figure 2.4, first published in R. A. Cook's Ph. D. thesis at Queen's University, as cited in [10]. (The data have also been analyzed in [20] with a hyperbolic transition model and with other multiphase regression models presented in [10].)

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 2.4) 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.

[3] The experiment was conducted according to a ramped workload protocol conducted on a treadmill at the Science North Science Centre in Sudbury, Ontario.

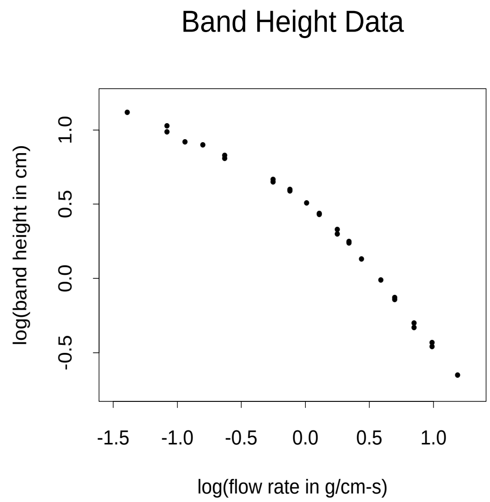

Figure 2.4: A plot of $\log_e$(stagnant-band-height) vs. $\log_e$(water flow rate), taken from [10]. Does the graph exhibit an abrupt break or a smooth transition?

The log-likelihood deviance surface (Figure 2.5 (a)) rules out any broken-stick fit at any reasonable confidence level such as 95%. The actual best-fitting cable has a bend ranging between log-flow rates of -0.373 and 0.484 (Figure 2.5 (b)). See Section 2.5.2 for more statistical details.

2.4 Implications

In typical biological situations where uncontrollable fluctuations are more substantial, data of this sort cannot provide evidence as definitive as did Cook's dataset. 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 model 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 if no laws of nature or auxiliary knowledge are available to support the abruptness notion. 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 an ocean predator known to feed on sockeye, 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 point to one or more sources whose influence took hold gradually or whose onset occurred at different instances over a range of years. Thus, interpreting the onset of the decline as abrupt without any solid evidence could lead to inappropriate conservation measures.

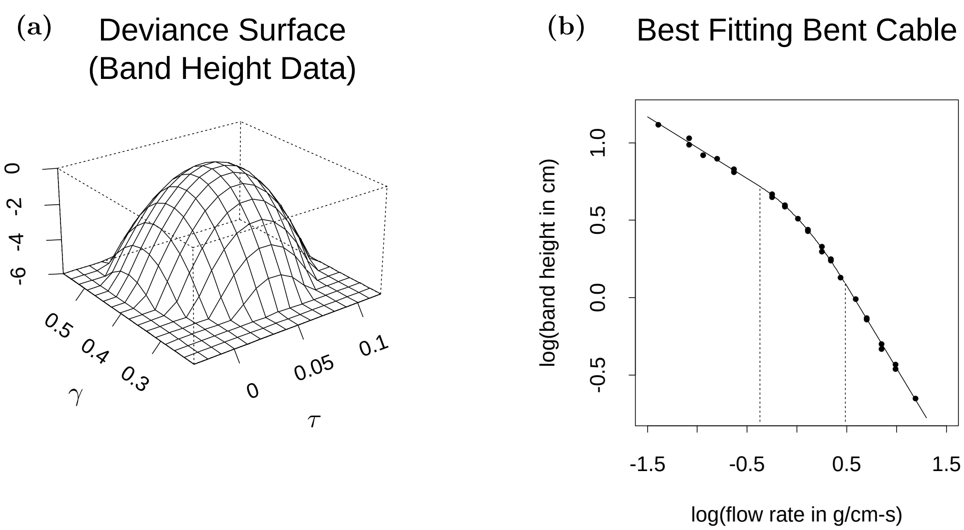

Figure 2.5: **(a)** The log-likelihood deviance surface vs. τ and γ. All deviance values above the horizontal plane at -5.99 are consistent with the data at an approximate 95% confidence level. (Refer to Section 2.5.2 for details.) All these values lead to cables with a long smooth bend. **(b)** The data are presented with the best-fitting bent cable. The bend ranges over log-flow rates of -0.373 to 0.484 (dotted lines).

2.5 Statistical Details

For the three examples presented in this chapter, we assume i.i.d. normal random errors, and use maximum likelihood to estimate the bent-cable regression parameters, as well as the constant error variance.[4] The regression function, evaluated at the covariate value x, is $f(x;\, \beta_0, \beta_1, \beta_2, \tau, \gamma) = \beta_0 + \beta_1\, x + \beta_2\, q(x;\, \tau, \gamma)$, where $q(x;\, \tau, \gamma) = \big[(x-\tau+\gamma)^2/(4\gamma)\big]\, \mathbf{1}\{|x-\tau| \le \gamma\} + (x-\tau)\, \mathbf{1}\{x > \tau+\gamma\}$ is the so-called basic bent cable of Chapter 4. Of the five model parameters, the center, τ, and half-width, γ, of the bend are the crucial transition parameters. A confidence region for (τ, γ) is particularly useful in assessing the nature of the underlying change. To this end, the method of profiling likelihoods in [11] is potentially valuable.

2.5.1 Computational Methods

The bent-cable model is partially linear in three parameters, namely, β_0, β_1, and β_2. The transition parameters, τ and γ, are non-linear. Denote the vector of linear parameters by $\boldsymbol{\beta}$, and that of the non-linear ones by $\boldsymbol{\alpha}$. In most cases, the error variance, σ^2, is unknown, and the regression problem therefore involves a sixth parameter. Denote the full log-likelihood by $\ell(\boldsymbol{\beta}, \boldsymbol{\alpha};\, \sigma^2)$, and the vector of maximum likelihood estimates (MLE's) by $\big(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}, \widehat{\sigma^2}\big)$. Due to the normality assumption, the MLE's and LSE's (least-squares estimates) for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are equivalent. In particular, $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ are both σ^2-free quantities. However, the MLE, $\widehat{\sigma^2} \equiv \widehat{\sigma^2}\big(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}\big)$, is the minimized value of the error mean-square (p. 72, formula (5.3)), and thus depends on the other MLE's.

Computing the conditional $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma^2}$ given $\boldsymbol{\alpha}$

As the bent-cable model is partially linear in $\boldsymbol{\beta}$, fixing $\boldsymbol{\alpha}$ allows us to maximize ℓ over $\boldsymbol{\beta}$ via linear least-squares. To do so, we can use a classic matrix-inversion method and compute $\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ — the MLE for $\boldsymbol{\beta}$-given-$\boldsymbol{\alpha}$. The details are as follows.

Denote the covariate-response pairs by $(x_1, Y_1), \ldots, (x_n, Y_n)$. Given any $\boldsymbol{\alpha}$, denote the $n \times 3$ "linear design matrix" (for estimating the linear parameters) by

[4] See footnote 2 on page 7.

$$\mathbb{X}(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & x_1 & q(x_1;\boldsymbol{\alpha}) \\ 1 & x_2 & q(x_2;\boldsymbol{\alpha}) \\ \vdots & \vdots & \vdots \\ 1 & x_n & q(x_n;\boldsymbol{\alpha}) \end{bmatrix} . \tag{2.1}$$

We assume $Y_i = f(x_i;\, \boldsymbol{\beta}, \boldsymbol{\alpha}) + \varepsilon_i$, where ε_i is the i-th random error in the response. Thus, $\boldsymbol{Y} = \big[\mathbb{X}(\boldsymbol{\alpha})\big]\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, and similarly for the vector $\boldsymbol{\varepsilon}$. Therefore, if $\mathbb{X}(\boldsymbol{\alpha})$ has full rank, then ℓ is maximized (and the error sum-of-squares minimized) over $\boldsymbol{\beta}$ at the conditional MLE

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \Big(\big[\mathbb{X}(\boldsymbol{\alpha})\big]^T \big[\mathbb{X}(\boldsymbol{\alpha})\big] \Big)^{-1} \big[\mathbb{X}(\boldsymbol{\alpha})\big]^T \boldsymbol{Y} \, . \tag{2.2}$$

To maximize ℓ also over σ^2, we take $\widehat{\sigma^2}$ to be the error mean-square minimized over $\boldsymbol{\beta}$ given $\boldsymbol{\alpha}$. That is,

$$\widehat{\sigma^2}\big(\boldsymbol{\alpha}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\big) = \frac{\big[\widehat{\boldsymbol{\varepsilon}}(\boldsymbol{\alpha})\big]^T \big[\widehat{\boldsymbol{\varepsilon}}(\boldsymbol{\alpha})\big]}{n} \, , \quad \text{where} \quad \widehat{\boldsymbol{\varepsilon}}(\boldsymbol{\alpha}) = \boldsymbol{Y} - \big[\mathbb{X}(\boldsymbol{\alpha})\big]\, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \, . \tag{2.3}$$

When maximized over both $\boldsymbol{\beta}$ and σ^2, the log-likelihood becomes a function of $\boldsymbol{\alpha}$, with the form $\ell\Big(\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}),\, \boldsymbol{\alpha};\, \widehat{\sigma^2}\big(\boldsymbol{\alpha}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\big)\Big)$. This is the profile log-likelihood surface over the (τ, γ)-plane.

The linear design matrix, $\mathbb{X}(\boldsymbol{\alpha})$, may have less than full rank at some $\boldsymbol{\alpha}$-values. This happens when the column of q's is collinear with the first two columns. In the case of a continuous covariate, this collinearity is likely to happen only if $x_i < \tau - \gamma$ for all i, or if $x_i > \tau + \gamma$ for all i, so that the given $\boldsymbol{\alpha}$ corresponds to having the observed data go through only one linear phase of the bent cable. Such an $\boldsymbol{\alpha}$ is, of course, not very meaningful in the context of bent-cable regression. For practical purposes, we shall only compute $\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ and $\widehat{\sigma^2}\big(\boldsymbol{\alpha}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\big)$ at those $\boldsymbol{\alpha}$'s such that the dataset goes through at least two segments of the corresponding bent cable. Visual inspection of a scatter plot usually provides a good sense for such $\boldsymbol{\alpha}$-values.

Computing the Profile Deviance

We multiply by 2 the profile log-likelihood as a function of $\boldsymbol{\alpha}$, then apply a vertical translation so that the surface has a peak value of 0. The height at any point on the resulting surface is thus a *deviance drop* (negative) of magnitude

$$D_{\text{prof}}(\boldsymbol{\alpha}) \;\equiv\; D\Big(\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}),\, \boldsymbol{\alpha};\, \widehat{\sigma^2}\big(\boldsymbol{\alpha}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\big)\Big) \;=\; 2\Big[\ell^{\max} - \ell\Big(\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}),\, \boldsymbol{\alpha};\, \widehat{\sigma^2}\big(\boldsymbol{\alpha}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\big)\Big)\Big]$$

$$\text{where}\quad \ell^{\max} \;=\; \ell\Big(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}}),\, \widehat{\boldsymbol{\alpha}};\, \widehat{\sigma^2}\big(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})\big)\Big) \;=\; \ell\Big(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}; \widehat{\sigma^2}\Big).$$

That is, $\ell^{\max}$ is the overall maximum of the full log-likelihood. For computational purposes, we can take

$$D_{\text{prof}}(\boldsymbol{\alpha}) = n \ln \frac{\widehat{\sigma^2}\big(\boldsymbol{\alpha}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\big)}{\widehat{\sigma^2}}$$

where $\widehat{\sigma^2}\big(\boldsymbol{\alpha}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\big)$ is the variance MLE from the linear fit given $\boldsymbol{\alpha}$, while $\widehat{\sigma^2}$ is the variance MLE from the full bent-cable fit (see Chapter 5, page 81, Theorem 5.3). Of course, the optimum value of D_{prof} is 0, and it occurs at the MLE, $\widehat{\boldsymbol{\alpha}}$.

At this stage, the actual value of $\widehat{\boldsymbol{\alpha}}$ (and hence, the full model variance MLE, $\widehat{\sigma^2}$) has yet to be determined. Indeed, the surfaces shown in Figures 2.2 and 2.5 (and 2.7 in the next section) are "approximate" profile deviance surfaces constructed by a "gridding" technique, as follows. Compute the conditional variance MLE, $\widehat{\sigma^2}\big(\boldsymbol{\alpha}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\big)$, at $\boldsymbol{\alpha} \in \big\{\boldsymbol{\alpha}_{11}, \ldots, \boldsymbol{\alpha}_{1k}, \boldsymbol{\alpha}_{21}, \ldots, \boldsymbol{\alpha}_{2k}, \ldots, \boldsymbol{\alpha}_{k1}, \ldots, \boldsymbol{\alpha}_{kk}\big\}$ — a set of plausible (τ_i, γ_j)-values which form a $k \times k$ Euclidean grid. Denote the (i, j)-th computed value by $\widehat{\sigma^2_{ij}}$. Let $\tilde{\sigma}^2 = \min_{1 \le i,j \le k} \widehat{\sigma^2_{ij}}$, which approximates the full model variance MLE, $\widehat{\sigma^2}$. Denote by $\tilde{\boldsymbol{\alpha}}$ the grid point at which $\tilde{\sigma}^2$ occurs, i.e. $\tilde{\sigma}^2 = \widehat{\sigma^2}\big(\tilde{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}(\tilde{\boldsymbol{\alpha}})\big)$. The approximated profile deviance surface is

$$-\tilde{D}_{\text{prof}}(\boldsymbol{\alpha}) = -n \ln \frac{\widehat{\sigma^2}\big(\boldsymbol{\alpha}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\big)}{\tilde{\sigma^2}}\,. \tag{2.4}$$

The actual plot of $-\tilde{D}_{\text{prof}}$ (such as appear in Figures 2.2 and 2.5) is constructed over the preset grid of $\boldsymbol{\alpha}_{ij}$'s. Of course, a different set of grid points leads to a slightly different $-\tilde{D}_{\text{prof}}$ surface.

For regular maximum likelihood problems, the (unapproximated) absolute deviance drop, D_{prof}, resembles a χ^2 statistic. In fact, by Chapter 5, Theorem 5.3, D_{prof} has an asymptotic χ^2_2 distribution when it is evaluated at the (true) $\boldsymbol{\alpha}$ of the underlying bent cable, provided that the true γ is strictly positive. Therefore, in the context of statistical confidence, an $\boldsymbol{\alpha}$ which yields a deviance drop that is too far below 0 is less plausible to be the true value, than one whose associated deviance drop is closer to 0. Thus, we can truncate the profile deviance surface ($-D_{\text{prof}}$ or $-\tilde{D}_{\text{prof}}$) along the vertical axis at say, -5.99 — the negative of the 5%-tail χ^2_2 critical value — and an approximate 95% confidence region for $\boldsymbol{\alpha}$ is formed by all those (τ, γ)-pairs enveloped under the truncated surface. Equivalently, we can construct a profile deviance contour plot projected onto the (τ, γ)-plane (Figure 2.6), and look for the region enclosed by the contour with a value of -5.99.

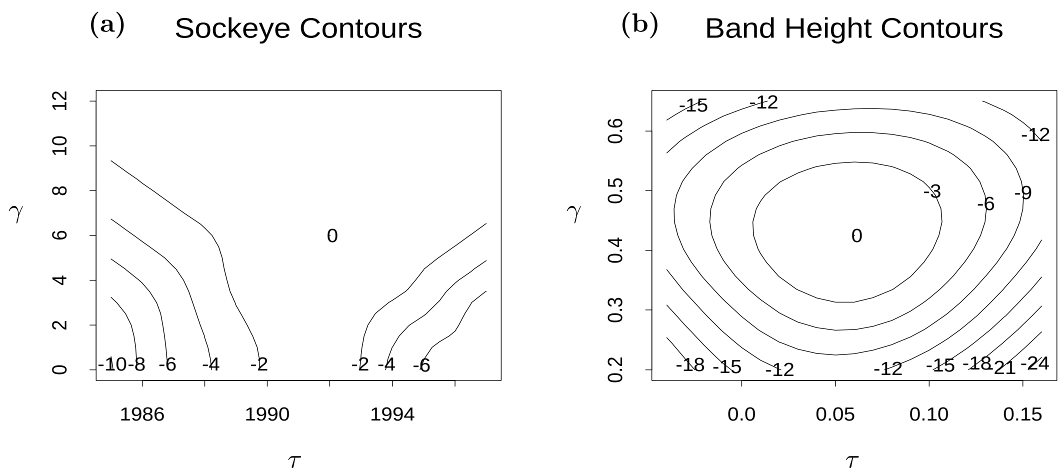

Figure 2.6: Profile deviance contours for **(a)** the sockeye data (Figures 2.1 and 2.2), and **(b)** the band height data (Figure 2.5). Both plots are constructed by the "gridding" technique described above. That is, they are projections of the "approximate" profile deviance surface, $-\tilde{D}_{\text{prof}}$, onto the $k \times k$ grid of preset (τ_i, γ_j)'s. The S-Plus function '`contour()`' automatically places '0' on the plot, marking $\tilde{\boldsymbol{\alpha}}$ — the grid point which gives the maximum value among the k^2 values of $-\tilde{D}_{\text{prof}}(\boldsymbol{\alpha}_{ij})$.

Whether by truncating the surface or identifying a particular contour, the reliability of the resulting confidence region depends heavily on the χ^2_2 approximation to the deviance statistic. Indeed, a paraboloidal deviance surface is an indication that the approximation is reasonable. Mathematically, the approximation is based on a two-term Taylor expansion of the log-likelihood about the true parameter value (see Chapter 4, page 47, relations (4.34) and (4.35)). This expansion is quadratic in the varying parameter, plus a remainder term. The quadratic term is exactly or very close to being χ^2-distributed (see Chapter 5, Theorem 5.3 (p. 81) and Lemma 5.5 (p. 90)).

If the remainder is negligible, then the log-likelihood or D_{prof} surface is quadratic, i.e. paraboloidal over the (τ, γ)-plane.

Computing the Global MLE's

From Figures 2.2 and 2.5 (and 2.7 in the next section), we see that even when the profile deviance surface is non-paraboloidal, it is nonetheless smooth everywhere. In fact, the log-likelihood function is continuously differentiable with respect to each model parameter (see, e.g. Chapter 1). Thus, given a reasonable initial estimate of $\boldsymbol{\alpha}$, a standard Gauss-Newton numerical algorithm can be applied to locate $\widehat{\boldsymbol{\alpha}}$ — the global maximizer of the profile deviance surface. The crucial element of the exercise is therefore identifying a reasonable initial value, $\widehat{\boldsymbol{\alpha}}^0$.

Given a $k \times k$ grid of preset $\boldsymbol{\alpha}_{ij}$'s, we can take $\widehat{\boldsymbol{\alpha}}^0$ to be $\tilde{\boldsymbol{\alpha}}$, which, among the $\boldsymbol{\alpha}_{ij}$'s, minimizes $\widehat{\sigma^2}\big(\boldsymbol{\alpha}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\big)$. However, the accuracy of this $\widehat{\boldsymbol{\alpha}}^0$ depends on the grid resolution. In particular, if the deviance surface is highly irregular (e.g. numerous local maxima and/or ridges), then a sparse grid may lead to an $\widehat{\boldsymbol{\alpha}}^0$ that is too imprecise for the Gauss-Newton algorithm to converge successfully on a final $\widehat{\boldsymbol{\alpha}}$. In some instances, the gridding may need several refinements to produce a good $\widehat{\boldsymbol{\alpha}}^0$. Upon convergence, the global MLE's are $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}} \equiv \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})$, and $\widehat{\sigma^2} \equiv \widehat{\sigma^2}\big(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})\big)$.

Occasionally, the numerical procedure fails to converge altogether, despite an impeccably high resolution for the (τ, γ)-grid. From our experience, we have identified the usual source of this non-convergence problem. It usually arises from an inadequate sample size or a poor design (such as observations being placed too sparsely apart). Either may be combined with a large response variation, resulting in a dataset which contains insufficient information about the model parameters. In Sections 4.4 and 5.2.1, we provide a more in-depth discussion of design conditions which provide reliable parameter estimates in bent-cable regression.

Computational Algorithm and S-Plus Code

Below is a summary of the computational methods presented thus far.

[CA1] Construct a scatter plot for the data, and assign accordingly a grid of plausible values of $\boldsymbol{\alpha}_{ij} = (\tau_i, \gamma_j)$, where $i, j = 1, \ldots, k$ (= grid dimension).

[CA2] Do the following for each (i, j):

(a) compute the linear design matrix, $\mathbb{X}(\boldsymbol{\alpha}_{ij})$, as given by (2.1);

(b) compute the conditional MLE's, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}_{ij})$ and $\widehat{\sigma^2}\big(\boldsymbol{\alpha}_{ij}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}_{ij})\big)$, as given by (2.2) and (2.3); equivalently, regress the observed responses, $\boldsymbol{y}$, on $\mathbb{X}(\boldsymbol{\alpha}_{ij})$ by using a standard linear least-squares algorithm, then obtain the coefficient and variance estimates from the output.

[CA3] Locate $\tilde{\boldsymbol{\alpha}} = \arg\min_{1 \le i,j \le k}\big\{\widehat{\sigma^2}\big(\boldsymbol{\alpha}_{ij}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}_{ij})\big)\big\}$.

[CA4] If so desired, a plot of the approximate profile deviance surface and/or contours may be produced by computing

(a) the approximate variance MLE, $\tilde{\sigma}^2 = \widehat{\sigma^2}\big(\tilde{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}(\tilde{\boldsymbol{\alpha}})\big)$, and

(b) $-\tilde{D}_{\text{prof}}(\boldsymbol{\alpha}_{ij})$ for all $i, j = 1, \ldots, k$, according to (2.4).

[CA5] Assign the following initial values:

$$\widehat{\boldsymbol{\alpha}}^0 := \tilde{\boldsymbol{\alpha}}\,, \quad \widehat{\boldsymbol{\beta}}^0 := \widehat{\boldsymbol{\beta}}(\tilde{\boldsymbol{\alpha}})\,.$$

[CA6] Apply a standard non-linear least-squares algorithm (e.g. Gauss-Newton) to the data, $(x_1, y_1), \ldots, (x_n, y_n)$, by specifying the regression model (5.1) of Chapter 5, and the initial estimates, $(\widehat{\tau}^0, \widehat{\gamma}^0) = \widehat{\boldsymbol{\alpha}}^0$ and $(\widehat{\beta_0}^0, \widehat{\beta_1}^0, \widehat{\beta_2}^0) = \widehat{\boldsymbol{\beta}}^0$.

[CA7] If the algorithm from [CA6] fails to converge, refine the grid of [CA1] and repeat [CA2] to [CA7] until convergence (in which case the global MLE's are obtained), or upon meeting a repetition tolerance (in which case the dataset is deemed uninformative about the bent-cable parameters).

All bent-cable regressions in this chapter utilize a code, written in S-Plus, that corresponds to the above algorithm. As an example, we provide the application of this code to the sockeye data of Figure 2.1.

```
> t(sockeye)       # data frame 'sockeye' already exists
             1      2      3      4      5      6       7      8      9     10
   Year   1980   1981   1982   1983   1984   1985    1986   1987   1988   1989
Returns 313522 851781 862178 671662 268180 684974 1163000 920563 875025 438926

            11     12     13     14     15     16    17     18    19   20    21
   Year   1990   1991   1992   1993   1994   1995  1996   1997  1998 1999  2000
Returns 820777 514726 851073 394146 131820 117197 65000 276000 52000 3600 20000

> dat <- sockeye       # rename data frame

> dat[,2] <- log(dat[,2])       # log-transform ``Returns''

> dimnames(dat)[[2]] <- c("x","y")      # assign robust names to attributes

> attach(dat)

#>>>>>>>>>>>>>>>>> Step [CA 1] >>>>>>>>>>>>>>>>>>>

> plot(x,y)      # scatter plot

> tau <- seq(1985,1997,.5); gamm <- seq(0,12,.5)
      # grid of 25-by-25 plausible values obtained based on scatter plot

#<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<


> design.mat <- cbind(rep(1,n),x,rep(NA,n))
      # linear design matrix: 3rd column to be computed

> dimnames(design.mat) <- list(NULL,c("int","x","cable.x"))
      # assign column names

> design.mat <- data.frame(design.mat)        # convert into data frame

> k <- length(tau)       # grid dimension


#>>>>>>>>>>>>>>>>> Step [CA 2] >>>>>>>>>>>>>>>>>>>

> loglik.mat <- loglik.surf(c(1:k),c(1:k),design.mat)
      # compute over grid the numerator of profile deviance, which is
      #       proportional to the log-likelihood
      # see code for 'loglik.surf()' below

#<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
```

```
#>>>>>>>>>>>>>>>>>>> Steps [CA 3], [CA 4] >>>>>>>>>>>>>>>>>>

> dev.mat <- loglik.mat-max(na.omit(loglik.mat))
      # first remove NA's (negative infinity), then
      #    subtract from each 'loglik.mat' value the maximum to get
      #        approximate profile deviance drop over grid

> persp(tau,gamm,dev.mat,xlab="tau",ylab="gamma",zlab="chi-sq")
      # perspective plot of approximate deviance surface

> contour(tau,gamm,dev.mat,xlab="tau",ylab="gamma")
      # contour plot of same surface

#<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<



#>>>>>>>>>>>>>>>> Step [CA 5] >>>>>>>>>>>>>>>>>>>

> dev.mat==0     # look for co-ordinate that gives peak value (=0) to deviance
      # full output snipped

> dev.mat[15,13]      # peak co-ordinate is (i,j)=(15,3)
[1] 0

> tau[15]; gamm[13]       # get initial value (approx. MLE) for alpha
[1] 1992
[1] 6

> t0 <- 1992; g0 <- 6       # assign inital alpha

> design.mat[,3] <- basic.q(x-t0,g0)
     # 3rd column of corresponding design matrix
     # 'basic.q()' is the one-parameter basic bent cable: see below for code

> fit0 <- lm(formula=y~x+cable.x,data=design.mat)
      # ordinary least squares given initial alpha

> b00 <- fit0$coef[1]; b10 <- fit0$coef[2]; b20 <- fit0$coef[3]
      # assign initial beta

#<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
```

```
#>>>>>>>>>>>>>>>>> Step [CA 6] >>>>>>>>>>>>>>>>>>>

> fit <- nls(formula=y~f.cable(x,b0,b1,b2,tau,gamm),data=dat,trace=T,
                                 start=list(b0=b00,b1=b10,b2=b20,tau=t0,gamm=g0))
     # use built-in non-linear least-squares algo 'nls()' to find global MLE's
     # 'f.cable()' is full bent-cable formula: see code below
     # supply initial values for all parameters in 'f.cable()'
     # 'trace' outputs all iterations of the 'nls()' algo (Gauss-Newton)

8.68715 : -146.967 0.0808324 -0.680622 1992 6
8.68203 : -148.631 0.0816722 -0.701644 1992.19 6.2795
                .
                .
                .
8.68046 : -145.149 0.0799159 -0.695169 1992.17 6.16178
8.68046 : -145.116 0.0798991 -0.695231 1992.17 6.1618

>    # the global MLE's are shown on the last line with
     #    error sum-of-squares minimized at 8.68046

#<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<

#>>>>>>>>>>>>>>>>> put fitted bent cable on scatter plot >>>>>>>>>>>>>>>>>>>

> mle <- fit$param
> mle
 b0.(Intercept)        b1.x  b2.cable.x       tau      gamm
       -145.116  0.07989911  -0.6952314  1992.166  6.161802

> plot(x,y)

> lines(x,f.cable(x,mle[1],mle[2],mle[3],mle[4],mle[5]))     # see 'f.cable()' below

> abline(v=c(mle[4],mle[4]-mle[5],mle[4]+mle[5]))     # mark transition region

#<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<

#>>>>>>>>>>>>>>>>> functions used in computational algo >>>>>>>>>>>>>>>>>>>

loglik.surf <- function(i.vect, j.vect, design)
          # for Step [CA 2]
          # see code for 'loglik.given.g()' below
{
        sapply(j.vect, loglik.given.g, i.vect = i.vect, design = design)
               # loop through the tau.i's and gamma.j's
               # see code for 'loglik.given.g()' below
}
```

```
loglik.given.g <- function(i.vect, j, design)
         # submodule for 'loglik.surf()'
{
        sapply(i.vect, loop, j = j, design = design)
               # fix j-th gamma and loop through tau.i's
               # see code for 'loop()' below
}

loop <- function(i, j, design)
        # submodule for 'loglik.given.g()' to compute conditional beta.hat, etc.
{
        t.hat <- tau[i]
        g.hat <- gamm[j]
            # 'tau' and 'gamm' are globally defined grid points
        design[, 3] <- basic.q(x - t.hat, g.hat)
            # compute 3rd column of linear design matrix
            # see code for 'basic.q()' below
        fit <- try(lm(formula = y ~ x + cable.x, data = design))
                 # (i,j)-th linear least-squares fit
                 # function 'try()' prevents a core dump when
                 #     design matrix has less than full rank
        if(length(fit) == 11)     # a proper 'lm' object has 11 attributes
        {
                return( - dim(design)[1] * log(summary(fit)$sigma^2))
                     # numerator of profile deviance, which is
                     #     proportional to the log-likelihood
        }
        else return(NA)     # problem with rank
}

basic.q <- function(z, g)       # one-parameter basic bent-cable function
{
        if(g > 0)
                ((z + g)^2/(4 * g)) * (z >=  - g & z <= g) + z * (z > g)
        else z * (z > 0)
}


f.cable <- function(x, b0, b1, b2, tau, gamm)   # full bent-cable model
{
        b0 + b1 * x + b2 * basic.q(x - tau, gamm)
}
```

2.5.2 Analyses

Asymptotic theory from Chapter 5 suggests that any (τ, γ)-pair for which the profile deviance is above -5.99 (based on the χ^2_2 distribution — see Theorem 5.2) is consistent with the data at a 95% level of confidence. However, the asymptotic theory hinges on the surface's resemblance to a paraboloid. This surface is clearly not so for the sockeye data (Figure 2.2); therefore the sample size there is most likely too small for the asymptotics to apply. Meanwhile, the surface in Figure 2.5 (a) for the physics example provides a vivid contrast in its resemblance to a paraboloid. The cutoff value of -5.99 is more trustworthy in that application. In fact, the surface remains highly paraboloidal even for a cutoff value of -10, which corresponds to a confidence level of more than 99%. Moreover, the value of $\gamma = 0$ (a broken stick) is well excluded from this 99+ percent coverage. Hence, the evidence for a smooth transition is overwhelming.

Despite an irregular sockeye deviance surface, the values on the triangular plateau there are so close to 0 that the corresponding values of τ and γ are almost certainly consistent with the data. In particular, the plateau is roughly defined by the intersection of the regions $\tau - \gamma \leq 1986$ (i.e. a smooth transition beginning no later than 1986) and $\tau + \gamma \geq 1998$ (i.e. a transition ending no earlier than 1998). A point in this region is, for example, (1990, 10), which gives a bent-cable model whose bend stretches over the entire range of the dataset. Additionally, note that the surface along the τ-axis (i.e. $\gamma = 0$) has a peak at $\tau \approx 93$. This local peak is not far below the height of the triangular plateau, and hence, $(\tau, \gamma) = (1993, 0)$ — a sharp change in 1993 — is also highly consistent with the data.

Similarly, for the anaerobic data of Figure 2.3, the deviance surface (Figure 2.7) has an upper diagonal ridge roughly along $\tau - \gamma = 3350$ for $\tau \geq 3500$, suggesting that numerous models are consistent with the data at a high level of confidence. One such model is a cable whose bend begins at an oxygen uptake of about 3,350 and continues through the remainder of the O_2 values. Furthermore, 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.

If we further profile out τ from the deviance surface, γ becomes the only parameter

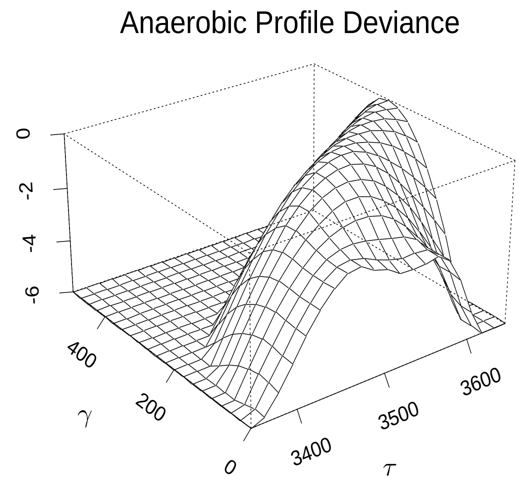

Figure 2.7: The profile log-likelihood deviance surface vs. τ and γ for the anaerobic data of Figure 2.3.

of the one-dimensional profile deviance function. One can visualize projecting the surfaces in Figures 2.2 and 2.7 onto the (γ, z)-plane (z being the variable along the vertical axis), by compressing the surface along the τ-axis to obtain the one-parameter profile deviances. We could then easily recognize another interesting aspect to the irregularity associated with widely scattered data: this profile deviance is entirely flat between $\gamma = 0$ and $\gamma = |x^* - \widehat{\tau}|$, where x^* is the observation closest to $\widehat{\tau}$. Indeed, this phenomenon agrees with the theoretical results of Lemma 4.6 (Chapter 4, Section 4.8.7). There, γ is the only unknown of the model, and it is 0, in which case the likelihood as a function of γ is flat everywhere between 0 and $|x^* - \tau|$.[5] As the same flatness is evident in Figures 2.2 and 2.7, we have some quantitative indication that each of those underlying response-covariate relationships could exhibit a sharp kink. That is, we may be unnecessarily modeling an underlying broken stick by a smooth bent cable. However, it is also possible that the true relationship has a smooth transition, but the data are too sparsely collected for us to detect this smoothness. In Sections 4.4 and 5.2.1 of the subsequent chapters, we will discuss the important role played by the covariate design in detecting the true nature of the transition.

[5]In Section 4.8.7, τ is known and is taken to be 0.

Inspecting an irregular profile deviance alone is, by no means, a likelihood-based formal test of $H_0 : \gamma = 0$ vs. $H_1 : \gamma > 0$. The latter is deemed too complex and impractical, due to the peculiar asymptotics in the case of an underlying zero γ (see Chapter 4, p. 54, Theorem 4.3). In Chapter 7, we will suggest alternative procedures for testing these hypotheses.

Chapter 3

Precedent Models and Theory

In Chapter 1, we have mentioned that the bent-cable model was originally developed by Tishler and Zang ([9]) to handle the non-differentiability of the first derivative when fitting a broken-stick model. Indeed, non-differentiability leads to more complex asymptotics, and thus has been a primary concern in (continuous) segmented regression theory.

In most contexts, the *change point*, or the x-coordinate of the join point between segments, is of main interest. Also of importance is the notion of whether the underlying response-covariate relationship truly exhibits a change. For a two-phase model, the latter can be formulated into a hypothesis test of one phase only against two phases joined at some change point. However, the theory of this testing problem is intrinsically complex due to the presence of unidentifiable model parameters under the null hypothesis (see [21]). In [22] and [23], Hinkley has suggested empirical evidence of an asymptotic χ_3^2 distribution for the classic F-test (deviance) statistic. In [24], Gallant and Fuller test for a (two-phase) quadratic-linear model against a (three-phase) quadratic-quadratic-linear model, also based on the notion of an asymptotically F-distributed deviance statistic. However, rigorous verifications have not been pursued in any of [22] to [24].

On the contrary, the problem of estimating the change point has been widely studied. One of the earliest articles on the subject of segmented regression was [25] by Sprent. It discusses estimation of piecewise-linear models, which are popular in the biological context. More development on the subject is provided in [22] and [23] by Hinkley. The natural extension to a piecewise-linear model is a segmented model comprising general continuous functions that are not necessarily straight lines. Hudson ([26]) was one of the pioneers on this front. However, most of these earlier articles deal with the estimation procedure without providing much details on asymptotic properties of the estimators.

Since then, Feder ([27]), Gallant ([28], [29]), and Ivanov ([30]), among others, have discussed asymptotics for segmented regression given a set of assumptions mostly for establishing model regularity.

In [27], Feder's principal model assumption is continuity of the underlying function. As no other smoothness constraints are imposed, a kinked fit is possible for a true relationship that happens to be smooth. In practice, while one cannot be sure if the underlying function is smooth or kinked, it appears unorthodox in model fitting to consider such a vast class in which the candidate models can be entirely different qualitatively. Moreover, if the underlying model has an odd order of smoothness (number of continuous derivatives plus one), then the asymptotics are radically different compared to those for an even order.

Thus, Feder's estimation problem is altogether different from ours. Here, we consider one that is more specialized. We have a once-differentiable underlying model, and consider only candidate models from within the same class. In [28] and [29], Gallant's model assumptions are similar, and he imposes elaborate regularity conditions. For example, for a once-differentiable quadratic-quadratic model, consistency and asymptotic normality of the least-squares estimator (LSE) for the unknown joint is established by, among other requirements, (i) taking the design points $x_1, \ldots, x_n$ to be "near" replicates of a basic design with five distinct covariate values (see [29], p. 26; [28], pp. 6–7), and (ii) restricting the search for candidate knots to within a compact subset trapped between two consecutive covariate values from (i) (see [29], p. 26).

Both Gallant and Ivanov (in [30]) establish consistency of LSE's by assuming a compact parameter space. In [30], Ivanov additionally assumes a condition which relates the candidate model to the variability of the response errors (see [30], p. 30, expression (3.13)). This latter condition may appear somewhat unintuitive. Asymptotic normality is established later in the book by assuming twice-differentiability of the regression function.

Segmented regression is a very practical tool for studying natural phenomena, and has been widely used for many decades. However, little rigorous theory has been developed without sacrificing practicality of the regularity conditions.

In Chapters 4 and 5 of this thesis, we provide some structurally simple design conditions for our once-differentiable bent-cable model. These conditions can be easily met in real-life settings, and are thus, in our view, very practical. We prove that they suffice to establish regularity for our specialized problem. Additionally, we show in Section 4.8 that, despite its common usage, the classic F-test is unreliable for assessing whether the underlying function exhibits a smooth transition or an abrupt change.

Chapter 4

The Basic Bent Cable

As a function of the regression covariate, x, the basic bent-cable model has an incoming linear phase of slope 0 and an outgoing linear phase of slope 1. A quadratic segment with half-width γ smoothly joins these two linear ones. When extrapolated, the two linear segments intersect at $x = \tau$. If γ is strictly positive, the resulting function is continuously differentiable with respect to each argument. The continuity constraint on the derivative forces the join points to occur at $x = \tau \pm \gamma$. The resulting regression function is then

$$q(x;\tau,\gamma) = \frac{(x-\tau+\gamma)^2}{4\gamma}\,\mathbf{1}\big\{|x-\tau| \le \gamma\big\} + (x-\tau)\,\mathbf{1}\big\{x-\tau > \gamma\big\} \tag{4.1}$$

where $\tau \in (-\infty, \infty)$ and $\gamma \in [0, \infty)$. Note that a bent cable with $\gamma = 0$ reduces to a non-differentiable broken stick. (See Figure 1.1 in Chapter 1). Therefore, with $\gamma > 0$, the bent cable exhibits a transition region that spans $[\tau-\gamma, \tau+\gamma]$, as opposed to a change point at τ for the broken stick.

Whether γ is zero or positive, the second partial derivatives of the bent-cable model do not exist everywhere. This irregularity prevents us from directly applying large-sample theory that is readily available for maximum likelihood problems that are regular.

For modeling natural phenomena, the two-parameter basic bent cable is often inadequate due to its fixed slopes. A five-parameter full bent-cable model is more sensible in most real-life settings, such as appear in Chapter 2. In Chapter 5, we will discuss in detail how the basic bent cable theory can be extended to the practical full model. In this chapter, we focus on laying down the framework for the more complicated model through the basic case.

4.1 Basic Bent-Cable Regression

We consider observations $\{(x_i, Y_i)\}_{i=1}^n$ generated by a basic bent cable. The regression model is

$$Y_i = q(x_i\,; \boldsymbol{\theta}_0) + \varepsilon_i\,, \qquad \text{where} \qquad \varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2) \text{ for a known } \sigma^2 \tag{4.2}$$

and q is the basic bent cable as in (4.1), $\boldsymbol{\theta}_0 = (\tau_0, \gamma_0)$ is the underlying bent-cable parameter.

For the estimation problem, we consider the unbounded parameter space, $\Omega = (-\infty, M] \times [0, \infty)$ for $\boldsymbol{\theta}_0$, and the open regression domain, $\mathcal{X} = \mathbb{R}$. Here, M is some large positive but finite upper bound for the candidate τ-values. (We defer the discussion of the use of this hard bound until we investigate the full model in Chapter 5.) The natural lower bound for candidate γ-values is, of course, zero. Any basic bent cable $q(x; \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Omega$ is a candidate model in estimating $\boldsymbol{\theta}_0$. While searching only within the class of basic bent cables (as opposed to the class of all three-phase models in [27] by Feder), we do allow more flexibility than does Gallant in [28] and [29], or Ivanov in [30]. (See Chapter 3 for a discussion of these works.)

Besides an unspecified upper bound on the τ-space, our only regularity conditions are placed on the design points, $x_1, \ldots, x_n$. The practical value in these conditions is that they indicate a design which enables the investigator to ensure that data are collected at appropriate x-locations for reliable estimation of the underlying parameters.

4.2 Results

We consider estimation of the bent-cable parameters via maximum likelihood. Given a set of conditions on the location of the design points, $x_1, \ldots, x_n$ (see Section 4.4 below), the MLE's for τ_0 and γ_0 (i) are consistent when $\gamma_0 \geq 0$, and (ii) asymptotically follow a bivariate normal distribution when $\gamma_0 > 0$. In the latter case, the likelihood ratio test statistic is shown to have an asymptotic chi-squared distribution with 2 degrees of freedom. Therefore, the case of $\gamma_0 > 0$ is regular under the given conditions. These results are formally restated as lemmas and theorems in Section 4.5. Asymptotic distributional properties for the case of $\gamma_0 = 0$ appear in Section 4.8.

In Chapter 5 where we investigate the full bent cable, we will see that the distributional results for $\gamma_0 > 0$ of this chapter remain valid even with normality of the ε_i's removed. In addition, the usual normal-theory MLE of σ^2 is consistent in the case of an unknown but constant error variance. For now, we will focus on the basic theory.

4.3 The Maximum Likelihood Estimator

With normally distributed errors, $\widehat{\boldsymbol{\theta}}_n$, the MLE of $\boldsymbol{\theta}_0$, is equivalent to the least-squares estimator. In particular, the log-likelihood function for $\boldsymbol{\theta} \in \Omega$ is

$$\ell_n(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\big[Y_i - q(x_i;\boldsymbol{\theta})\big]^2 + \text{constant}$$

which is linear in the error sum-of-squares function.

When a sample yields multiple maximizers of the log-likelihood function, we take $\widehat{\boldsymbol{\theta}}_n$ to be the one selected sequentially as follows: (i) pick out the one(s) with the least vector norm; (ii) keep the one(s) with the least γ-value; (iii) if necessary, select that with the least τ-value.

Before stating the results on the asymptotic properties of $\widehat{\boldsymbol{\theta}}_n$, we need to introduce some notation. Let us define

$$d_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(x) = q(x;\boldsymbol{\theta}_1) - q(x;\boldsymbol{\theta}_2) \quad , \quad T_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_1^n \Big| d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\Big|^2 ,$$

$$S_n(\boldsymbol{\theta}) = \sum_1^n \Big|\varepsilon_i + d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\Big|^2 , \quad \overline{S}_n(\boldsymbol{\theta}) = \frac{1}{n}S_n(\boldsymbol{\theta}) \quad , \quad H_n(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}_0}\Big[\overline{S}_n(\boldsymbol{\theta})\Big] ,$$

$$U_{n\tau}(\boldsymbol{\theta}) = \frac{\partial \ell_n}{\partial \tau}(\boldsymbol{\theta}) , \quad U_{n\gamma}(\boldsymbol{\theta}) = \frac{\partial \ell_n}{\partial \gamma}(\boldsymbol{\theta}) \quad , \quad \boldsymbol{U}_n(\boldsymbol{\theta}) = \big(U_{n\tau}(\boldsymbol{\theta}), U_{n\gamma}(\boldsymbol{\theta})\big)^T ,$$

$$\mathbb{V}_n(\boldsymbol{\theta}) = \boldsymbol{\nabla}\boldsymbol{U}_n(\boldsymbol{\theta}) \quad \text{(wherever defined)} \quad , \quad \mathbb{I}_n(\boldsymbol{\theta}) = \mathrm{Cov}_{\boldsymbol{\theta}_0}\big[\boldsymbol{U}_n(\boldsymbol{\theta})\big] .$$

Note that the function $d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}$ is a pointwise distance between the true and candidate bent cables. One can easily see that for a $\boldsymbol{\theta}$ which is distant from $\boldsymbol{\theta}_0$, the value of $d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)$ is large for some i's, and hence, so are $T_n(\boldsymbol{\theta})$ and $H_n(\boldsymbol{\theta})$. In turn, the sum of squares function $S_n(\boldsymbol{\theta})$ is large in probability. This supports the notion that $\widehat{\boldsymbol{\theta}}_n$, the minimizer of S_n, is probably close to $\boldsymbol{\theta}_0$ for all large n.

For a regular estimation problem in which the log-likelihood is at least twice-differentiable, the definition for the Fisher Information is the negative expected Hessian at $\boldsymbol{\theta}_0$. And by regularity, this negative expected value is equivalent to the covariance of the vector-valued score, $\boldsymbol{U}_n$, evaluated at $\boldsymbol{\theta}_0$.

With conditions [A] to [D] given in Section 4.4 below, the only irregularity for the case of $\gamma_0 > 0$ lies in a Hessian, $\mathbb{V}_n$, that is not well-defined everywhere. However, condition [D] preserves regularity at the true $\boldsymbol{\theta}_0$. Thus, the identity between the negative expected Hessian and the covariance matrix holds true at $\boldsymbol{\theta}_0$. As a result, the matrix $\mathbb{I}_n(\boldsymbol{\theta}_0)$ here is indeed the Fisher Information matrix in its usual sense.

4.4 Conditions for the Basic Bent Cable

Given $\delta > 0$ and sequence $\xi_n \downarrow 0$, define

$$
\begin{aligned}
c_0(\delta) &= \liminf_{n \longrightarrow \infty} \frac{1}{n} \sum \mathbf{1}\Big\{ |x_i - \tau_0| \le \gamma_0 + \delta \Big\} \\
c_-(\delta) &= \liminf_{n \longrightarrow \infty} \frac{1}{n} \sum \mathbf{1}\Big\{ |x_i - \tau_0| \le \gamma_0 - \delta \Big\} \\
c_+(\delta) &= \liminf_{n \longrightarrow \infty} \frac{1}{n} \sum \mathbf{1}\Big\{ x_i \ge \tau_0 + \gamma_0 + \delta \Big\} \\
\zeta(\xi_n) &= \frac{1}{n} \sum \mathbf{1}\left\{ \Big| |x_i - \tau_0| - \gamma_0 \Big| \le \xi_n \right\}
\end{aligned}
$$

The regularity conditions on the design are

[A_1] If $\gamma_0 = 0$, then $\forall\ \delta > 0$, $c_0(\delta) > 0$.

[A_2] If $\gamma_0 > 0$, then $\exists\ \delta_1 > 0$ such that $c_-^* \equiv c_-(\delta_1) > 0$.

[B] For $\gamma_0 \ge 0$, $\exists\ \delta_{11} > 0$ such that $c_+^* \equiv c_+(\delta_{11}) > 0$.

[C] If $\gamma_0 > 0$, then $\forall\ \xi_n \downarrow 0$, $\zeta(\xi_n) \longrightarrow 0$.

[D] If $\gamma_0 > 0$, then $x_i \neq \tau_0 \pm \gamma_0 \; \forall \; i = 1, \ldots, n$.

In practice, conditions [A] to [D] serve as a set of guidelines for designing the experiment. To reliably estimate the parameters of an underlying basic bent cable, a non-trivial fraction of observations must be collected from well inside the bend $[\tau_0 - \gamma_0, \tau_0 + \gamma_0]$ in the case of $\gamma_0 > 0$ ([A_2]), and from within all neighborhoods of the break point τ_0 if $\gamma_0 = 0$ ([A_1]). In addition, a non-trivial fraction must come from the outgoing linear phase ([B]). While the true parameters are unknown, the investigator's expertise in the subject matter should nonetheless suggest a range of values which are highly likely inside a certain phase of the model. Condition [C] prevents the accumulation of too many observations exactly at the join points of the underlying bent cable. Condition [D] is used to avoid technical difficulties in defining the Fisher Information, but it could be eliminated at the cost of some notational complexity. Technicality aside, it is not unreasonable, in the case of a continuous predictor variable, to rule out exact equality of any observed predictor and an underlying join point.

4.5 Formal Statements of Results

Lemma 4.1 (Identifiability). *Given two points (w_0, z_0) and (w_1, z_1) such that*

$$z_i = q(w_i \,; \boldsymbol{\theta}_0) \quad \forall \quad i = 0, 1 \tag{4.3}$$

$$w_0 \begin{cases} \in \; [\tau_0 - \gamma_0 + \delta_1, \tau_0 + \gamma_0 - \delta_1] & \text{if} \;\; \gamma_0 > 0 \\ = \; \tau_0 & \text{if} \;\; \gamma_0 = 0 \end{cases} \tag{4.4}$$

$$w_1 \in [\tau_0 + \gamma_0 + \delta_{11}, \infty) \tag{4.5}$$

with δ_1 and δ_{11} as given in [A] and [B], then

$$q(w_i \,; \boldsymbol{\theta}) = z_i \; \forall \; i = 0, 1 \quad \Longrightarrow \quad \boldsymbol{\theta} = \boldsymbol{\theta}_0 \; .$$

Furthermore, if $\gamma_0 = 0$, suppose (4.3) and (4.5) hold, but

$$w_0 \in [\tau_0 - \epsilon^*, \tau_0 + \epsilon^*] \tag{4.6}$$

where $\epsilon^* \in (0, \delta/4)$ and $\delta \in (0, \delta_{11}/2)$. Then

$$q(w_i \,; \boldsymbol{\theta}) = z_i \; \forall \; i = 0, 1 \quad \Longrightarrow \quad |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq \delta \, . \tag{4.7}$$

A consistent parameter estimator cannot be obtained for an unidentifiable model. Note that (4.3) to (4.6) in the lemma are directly related to conditions [A] and [B] which are required for the consistency of $\widehat{\boldsymbol{\theta}}_n$. For the boundary problem of $\gamma_0 = 0$, (4.7) is needed to prove Theorem 4.1.

Regarding the basic bent-cable as a regression function of x, we write $q_{\boldsymbol{\theta}}(x) = q(x; \boldsymbol{\theta})$. The proof of Lemma 4.1 relies on the following heuristic. For $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, the given (w_i, z_i)'s from $q_{\boldsymbol{\theta}_0}$ according to (4.3) to (4.5) cannot simultaneously satisfy the (i) convexity and (ii) smoothness and slope constraints that are associated with $q_{\boldsymbol{\theta}}$. The argument for (4.7) is similar. The mathematical details appear in Section 4.10.1 of the Chapter Appendix (Section 4.10).

Theorem 4.1 (Consistency). *Under conditions [A] and [B],* $\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}_0$ *as* $n \to \infty$.

Corollary 4.1. *Under conditions [A] and [B], there exists a decreasing sequence* $\xi_n \downarrow 0$ *such that* $P_{\boldsymbol{\theta}_0}\left\{\left|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right| \leq \xi_n\right\} \longrightarrow 1$ *as* $n \longrightarrow \infty$.

Corollary 4.1 is a technical consequence of Theorem 4.1. For example, take $\xi_n = m_n^{-1}$, where m_n is the largest integer such that $P_{\boldsymbol{\theta}_0}\left\{\left|\widehat{\boldsymbol{\theta}}_{n'} - \boldsymbol{\theta}_0\right| > m_n^{-1}\right\} \leq m_n^{-1} \ \forall \ n' \geq n$.

There are three main ideas in proving Theorem 4.1. (i) There is some large but finite number, M^*, such that as $n \longrightarrow \infty$, the probability for $\widehat{\boldsymbol{\theta}}_n$ to reside at a distance of at least M^* from $\boldsymbol{\theta}_0$ tends to 0. Thus, we can restrict our attention to the closed disc $\Theta^* \equiv \left\{\boldsymbol{\theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq M^*\right\}$. (ii) As $n \longrightarrow \infty$, the expected mean square function, H_n, does not become increasingly flat. (iii) As $n \longrightarrow \infty$, the mean square function, $\overline{S}_n$, converges uniformly on Θ^*, in probability, to its expectation, H_n.

Since H_n is minimized at $\boldsymbol{\theta}_0$ for all n, the idea in (ii) is crucial for showing that, for all sufficiently large n, every $\boldsymbol{\theta}$ which nearly minimizes H_n over Θ^* resides near $\boldsymbol{\theta}_0$. The idea in (iii) completes the logic of the proof by implying that, for all sufficiently large n, the minimizer of $\overline{S}_n$ (i.e. $\widehat{\boldsymbol{\theta}}_n$) is highly likely to nearly minimize H_n (and reside near $\boldsymbol{\theta}_0$).

The details for proving Theorem 4.1 are found in Section 4.7.

Before stating our next theorem, we provide a fact about the "square-root" matrix to avoid ambiguity.

For a symmetric positive definite matrix $\mathbb{A}$, there is a unique symmetric positive definite matrix $\mathbb{B}$ such that $\mathbb{A} = \mathbb{B}^2$ (see [31], Exercise P11.2.4). We call $\mathbb{B}$ the "square root" of $\mathbb{A}$ and write $\mathbb{B} = \mathbb{A}^{\frac{1}{2}}$.

Theorem 4.2 (Normal and χ^2-Limits). *Under conditions [A] to [D] and* $\gamma_0 > 0$,

(1.) $\frac{1}{n}\mathbb{I}_n(\boldsymbol{\theta}_0)$ *is positive definite for all sufficiently large n;*

(2.) $\sqrt{n}\left[\frac{1}{n}\mathbb{I}_n(\boldsymbol{\theta}_0)\right]^{\frac{1}{2}}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)$ *converges in distribution to a standard bivariate normal random variable;*

(3.) $P_{\boldsymbol{\theta}_0}\left\{\frac{1}{n}\mathbb{I}_n\left(\widehat{\boldsymbol{\theta}}_n\right) \text{is positive definite}\right\} \longrightarrow 1$;

(4.) $\sqrt{n}\left[\frac{1}{n}\mathbb{I}_n\left(\widehat{\boldsymbol{\theta}}_n\right)\right]^{\frac{1}{2}}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)$ *converges in distribution to a standard bivariate normal random variable;*

(5.) $G_n \equiv 2\left[\ell_n\left(\widehat{\boldsymbol{\theta}}_n\right) - \ell_n\left(\boldsymbol{\theta}_0\right)\right]$, *the deviance statistic, converges in distribution to a* χ^2 *random variable with 2 degrees of freedom.*

Assertions (1.) and (3.) of Theorem 4.2 permit an unambiguous definition of the square-root matrix in Assertions (2.) and (4.). In fact, we prove statement (4.26) on page 44, which is stronger than the statements of Assertions (1.) and (3.).

Assertions (2.) and (4.) are the main results of Theorem 4.2. In general, asymptotic normality of the MLE relies on the log-likelihood surface tending to a paraboloid over an ever-decreasing neighborhood of the true parameter value, with the surface's peak inside this neighborhood. We make use of the sequence ξ_n from Corollary 4.1 and examine the behavior of $\ell_n(\boldsymbol{\theta})$ over Θ_{ξ_n}, the ξ_n-neighborhood of $\boldsymbol{\theta}_0$.

The usual approach employs (i) the (asymptotic) normality of the score when evaluated at the true parameter value, and (ii) a one-term Taylor expansion of the score function evaluated at the MLE. This expansion is used to show that the log-likelihood is approximately paraboloidal, with the unique maximizer inside Θ_{ξ_n}. Then, normality of the score and Taylor expansion lead to approximate normality of the maximizer (i.e. the MLE).

The normality of the ε_i's implies

$$\boldsymbol{U}_n(\boldsymbol{\theta}_0) \sim \text{BVN}\left(\mathbf{0}, \mathbb{I}_n(\boldsymbol{\theta}_0)\right) \tag{4.8}$$

where "BVN" stands for a bivariate normal distribution. However, the tactic of (ii) cannot be directly employed in our problem, as the log-likelihood ℓ_n is only once-differentiable at some values of $\boldsymbol{\theta}$. Nevertheless, the spirit of the Taylor expansion is still apparent in our modified approach by using Lemma 4.2, which we state in the next section.

Asymptotic uniqueness (in probability) of $\widehat{\boldsymbol{\theta}}_n$ in Θ_{ξ_n} follows from virtual certainty (when n is large) for (a) $\widehat{\boldsymbol{\theta}}_n$ to exist in Θ_{ξ_n} (by Corollary 4.1), and (b) ℓ_n to be concave everywhere on Θ_{ξ_n}. To show concavity, we will prove that, for all large n, the Hessian is highly likely to be negative definite everywhere on Θ_{ξ_n}, except along some criss-crossing rays on the (τ, γ)-plane. This asymptotic negative definiteness, in probability, results from Assertion (1.) of the theorem and Lemma 4.3. (The lemma is given in the next section after Lemma 4.2. It states the component-wise uniform convergence of the negative Hessian to the Information matrix (at $\boldsymbol{\theta}_0$), after both matrices are divided by n.) Now, for each ray inside Θ_{ξ_n}, we will show that a reparametrized ℓ_n has a negative second derivative everywhere on this collapsed domain (i.e. the ray) except at isolated points. To complete the argument, we apply a result about the concavity of a function that is twice differentiable except at isolated points. The result is stated in Lemma 4.4 (p. 45) amidst the mathematical proof for Assertion (2.) in Section 4.7. This lemma implies that ℓ_n is also concave everywhere along any ray within Θ_{ξ_n}.

Normality of the score, consistency of the MLE, and Lemmas 4.2 to 4.4 thus provide the basis of asymptotic normality for our irregular estimation problem.

Finally, Assertion (5.) shows that the deviance statistic has the usual asymptotic χ^2 distribution. In a finite-sample practical setting, it is expected that the χ^2 approximation is better than normal approximations of Assertions (2.) and (4.). The standard approach to prove Assertion (5.) is to apply a two-term Taylor expansion to the log-likelihood function. Again, an undefined Hessian at some values in Ω forces us to apply our modified Taylor expansion in Lemma 4.2.

4.6 Auxiliary Lemmas for Proving Theorem 4.2

The first relevant lemma is a one-term expansion of the score function. It provides a well-defined "Taylor-type" expansion despite non-differentiability. Before we state the lemma, note that the ith summands of $U_{n\tau}(\boldsymbol{\theta})$ and $U_{n\gamma}(\boldsymbol{\theta})$ are

$$U_{\tau,i}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\left[\varepsilon_i + d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\right]\frac{\partial q_{\boldsymbol{\theta}}(x_i)}{\partial\tau} \quad , \quad U_{\gamma,i}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\left[\varepsilon_i + d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\right]\frac{\partial q_{\boldsymbol{\theta}}(x_i)}{\partial\gamma}$$

respectively, where

$$\begin{aligned}\frac{\partial q_{\boldsymbol{\theta}}(x_i)}{\partial\tau} &= -\frac{x_i-(\tau-\gamma)}{2\gamma}\,\mathbf{1}\big\{|x_i-\tau|\le\gamma\big\} - \mathbf{1}\{x_i>\tau+\gamma\}\,, \\ \frac{\partial q_{\boldsymbol{\theta}}(x_i)}{\partial\gamma} &= \frac{1}{4}\left[1-\left|\frac{x_i-\tau}{\gamma}\right|^2\right]\mathbf{1}\big\{|x_i-\tau|\le\gamma\big\}\,.\end{aligned}$$

Thus, for each i, both $U_{\tau,i}$ and $U_{\gamma,i}$ are continuous surfaces over the (τ,γ)-plane, but they have folds along the rays defined by

$$R_{i+} = \big\{\boldsymbol{\theta}\in\Omega : \gamma = \tau - x_i\big\}\,, \quad R_{i-} = \big\{\boldsymbol{\theta}\in\Omega : \gamma = x_i - \tau\big\}\,. \tag{4.9}$$

Summing these surfaces over i produces continuous score surfaces $U_{n\tau}$ and $U_{n\gamma}$, each with n pairs of folds indexed by the data $x_1,\ldots,x_n$ (Figure 4.1). As a result, the matrix $\mathbb{V}_n$, which is the Hessian for the log-likelihood, is well-defined everywhere on Ω except along the $R_{i\pm}$'s.

To avoid difficulties caused by an undefined $\mathbb{V}_n(\boldsymbol{\theta})$ value, we define a "directional" Hessian. That is, given any i and $\boldsymbol{\theta}\in\Omega$, we define directional derivatives for $U_{\tau,i}$ and $U_{\gamma,i}$ by taking

$$V^+_{\tau j,i}(\boldsymbol{\theta}) = \lim_{h\downarrow 0}\frac{\partial}{\partial\tau}U_{j,i}(\tau+h,\gamma) \quad , \quad V^+_{\gamma j,i}(\boldsymbol{\theta}) = \lim_{h\downarrow 0}\frac{\partial}{\partial\gamma}U_{j,i}(\tau,\gamma+h) \tag{4.10}$$

for all $i=1,\ldots,n$ and $j=\tau,\gamma$. Of course, these V^+_i's are merely regular derivatives when evaluated at some $\boldsymbol{\theta}\notin R_{k\pm}$, $k=1,\ldots,n$. Now, we can replace the Hessian matrix $\mathbb{V}_n$ by

$$\mathbb{V}^+_n(\boldsymbol{\theta}) = \sum_{i=1}^n \begin{bmatrix} V^+_{\tau\tau,i}(\boldsymbol{\theta}) & V^+_{\tau\gamma,i}(\boldsymbol{\theta}) \\ V^+_{\gamma\tau,i}(\boldsymbol{\theta}) & V^+_{\gamma\gamma,i}(\boldsymbol{\theta}) \end{bmatrix}$$

which is well-defined on Ω, and coincides with $\mathbb{V}_n(\boldsymbol{\theta})$ except along the $R_{k\pm}$'s.

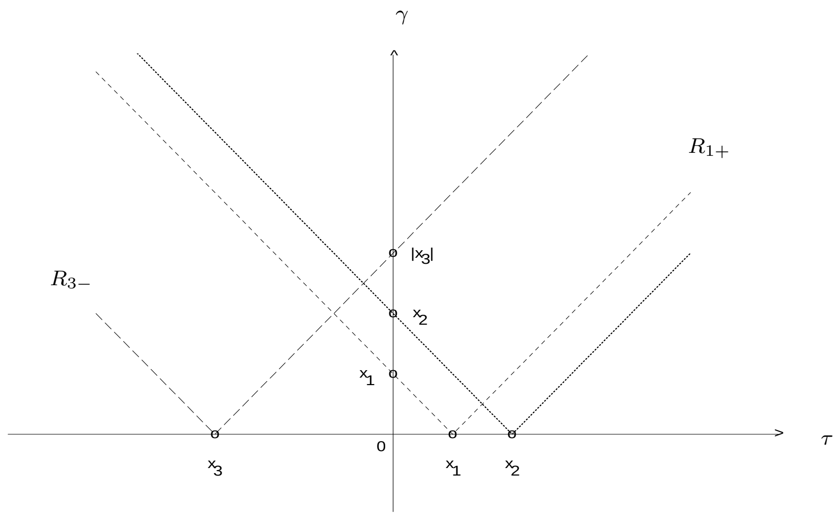

Figure 4.1: An example of the rays $R_{i\pm}$'s (on the (τ, γ)-plane) along which the score surfaces $U_{n\tau}$ and $U_{n\gamma}$ have folds. Each pair of $R_{i\pm}$ is indexed by a data point x_i. The "+" rays have slope 1 (i.e. 45 degrees), while the "−" rays have slope −1 (i.e. 135 degrees).

Lemma 4.2. *For all* $\boldsymbol{\theta} \in \Omega$*, we have*

$$\boldsymbol{U}_n(\boldsymbol{\theta}) = \boldsymbol{U}_n(\boldsymbol{\theta}_0) + \left[\int_0^1 \mathbb{V}_n^*(\boldsymbol{\theta}, t)\, dt\right]^T (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \tag{4.11}$$

where

$$\mathbb{V}_n^*(\boldsymbol{\theta}, t) = \sum_{i=1}^n \begin{bmatrix} V_{\tau\tau,i}^+\Big(\tau_0 + t(\tau - \tau_0), \quad \gamma_0 \Big) & V_{\tau\gamma,i}^+\Big(\tau_0 + t(\tau - \tau_0), \quad \gamma_0 \Big) \\ V_{\gamma\tau,i}^+\Big(\quad \tau \quad , \gamma_0 + t(\gamma - \gamma_0)\Big) & V_{\gamma\gamma,i}^+\Big(\quad \tau \quad , \gamma_0 + t(\gamma - \gamma_0)\Big) \end{bmatrix}$$

for all $t \in [0, 1]$ *and* $\boldsymbol{\theta} \in \Omega$.

Note that $\mathbb{V}_n^+$ is discontinuous. Thus, its integration to retrieve the score must be done with care. In particular, had the Hessian been continuous, the arguments for all four Hessian components in the integrand would have been simply $\big(\tau_0 + t(\tau - \tau_0), \gamma_0 + t(\gamma - \gamma_0)\big) = \boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$. Instead, we take slightly different arguments for the integrand, as given by $\mathbb{V}_n^*$.

The detailed proof for Lemma 4.2 is found in Section 4.10.1 of the Chapter Appendix.

Our next lemma states the uniform convergence, in probability, of the negative Hessian to the Information matrix.

Lemma 4.3. *Given are conditions [A] to [D], and a sequence* $\delta_n \downarrow 0$*. As* $n \longrightarrow \infty$*, the negative directional Hessian,* $-\mathbb{V}_n^+$*, converges component-wise (subject to proper scaling), in probability, to the Information matrix* $\mathbb{I}_n(\boldsymbol{\theta}_0)$ *uniformly over the disc* $\Theta_{\delta_n} = \{\boldsymbol{\theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \le \delta_n\}$*. More specifically, for all* $j, k = \tau, \gamma$*, we have*

$$\sup_{\boldsymbol{\theta} \in \Theta_{\delta_n}} \left| 1 + \frac{V_{n,jk}^+(\boldsymbol{\theta})}{I_{n,jk}(\boldsymbol{\theta}_0)} \right| \xrightarrow{\text{P}} 0 \qquad \text{as } n \longrightarrow \infty \tag{4.12}$$

where $V_{n,jk}^+$ *denotes the* (j,k)*-th component of* $\mathbb{V}_n^+$*, and similarly for* $I_{n,jk}$*.*

This lemma is central to the regularity of our regression problem. Its detailed proof appears in the Chapter Appendix in Section 4.10.1. There, inequalities (4.63) to (4.65) show that the Fisher Information, $\mathbb{I}_n(\boldsymbol{\theta}_0)$, is asymptotically well-behaved under conditions [A] and [B], in the sense that, for all sufficiently large n, it is guaranteed to contain information about the unknown parameter $\boldsymbol{\theta}_0$, and this information is no less than a non-trivial multiple of the sample size. As $-\mathbb{V}_n^+(\boldsymbol{\theta})$ is the negative (directional) Hessian of the log-likelihood function, its uniform convergence on Θ_{ξ_n} to a well-behaved Information matrix ensures that the log-likelihood surface itself is uniformly well-behaved on this ever-decreasing neighborhood of $\boldsymbol{\theta}_0$. This is the basis of the "regularity" of our problem, despite score surfaces that have folds.

4.7 Theorem Proofs

4.7.1 Proof of Theorem 4.1

To find M^* such that

$$P_{\boldsymbol{\theta}_0}\left\{\left|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right| > M^*\right\} = P_{\boldsymbol{\theta}_0}\left\{\widehat{\boldsymbol{\theta}}_n \notin \Theta^*\right\} \longrightarrow 0\,, \tag{4.13}$$

let us first find a lower τ-bound on Ω.

Consider a candidate cable $q_{\boldsymbol{\theta}}$ whose $\tau < \tau_0 - \gamma_0$. Thus, $q_{\boldsymbol{\theta}}$ lies entirely above $q_{\boldsymbol{\theta}_0}$, regardless of the size of γ. Consider $x \ge \tau_0 + \gamma_0$, at which point $q_{\boldsymbol{\theta}_0}$ is in its slope-one phase, while $q_{\boldsymbol{\theta}}$ can be in its quadratic or slope-one phase. In either case for $q_{\boldsymbol{\theta}}$, the distance between the two cables evaluated at this x is no less than the distance between the two slope-one phases. Therefore, for all $x \geq \tau_0 + \gamma_0$, we have $\left|d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x)\right| \geq \tau_0 - \tau$, which diverges to $+\infty$ as $\tau \longrightarrow -\infty$. This divergence is uniform as x varies over any compact set and as γ varies over $[0,\infty)$.

$$\tau < -M_1 \implies \left|d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x)\right|^2 \geq \frac{6\sigma^2}{c_+^*} \quad \forall\, x \geq \tau_0 + \gamma_0 \quad \forall\, \gamma \geq 0 \tag{4.14}$$

where c_+^* is from condition [B]. Now, take N such that $\sum \mathbf{1}\{x_i > \tau_0 + \gamma_0 + \delta_{11}\} \geq nc_+^*$ for all $n > N$. Then, for all $\gamma \geq 0$ and $n > N$, we have

$$\tau < -M_1 \Longrightarrow T_n(\boldsymbol{\theta}) \geq \frac{1}{n}\sum_i \left|d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\right|^2 \mathbf{1}\{x_i > \tau_0 + \gamma_0 + \delta_{11}\} \geq \frac{6\sigma^2}{c_+^*}\, c_+^* = 6\sigma^2. \tag{4.15}$$

That is, for $n > N$,

$$\inf_{\tau < -M_1} T_n(\boldsymbol{\theta}) \geq 6\sigma^2 \quad \forall\, \gamma \geq 0\,. \tag{4.16}$$

Next, note that

$$0 \leq \left[\frac{u}{\sqrt{2}} + \sqrt{2}v\right]^2 = \frac{u^2}{2} + 2uv + 2v^2 = (u+v)^2 - \left(\frac{u^2}{2} - v^2\right) \implies (u+v)^2 \geq \frac{u^2}{2} - v^2\,.$$

Take $u = d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)$ and $v = \varepsilon_i$. Then, apply (4.16) and the Strong Law of Large Numbers to see that, for $n > N$,

$$\begin{aligned} \forall\, \gamma \geq 0\,, \ \inf_{\tau < -M_1} \overline{S}_n(\boldsymbol{\theta}) - \overline{S}_n(\boldsymbol{\theta}_0) \ &\geq \ \inf_{\tau < -M_1} \left\{\frac{1}{2}T_n(\boldsymbol{\theta}) - \frac{1}{n}\sum_i \varepsilon_i^2\right\} - \overline{S}_n(\boldsymbol{\theta}_0) \\ &\geq \ 3\sigma^2 - 2\overline{S}_n(\boldsymbol{\theta}_0) \ \overset{\text{a.s.}}{\longrightarrow} \ \sigma^2 > 0\,. \end{aligned} \tag{4.17}$$

Hence,

$$\begin{aligned} &\forall\, \gamma \geq 0\,, \ \inf_{\tau < -M_1} \overline{S}_n(\boldsymbol{\theta}) \ \overset{\text{a.s.}}{>} \ \overline{S}_n(\boldsymbol{\theta}_0) \ \geq \ \overline{S}_n\big(\widehat{\boldsymbol{\theta}}_n\big) \quad \text{as } n \longrightarrow \infty\,, \\ \therefore \quad &P_{\boldsymbol{\theta}_0}\left\{\widehat{\boldsymbol{\theta}}_n \in [-M_1, M] \times [0,\infty)\right\} \longrightarrow 1\,. \end{aligned} \tag{4.18}$$

We can now restrict our attention to the reduced parameter space $\Omega_1 \equiv [-M_1, M] \times [0,\infty)$. To complete the argument for (i), we find an upper γ-bound on Ω_1.

Without loss of generality, assume $[\tau_0 - \gamma_0 - 1, \tau_0 + \gamma_0 + 1] \subset [-M_1, M]$. For τ confined in $[-M_1, M]$, take γ larger than the range of this τ-space, i.e. take $\gamma > M + M_1$. Thus, the bend of $q_{\boldsymbol{\theta}}(x)$ (as a function of x) is guaranteed to stretch over the entire $[-M_1, M]$ interval. In particular, this bend lies entirely above $q_{\boldsymbol{\theta}_0}(x)$ over $[\tau_0 - \gamma_0 - 1, \tau_0 + \gamma_0 + 1]$. Moreover, as $\gamma \longrightarrow \infty$, $\partial \left| d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x) \right| / \partial \gamma$ converges, uniformly for $x \in [\tau_0 - \gamma_0 - 1, \tau_0 + \gamma_0 + 1]$ and $\tau \in [-M_1, M]$, to $1/4$. Therefore, $\left| d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}} \right|$ diverges, as $\gamma \longrightarrow \infty$, to $+\infty$ uniformly on this x-range and τ-domain. Hence, there is a large but finite $M_2 > 0$ such that

$$\gamma > M_2 \implies \left| d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x) \right|^2 \geq \frac{6\sigma^2}{\min\{c_0(1), c_-^*\}} \quad \forall \quad x \in [\tau_0 - \gamma_0 - 1, \tau_0 + \gamma_0 + 1]$$
$$\forall \quad \tau \in [-M_1, M] \, .$$

To find a lower bound for the infimum of $T_n(\boldsymbol{\theta})$ over $\gamma > M_2$, we apply condition $[\mathrm{A}_1]$ to $\sum \mathbf{1}\{|x_i - \tau_0| \leq 1\}$, and $[\mathrm{A}_2]$ to $\sum \mathbf{1}\{|x_i - \tau_0| \leq \gamma_0 - \delta_1\}$. Then, we follow the logic of equations (4.15) through (4.18), and arrive at the conclusion that $P_{\boldsymbol{\theta}_0}\big\{\widehat{\boldsymbol{\theta}}_n \in [-M_1, M] \times [0, M_2]\big\} \longrightarrow 1$. Finally, take M^* such that $[-M_1, M] \times [0, M_2] \subset \Theta^* = \{\boldsymbol{\theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq M^*\}$. Then, $P_{\boldsymbol{\theta}_0}\big\{\widehat{\boldsymbol{\theta}}_n \notin \Theta^*\big\} \leq P_{\boldsymbol{\theta}_0}\big\{\widehat{\boldsymbol{\theta}}_n \notin [-M_1, M] \times [0, M_2]\big\} \longrightarrow 0$.

To show that H_n is not asymptotically flat, consider $\Theta_\delta \equiv \big\{\boldsymbol{\theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq \delta\big\}$, where $\delta \in (0, \delta_{11}/2)$. We will show that for a large n, the minimum of H_n over Θ^* cannot be achieved within the doughnut $D_\delta \equiv \text{closure}\{\Theta^* \setminus \Theta_\delta\} = \big\{\boldsymbol{\theta} : \delta \leq |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq M^*\big\}$. To this end, we show that there are N_1 and $\epsilon_1 > 0$ such that

$$n > N_1 \Longrightarrow \inf_{\boldsymbol{\theta} \, \in \, D_\delta} T_n(\boldsymbol{\theta}) \geq \epsilon_1 \, . \tag{4.19}$$

Continuity of $d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}$, compactness of D_δ, and Lemma 4.1 prove (4.19) as follows. First, note that $\tau - \gamma \leq (\tau_0 + M^*) + (\gamma_0 + M^*) = \tau_0 + \gamma_0 + 2M^*$ for all $\boldsymbol{\theta} \in D_\delta$. Pick $R \in (\tau_0 + \gamma_0 + 2M^*, \infty)$. If $\gamma_0 = 0$, also pick $\epsilon^* \in (0, \delta/4)$. Now, define

$$T^*(w_0, w_1, \boldsymbol{\theta}) = \left| d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(w_0) \right|^2 + \left| d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(w_1) \right|^2$$

over the compact domain,

$$C_{\delta,R} = \left\{ (w_0, w_1, \boldsymbol{\theta}) : \begin{array}{rcl} w_0 & \in & \begin{cases} [\tau_0 - \gamma_0 + \delta_1, \tau_0 + \gamma_0 - \delta_1] & \text{if} \quad \gamma_0 > 0 \\ [\tau_0 - \epsilon^*, \tau_0 + \epsilon^*] & \text{if} \quad \gamma_0 = 0 \end{cases} \\ w_1 & \in & [\tau_0 + \gamma_0 + \delta_{11}, R] \\ \boldsymbol{\theta} & \in & D_\delta \end{array} \right\}.$$

Also define $\eta = \inf_{C_{\delta,R}} T^*$, which is achieved within $C_{\delta,R}$, due to the continuity of T^* over this compact set. By Lemma 4.1, η is strictly positive.

Now, since both cables $q_{\boldsymbol{\theta}_0}$ and $q_{\boldsymbol{\theta}}$ must be in their slope-one phase beyond the value $\sup_{\boldsymbol{\theta} \in D_\delta} \{\tau + \gamma\} \le \tau_0 + \gamma_0 + 2M^*$, we have $d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(w_1) = \tau_0 - \tau$ for all $w_1 > R$. That is, for any given $\boldsymbol{\theta} \in D_\delta$ and w_0 satisfying (4.4), the function T^* is constant when w_1 is varied over (R, ∞). Hence,

$$\lim_{K \to \infty} \left\{ \inf_{C_{\delta,K}} T^* \right\} = \inf_{C_{\delta,R}} T^* = \eta > 0\,.$$

Next, re-label and group the data $\{x_1, \ldots, x_n\}$ into the following three sets:

$$\begin{array}{rcl} \mathcal{A} = \{x_{01}, x_{02}, \ldots, x_{0a}\} & = & \left\{ x_i : x_i \in \begin{array}{cl} [\tau_0 - \gamma_0 + \delta_1, \tau_0 + \gamma_0 - \delta_1] & \text{if} \quad \gamma > 0 \\ [\tau_0 - \epsilon^*, \tau_0 + \epsilon^*] & \text{if} \quad \gamma_0 = 0 \end{array} \right\} \\ \mathcal{B} = \{x_{11}, x_{12}, \ldots, x_{1b}\} & = & \big\{ x_i : x_i \in [\tau_0 + \gamma_0 + \delta_{11}, \infty) \big\} \\ \mathcal{C} = \{x_{21}, x_{22}, \ldots, x_{2c}\} & = & \big\{ x_i : x_i \notin \mathcal{A} \cup \mathcal{B} \big\} \end{array}$$

where the total cardinality is $a + b + c = n$. Now, by conditions [A] and [B], there is an N_1 such that

$$n > N_1 \Longrightarrow \begin{cases} a \ \ge \ \begin{cases} n\, c_-^* & \text{if} \quad \gamma_0 > 0 \\ n\, c_0(\epsilon^*) & \text{if} \quad \gamma_0 = 0 \end{cases} \\ b \ \ge \ n\, c_+^* \end{cases}.$$

Define $m = \min\{a, b\}$ and $c^* = \min\{\, c_-^*\,,\, c_0(\epsilon^*)\,,\, c_+^*\,\}$, and so $(m/n) \ge c^*$. Thus, for $n > N_1$ and $\boldsymbol{\theta} \in D_\delta$, we have

$$
\begin{aligned}
T_n(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n}\left|d_{\boldsymbol{\theta}_0,\,\boldsymbol{\theta}}(x_i)\right|^2 \\
&\geq \frac{1}{n}\sum_{j=1}^{m} T^*(x_{0j}, x_{1j}, \boldsymbol{\theta}) \\
&\geq \frac{1}{n}\sum_{j=1}^{m} \lim_{K\to\infty}\left\{\inf_{C_{\delta,K}} T^*\right\} \;=\; \eta\left(\frac{m}{n}\right) \;\geq\; \eta\, c^*\,.
\end{aligned}
$$

Now, take $\epsilon_1 = \eta\, c^*$ which is positive and $\boldsymbol{\theta}$-free. Then, for the given doughnut D_δ (and an arbitrary ϵ^* in the case of $\gamma_0 = 0$), (4.19) is true.

Lastly, we will show that the mean square function, $\overline{S}_n$, converges, in probability, to its expectation, H_n, uniformly on Θ^*. That is,

$$
P_{\boldsymbol{\theta}_0}\left\{\sup_{\boldsymbol{\theta}\,\in\,\Theta^*}\left|\overline{S}_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta})\right| \leq \epsilon\right\} \longrightarrow 1 \;\;\forall\; \epsilon > 0\,. \tag{4.20}
$$

To see this, first it can be shown that for all $x \in \mathbb{R}$ and $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Omega$,

$$
\left|d_{\boldsymbol{\theta}_1,\,\boldsymbol{\theta}_2}(x)\right| \leq \max\left\{|\tau_1 - \tau_2|, |\gamma_1 - \gamma_2|\right\}. \tag{4.21}
$$

Next, define $\mu^* = \sigma\sqrt{2/\pi}$, the mean of a $\left|N(0,\sigma^2)\right|$ random variable. Since Θ^* is compact, there is, for every $\epsilon > 0$, a finite collection of balls B_k, $k = 1, \ldots, R < \infty$, each centered at some $\boldsymbol{\theta}_k \in \Theta^*$, and with radius $\delta'/2$, where $0 < \delta'(2M^* + \mu^*) < \epsilon$, such that $\Theta^* \subset \bigcup_k B_k$. Then, for all $i = 1, \ldots, n$, we have

$$
\left|d_{\boldsymbol{\theta}_0,\,\boldsymbol{\theta}}(x_i)\right| \leq M^* \quad \forall \quad \boldsymbol{\theta} \in \Theta^*\,, \quad \text{and} \tag{4.22}
$$

$$
\left|d_{\boldsymbol{\theta}_k,\,\boldsymbol{\theta}}(x_i)\right| \leq \delta'/2 \quad \forall \quad \boldsymbol{\theta} \in B_k \;\forall\; k = 1, \ldots, R \tag{4.23}
$$

Apply the Weak Law of Large Numbers, Chebyshev's inequality, and (4.22) to obtain convergence, in probability, of $\overline{S}_n(\boldsymbol{\theta}_k)$ to $H_n(\boldsymbol{\theta}_k)$ for each k. Next, apply (4.22) and (4.23) for $\boldsymbol{\theta} \in B_k \bigcap \Theta^*$ for each k to get

$$
\begin{aligned}
\left|\overline{S}_n(\boldsymbol{\theta}) - \overline{S}_n(\boldsymbol{\theta}_k)\right| &= \left|\frac{1}{n}\sum 2\varepsilon_i\left\{d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i) - d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_k}(x_i)\right\} + H_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta}_k)\right| \\
&\le \frac{2}{n}\sum\left|\varepsilon_i\right|\left|d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)\right| + \left|H_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta}_k)\right| \\
&\le 2\left(\frac{1}{n}\sum|\varepsilon_i|\right)\frac{\delta'}{2} + \left|H_n(\boldsymbol{\theta}_k) - H_n(\boldsymbol{\theta})\right|, \qquad (4.24) \\
\left|H_n(\boldsymbol{\theta}_k) - H_n(\boldsymbol{\theta})\right| &= \left|T_n(\boldsymbol{\theta}_k) - T_n(\boldsymbol{\theta})\right| \\
&= \frac{1}{n}\left|\sum\left\{d^2_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_k}(x_i) - d^2_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\right\}\right| \\
&= \frac{1}{n}\left|\sum\left\{d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_k}(x_i) + d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\right\} d_{\boldsymbol{\theta},\boldsymbol{\theta}_k}(x_i)\right| \\
&\le \frac{1}{n}\sum\left(\left|d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_k}(x_i)\right| + \left|d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\right|\right)\left|d_{\boldsymbol{\theta},\boldsymbol{\theta}_k}(x_i)\right| \;\le\; M^*\delta'. \qquad (4.25)
\end{aligned}
$$

Substitute this $\boldsymbol{\theta}$- and k-free bound into (4.24), and (4.20) follows from the triangle inequality and the Weak Law of Large Numbers. ■

4.7.2 Proof of Theorem 4.2

Assertion (1.)

Denote the eigenvalues of the covariance matrix $n^{-1}\,\mathbb{I}_n(\boldsymbol{\theta}_0)$ by λ_{n1} and λ_{n2}, where $0 \le \lambda_{n1} \le \lambda_{n2}$. We show

$$\exists \text{ integer } N \text{ and } \epsilon > 0 \quad \text{s.t.} \qquad n > N \implies \lambda_{n1} \ge \epsilon \qquad (4.26)$$

(so that the matrix $n^{-1}\,\mathbb{I}_n(\boldsymbol{\theta}_0)$ is positive definite for all $n > N$). Note the relations

$$\lambda_{n1} + \lambda_{n2} = \text{trace}\left(\frac{1}{n}\mathbb{I}_n(\boldsymbol{\theta}_0)\right) \;,\; \lambda_{n1}\lambda_{n2} = \det\left(\frac{1}{n}\mathbb{I}_n(\boldsymbol{\theta}_0)\right)$$

where “det” stands for “determinant.” It can be easily verified that

$$0 \le \lambda_{n1} + \lambda_{n2} \le \frac{1}{\sigma^2} + \frac{1}{16\sigma^2} = \frac{17}{16\sigma^2}\,, \qquad (4.27)$$

and, by conditions [A], [B], and the Cauchy-Schwarz inequality, that

$$\lambda_{n1}\lambda_{n2} \ge \frac{c_+^* c_-^*}{16\sigma^4}\left[1 - \left(\frac{\gamma_0 - \delta_1}{\gamma_0}\right)^2\right]^2 \quad (> 0) \qquad (4.28)$$

for all $n > N$ for some N.

Note that the set $\{(\lambda_{n1}, \lambda_{n2}) : \lambda_{n1} \le \lambda_{n2}, (4.27) \text{ and } (4.28) \text{ satisfied } \forall\, n > N\}$ is compact. Hence, the infimum of λ_{n1} is achieved over this set, and is positive by (4.28). Thus, we have proved Assertion (1.) of Theorem 4.2.

Assertion (2.)

We begin by showing that there exists an integer N such that for $n > N$, the surface ℓ_n is probably strictly concave over Θ_{ξ_n}, so that there is no $\boldsymbol{\theta} \in \Theta_{\xi_n}$ other than $\widehat{\boldsymbol{\theta}}_n$ (as defined at the beginning of Section 4.3) at which ℓ_n is maximized.

As in the proof of Assertion (1.), we consider the eigenvalues of the random matrix $-n^{-1}\,\mathbb{V}_n^+$. As eigenvalues depend continuously on the matrix entries, Lemma 4.3 implies that the smaller eigenvalue of $-n^{-1}\,\mathbb{V}_n^+$ converges uniformly over Θ_{ξ_n}, in probability, to $\lambda_{n1} > 0$. Thus,

$$P_{\boldsymbol{\theta}_0}\left\{\frac{1}{n}\mathbb{V}_n^+(\boldsymbol{\theta}) \text{ is negative definite for all } \boldsymbol{\theta} \in \Theta_{\xi_n}\right\} \longrightarrow 1\,. \tag{4.29}$$

Now, we introduce a lemma about the concavity of a once-differentiable function. Its first assertion is due to Theorem 4.4.10 in [32], and the second, the definition of concavity. We omit the proof.

Lemma 4.4 (Concavity). *(1) Let $\mathcal{N}$ be a subset of [0,1] consisting of isolated points. Suppose that a differentiable function $f : [0,1] \longrightarrow \mathbb{R}$ has a continuous second derivative, f'', in $[0,1] \setminus \mathcal{N}$. Then, f is strictly concave on $[0,1]$ if $f''(t) < 0$ for all but those isolated values of $t \in \mathcal{N}$. (2) A differentiable function $g : \Theta_{\xi_n} \longrightarrow \mathbb{R}$ is strictly concave if and only if $h(t) \equiv g\big(\boldsymbol{\theta}_1 + t(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)\big)$ is concave on $[0,1]$ for each $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta_{\xi_n}$, where $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$.*

For our problem, we study

$$h_n(t) \equiv \ell_n\big(\boldsymbol{\theta}_1 + t(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)\big)\,, \qquad \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta_{\xi_n}\,,\ \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1\,,\ t \in [0,1]\,.$$

In what follows, we restrict our attention to the event defined in (4.29).

First, consider $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ which do not both lie on a ray $R_{k\pm}$ for any k. They define a line segment along which ℓ_n is piecewise twice differentiable, with discontinuities in the second partial derivatives at the isolated intersections of the line segment with the rays $R_{k\pm}$'s. By the chain rule, we have

$$\frac{1}{n}h_n''(t) = (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)^T \left[\frac{1}{n}\mathbb{V}_n\big(\boldsymbol{\theta}_1 + t(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)\big)\right] (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1) \tag{4.30}$$

$$= (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)^T \left[\frac{1}{n}\mathbb{V}_n^+\big(\boldsymbol{\theta}_1 + t(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)\big)\right] (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1) < 0 \tag{4.31}$$

for all but isolated values of $t \in [0,1]$. By Lemma 4.4, it follows that $n^{-1}\,\ell_n$ — hence, ℓ_n — is concave along this line segment.

We are left to examine the case where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ both lie on a ray. By symmetry, we need only to consider, say, the 45-degree R_{k+}'s. That is, we consider $\boldsymbol{\theta}_j = (\tau_j, \tau_j - x_k)$ for $j = 1, 2$. Here, the chain rule cannot be applied to yield relation (4.30), since $\mathbb{V}_n$ is not defined at $\boldsymbol{\theta}_1 + t(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)$ for any $t \in [0,1]$. Instead, one can expand $h_n(t)$, differentiate it twice, and verify, after some lengthy algebra, that $n^{-1}h_n(t)$ is indeed equal to the expression in (4.31) for all but isolated values of $t \in [0,1]$. These isolated values correspond to the intersections of the line segment (joining $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ on R_{k+}) with the 135-degree R_{i-}'s.

Thus, we have shown that every cross-section, and hence, the entire surface, of ℓ_n over Θ_{ξ_n} is strictly concave on the event defined in (4.29). By consistency,

$$P_{\boldsymbol{\theta}_0}\left\{\widehat{\boldsymbol{\theta}}_n \text{ is the unique maximizer of } \ell_n \text{ over } \Theta_{\xi_n}\right\} \longrightarrow 1\,. \tag{4.32}$$

Having argued for uniqueness of $\widehat{\boldsymbol{\theta}}_n$, we are ready to apply the Taylor-expansion-type relation of Lemma 4.2 to conclude asymptotic normality. First, define

$$\mathbb{D}_n(\boldsymbol{\theta}, t) = \mathbb{V}_n^*(\boldsymbol{\theta}, t) + \mathbb{I}_n(\boldsymbol{\theta}_0) \qquad (\boldsymbol{\theta} \in \Theta_{\xi_n},\ t \in [0,1])$$

$$\mathbb{M}_n(\boldsymbol{\theta}) = \left[\int_0^1 \mathbb{D}_n(\boldsymbol{\theta}, t)\, dt\right]^T \qquad (\boldsymbol{\theta} \in \Theta_{\xi_n})$$

By Lemma 4.2, we have

$$\boldsymbol{U}_n(\boldsymbol{\theta}) = \boldsymbol{U}_n(\boldsymbol{\theta}_0) + \frac{1}{n}\Big[\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{I}_n(\boldsymbol{\theta}_0)\Big]\, n(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\,. \tag{4.33}$$

Apply Lemma 4.3, Assertion (1.) of the theorem, (4.32), and (4.8), and the subsequent algebra is standard for concluding asymptotic normality of Assertion (2.).

Assertion (3.)

On page 63 of the Chapter Appendix, we prove relation (4.67), which asserts that $\left|I_{n,jk}(\boldsymbol{\theta}) - I_{n,jk}(\boldsymbol{\theta}_0)\right|$ has order $o(n)$ uniformly for $\boldsymbol{\theta} \in \Theta_{\delta_n}$ for any $\delta_n \downarrow 0$. Together with the consistency of $\widehat{\boldsymbol{\theta}}_n$, we have

$$\frac{1}{n} I_{n,jk}(\widehat{\boldsymbol{\theta}}_n) \;=\; \frac{1}{n} I_{n,jk}(\boldsymbol{\theta}_0) \pm o_p(1)\,.$$

Hence, the result of Assertion (3.) follows directly from Assertion (1.) of the theorem.

Assertion (4.)

This assertion follows directly from a simple application of Slutsky's Theorem to the algebra in the proof of Assertion (2.).

Assertion (5.)

Since ℓ_n is once-differentiable everywhere on Ω, we can apply the standard one-term Taylor expansion to get

$$\ell_n(\boldsymbol{\theta}) = \ell_n(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \int_0^1 \boldsymbol{U}_n\big(\boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\big)\, dt\,. \tag{4.34}$$

By Lemma 4.2,

$$\boldsymbol{U}_n\big(\boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\big) = \boldsymbol{U}_n(\boldsymbol{\theta}_0) + \left[\int_0^1 \mathbb{V}_n^*\big(\boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0), s\big)\, ds\right]^T t(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \tag{4.35}$$

for all $t \in [0,1]$, where

$$\mathbb{V}_n^*\big(\boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0), s\big) =$$
$$\sum_i \begin{bmatrix} V^+_{\tau\tau,i}\big(\tau_0 + st(\tau - \tau_0), \gamma_0\big) & V^+_{\tau\gamma,i}\big(\tau_0 + st(\tau - \tau_0), \gamma_0\big) \\ V^+_{\gamma\tau,i}\big(\tau_0 + t(\tau - \tau_0), \gamma_0 + st(\gamma - \gamma_0)\big) & V^+_{\gamma\gamma,i}\big(\tau_0 + t(\tau - \tau_0), \gamma_0 + st(\gamma - \gamma_0)\big) \end{bmatrix}.$$

In what follows, we denote by $\mathbb{Z}_n$ a 2×2 random matrix whose (a) value changes from use to use, and (b) elements may depend on any or all of $s, t \in [0,1]$ and $\boldsymbol{\theta} \in \Theta_{\xi_n}$, but are each of order $o_p(n)$ uniformly over these domains of s, t, and $\boldsymbol{\theta}$.

By Lemma 4.3, there is an integer N such that $n > N$ yields $\mathbb{V}_n^*\big(\boldsymbol{\theta}_0+t(\boldsymbol{\theta}-\boldsymbol{\theta}_0), s\big) = -\mathbb{I}_n(\boldsymbol{\theta}_0) + \mathbb{Z}_n$ uniformly over the set $\{(\boldsymbol{\theta}, s, t) : \boldsymbol{\theta} \in \Theta_{\xi_n},\ s, t \in [0,1]\}$. Hence,

$$\begin{aligned}\int_0^1 \boldsymbol{U}_n\big(\boldsymbol{\theta}_0 + t(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\big)\, dt &= \int_0^1 \Big\{\boldsymbol{U}_n(\boldsymbol{\theta}_0) - t\big[\mathbb{I}_n(\boldsymbol{\theta}_0) + \mathbb{Z}_n\big]\big(\boldsymbol{\theta}-\boldsymbol{\theta}_0\big)\Big\}\, dt \\ &= \boldsymbol{U}_n(\boldsymbol{\theta}_0) - \frac{1}{2}\boldsymbol{U}_n^\star(\boldsymbol{\theta}) \qquad (4.36)\end{aligned}$$

where $$\boldsymbol{U}_n^\star(\boldsymbol{\theta}) = \big[\mathbb{I}_n(\boldsymbol{\theta}_0) + \mathbb{Z}_n\big](\boldsymbol{\theta}-\boldsymbol{\theta}_0) = \frac{1}{n}\big[\mathbb{I}_n(\boldsymbol{\theta}_0) + \mathbb{Z}_n\big]\, n(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\,.$$

Apply Lemma 4.3, Assertion (1.) of the theorem, (4.32), and Theorem 4.1 to verify that

$$\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big)^T\Big[\boldsymbol{U}_n\big(\boldsymbol{\theta}_0\big) - \boldsymbol{U}_n^\star\big(\widehat{\boldsymbol{\theta}}_n\big)\Big] \;=\; o_p(1)\,. \qquad (4.37)$$

Finally, apply Assertion (2.) of the theorem and expression (4.8) to see that

$$\begin{aligned}G_n &= \big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big)^T\boldsymbol{U}_n(\boldsymbol{\theta}_0) + \big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big)^T\Big[\boldsymbol{U}_n(\boldsymbol{\theta}_0) - \boldsymbol{U}_n^\star\big(\widehat{\boldsymbol{\theta}}_n\big)\Big] \qquad (4.38)\\ &= \sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big)^T\left[\frac{1}{n}\mathbb{I}_n(\boldsymbol{\theta}_0)\right]^{\frac{1}{2}}\left[\frac{1}{n}\mathbb{I}_n(\boldsymbol{\theta}_0)\right]^{-\frac{1}{2}}\frac{1}{\sqrt{n}}\boldsymbol{U}_n(\boldsymbol{\theta}_0) + o_p(1) \;\xrightarrow{\mathcal{L}}\; \chi_2^2\,. \qquad (4.39)\end{aligned}$$

This concludes the proof of Theorem 4.2. ∎

4.8 A Kinked Truth

So far, we have shown the asymptotic properties of the MLE, $\widehat{\boldsymbol{\theta}}_n = (\widehat{\tau}_n, \widehat{\gamma}_n)^T$, for a basic bent cable whose quadratic bend has a non-trivial half-width γ_0. In this section, we shed some light on the distributional theory for the remaining case of a missing bend (i.e. when a bent cable is fitted to an underlying broken stick). We discuss only the one-parameter model, where τ_0, the bend location, is known. Again, we assume normally distributed random errors with a known, constant variance.

Despite the apparent simplicity of this model, the associated maximum likelihood estimation problem is far from regular. We will see that the asymptotic distribution for the MLE of γ_0, when in fact $\gamma_0 = 0$, is very peculiar, with a slow convergence rate (see Sections 4.8.2 and 4.8.5). These negative results suggest that no practical benefits can be gained by pursuing the case of $\gamma_0 = 0$ in the more general full bent cable of Chapter 5.

4.8.1 The One-Parameter Basic Bent Cable

Suppose the location of the bend, τ_0, is known. Without loss of generality, take $\tau_0 = 0$. The underlying basic bent cable is then

$$q(x; \gamma_0) = \frac{(x + \gamma_0)^2}{4\gamma_0} \mathbf{1}\big\{|x| \le \gamma_0\big\} + x\,\mathbf{1}\big\{x > \gamma_0\big\} \tag{4.40}$$

on the regression domain $\mathcal{X} = \mathbb{R}$ and the parameter space $\Omega = [0, \infty)$. Again, we consider maximum likelihood estimation of γ_0 from a dataset $\{(x_i, Y_i)\}_{i=1}^n$ generated by (4.2).

The "kinked truth" arises when the unknown γ_0 is 0, and is a boundary problem — 0 being a boundary value of Ω. This is also an "unidentified model" in the sense of [21], in which a segment of the regression model is missing. Feder has shown in [21], through an example, that the limiting distribution of the MLE for the unidentified model is not normal. However, this sense of "unidentifiability" is somewhat different from that of Lemma 4.1. In particular, all the join points of Feder's segmented model are regression parameters, so that a missing segment corresponds to coinciding joints, and the estimation problem becomes overspecified. On the contrary, parameter estimation for the bent cable with no bend is a well-defined, albeit irregular, problem.

Note also that since $\gamma_0 = 0$ is on the boundary, the cube root asymptotics of [33] are not applicable.

4.8.2 Results for $\gamma_0 = 0$

Under the assumption of normally distributed errors with a known variance, and some conditions on the design (see Section 4.8.4), we examine the long-run behavior of the error sum-of-squares function $S_n(\gamma)$, from which the asymptotic properties of its (global) minimizer, $\widehat{\gamma}_n$ (the MLE), are deduced.

Given any $M > 0$, denote the minimizer of S_n locally on $[0, Mn^{-1/3}]$ by $\widehat{\gamma}^*_{n,M}$. Hence, $\widehat{\gamma}^*_{n,M} \equiv \widehat{\gamma}_n$ if S_n is indeed globally minimized somewhere on $[0, Mn^{-1/3}]$. We show that the limiting distribution of $n^{1/3}\big(\widehat{\gamma}^*_{n,M} - \gamma_0\big) = n^{1/3}\,\widehat{\gamma}^*_{n,M}$ is the distribution of T_M, the minimizer over $[0, M]$ of $W_t \equiv \mu_t - 2\mathcal{G}_t$. Here, $\mathcal{G}_t$ is a mean-zero Gaussian process, while μ_t is a drift function proportional to t^3. Compared to μ_t, the process $\mathcal{G}_t$ is asymptotically negligible in probability, as $t \longrightarrow \infty$. This ensures that for any $\epsilon > 0$, there is a κ such that the probability is at least $1 - \epsilon$ for T_∞ (the global minimizer on $[0, \infty)$ of W_t) to reside within $[0, \kappa]$. In turn, $\lim_{M\to\infty}\left\{\lim_{n\to\infty}\ \mathcal{L}\left(n^{1/3}\,\widehat{\gamma}^*_{n,M}\right)\right\} = \mathcal{L}\big(T_\infty\big)$, where $\mathcal{L}$ denotes the distributional law. Such a limit for the estimation error is exceptional, and so is the convergence rate of $n^{-1/3}$, which is substantially slower than $n^{-1/2}$ — the rate in a regular problem, as well as in the case of $\gamma_0 > 0$. On the other hand, if $\widehat{\gamma}_n \in (Mn^{-1/3}, \infty)$, then its convergence rate must be no better than $n^{-1/3}$. Finally, as $n \longrightarrow \infty$, then $M \longrightarrow \infty$, the likelihood ratio test (deviance) statistic, computed based on $\widehat{\gamma}^*_{n,M}$, converges in distribution to $-W_{T_\infty}/\sigma^2$.

These results are formally restated with proofs in Section 4.8.5.[1]

[1] We have examined conditions under which $\widehat{\gamma}^*_{n,M} = \widehat{\gamma}_n$ with high probability, and they are complex. We omit them here, as they are not needed to establish the poor performance of $\widehat{\gamma}_n$.

Implications

Due to the peculiar rates and the form of the limit associated with the best rate, formal inference based on maximum likelihood estimation appears to be impractical in finite-sample settings, whether the bent-cable model has one or more unknown parameters. Indeed, we have already shed some light on this impracticality in Chapter 2 through analyzing observed datasets with the full bent-cable model.

The common F-test for $H_0 : \gamma_0 = 0$ is also highly questionable. For σ known, our results suggest that there do not exist reasonable regularity conditions under which the likelihood ratio statistic is asymptotically χ^2 distributed under H_0. Yet, some authors who applied bent-cable regression have included formal hypothesis testing of this null based on the F distribution without formal justification. For example, see [10], p. 478.

4.8.3 Notation for Estimation of the Kinked Truth

For an underlying bent cable with no bend (i.e. $\gamma_0 = 0$) and break point at $\tau_0 = 0$, we have the following definitions based on Section 4.3:

$$d_{\gamma,0}(x) = \frac{(\gamma - |x|)^2}{4\gamma}\,\mathbf{1}\{|x| \le \gamma\} \quad , \quad S_n(\gamma) = \sum_1^n \Big|\varepsilon_i - d_{\gamma,0}(x_i)\Big|^2$$

$$\ell_n(\gamma) = -\frac{1}{2\sigma^2} S_n(\gamma) + \text{constant} \quad , \quad I_n(\gamma) = \frac{1}{16\sigma^2}\sum_1^n \left[1 - \frac{x_i^2}{\gamma^2}\right]^2 \mathbf{1}\{|x_i| \le \gamma\}\,.$$

Note that for a given x, the distance function $d_{\gamma,0}(x)$ is non-decreasing in $\gamma \in \Omega$.

Recall from Section 4.3 that in the case of multiple minimizers of S_n, we take $\widehat{\gamma}_n$ to be the least of all such values. Here, we also do so for $\widehat{\gamma}^*_{n,M}$.

Our goal is to study the large-n behavior of S_n to deduce the asymptotic properties of its minimizer, $\widehat{\gamma}^*_{n,M}$, over $[0, Mn^{-1/3}]$. For this, we reparametrize $S_n(\gamma)$ by defining the translated sum-of-squares process as $S_{tn} = S_n(tn^{-1/3}) - S_n(0)$. We will then see that S_{tn} is a linear combination of a deterministic mean process, g_{tn}, and a stochastic deviation process, R_{tn}, that can be decomposed into processes $A_{tn,p}$, $p = 0, 1, 2$. Here, the $A_{tn,p}$'s are mean-zero continuous time Gaussian Markov processes. We introduce this decomposition through the following definitions:

$$
\begin{aligned}
a_i(u,p) &= |x_i|^p \,\mathbf{1}\big\{|x_i| \le u\big\} \\
b_i(u) &= d_{u,0}(x_i) \\
A_{tn,p} &= n^{\frac{p-1}{3}} \sum_1^n \varepsilon_i \; a_i\big(tn^{-1/3}, p\big) \\
R_{tn} &= \sum_1^n \varepsilon_i \, b_i\big(tn^{-1/3}\big) \;=\; \frac{t}{4} A_{tn,0} - \frac{1}{2} A_{tn,1} + \frac{1}{4t} A_{tn,2} \\
g_{tn} &= \sum_1^n \Big| b_i\big(tn^{-1/3}\big)\Big|^2 \\
S_{tn} &= S_n\big(tn^{-1/3}\big) - S_n(0) \;=\; g_{tn} - 2R_{tn}
\end{aligned}
\tag{4.41}
$$

4.8.4 The Design

Throughout the remainder of Section 4.8, we assume the following design conditions in the estimation of $\gamma_0 = 0$ in model (4.40):

[a] For all $\delta > 0$, $\liminf_{n\to\infty} \; (1/n)\sum_1^n \mathbf{1}\big\{|x_i| \le \delta\big\} > 0$.

[b] For all $t \ge 0$ and $p = 0, 1, 2, 3, 4$, $\lim_{n\to\infty} \; n^{\frac{p-2}{3}} \sum_1^n a_i\big(tn^{-1/3}, p\big) = \; c_p \, t^{p+1}$, where c_p is some positive constant (independent of t or n).

Condition [a] ensures a non-trivial fraction of data exists in any neighborhood around the kink, leading to a consistent $\widehat{\gamma}_n$. Indeed, condition [a] is equivalent to [A_1] of Section 4.4. On the other hand, we replace conditions [B] to [D] from the earlier section by the new condition [b] here. To motivate it, first consider the conditions needed for S_{tn}, the translated sum-of-squares process, to converge. Through the decomposition of S_{tn} as defined above, we will see that the convergence of $\sum a_i$, subject to proper normalization, is needed for the covariance structure of S_{tn} to converge. Now, condition [b] holds automatically in the case of randomly generated x_i's from a continuous distribution, whose density function f is such that $f(0) > 0$. In this case, one can show, for each i, that the (random) quantity $n^{\{(p-2)/3\}} \, a_i\big(tn^{-1/3}, p\big)$ has mean $[2\, f(0)/n(p+1)] \, t^{p+1} + r_{n,i}$ (where $r_{n,i}$ is dominated by the preceding term), and has variance of order $O\big(n^{-5/3}\big)$. Therefore, summing over i in this instance gives an asymptotic mean of $c_p\, t^{p+1}$, where $c_p = 2\, f(0)/(p+1)$, and a standard error of order $O\big(n^{-1/3}\big) \longrightarrow 0$.

Perhaps the most mysterious part of [b] is the division by $n^{1/3}$, which we will prove to be the best convergence rate of $\widehat{\gamma}_n$. Heuristically, this rate is due to the size of the Fisher Information as a function of γ, which is $I_n(\gamma) = O\big(\sum \mathbf{1}\{|x_i| \le \gamma\}\big)$. Consider spreading $x_1, \ldots, x_n$ uniformly over $[-1/2, 1/2]$, so that for all $\gamma \in [0, 1/2]$, $I_n(\gamma)$ is proportional to $n\gamma$. One would then expect $\widehat{\gamma}_n$ to be unlikely to exceed a value that is proportional to $\big[I_n(\widehat{\gamma}_n)\big]^{-1/2} \overset{\mathrm{P}}{\propto} (n\widehat{\gamma}_n)^{-1/2}$. The worst-case scenario for such $\widehat{\gamma}_n$ is when it takes on its largest "likely value," i.e. $\widehat{\gamma}_n \overset{\mathrm{P}}{\propto} (n\widehat{\gamma}_n)^{-1/2}$, or, $\widehat{\gamma}_n \overset{\mathrm{P}}{\propto} n^{-1/3}$.

However, convergence of S_{tn} does not require the randomness of the x_i's, so long as they satisfy condition [b]. The condition captures the properties of randomly generated x_i's according to any density as described, while it conveniently allows the data to be placed in a deterministic fashion without specifying the particular density.

4.8.5 Formal Results for Kinked Truth Estimation

We formally present the results in a lemma and a theorem.

Lemma 4.5. *Let c_p's be as given in condition [b]. Define $\mu_t = \big\{c_0 - 4c_1 + 6c_2 - 4c_3 + c_4\big\} \times t^3/16$ for all $t \ge 0$, and let $\mathcal{G}_t$ be a mean-zero Gaussian process with covariance structure*

$$
\begin{aligned}
Cov\big[\mathcal{G}_s - \mathcal{G}_r, \mathcal{G}_r\big] &= \Big\{(c_0 - 2c_1)s + c_2(s-r) + 2(c_3 - c_4)r\Big\}\frac{(s-r)r^2\sigma^2}{16s} && (4.42)\\
Cov\big[\mathcal{G}_r, \mathcal{G}_s\big] &= \Big\{(c_0 + c_2 - 2c_1)s^2 + 2(2c_2 - c_1 - c_3)rs \\
&\qquad + (c_2 + c_4 - 2c_3)r^2\Big\}\frac{r^2\sigma^2}{16s} && (4.43)
\end{aligned}
$$

for all $s \ge r \ge 0$. Let $n \longrightarrow \infty$. Under condition [b], the covariance structure of R_{tn} (and hence, of $-S_{tn}/2$) converges to that of $\mathcal{G}_t$, and g_{tn} $\big(= E_0[S_{tn}]\big)$ converges uniformly to μ_t on any closed interval of t.

One can verify that the constant $c_0 - 4c_1 + 6c_2 - 4c_3 + c_4$ in the expression of μ_t is strictly positive under conditions [a] and [b].

As a consequence of this lemma and relation (4.41), the expectation and covariance structure of S_{tn} converge respectively to those of $W_t \equiv \mu_t - 2\mathcal{G}_t$. For $r = s = t$, relation (4.43) yields a variance which is proportional to t^3. Hence, S_{tn} is proportional in probability to t^3 as $n, t \longrightarrow \infty$.

The proof of Lemma 4.5 appears in Section 4.10.1 of the Chapter Appendix.

Theorem 4.3 (Main Results).

1. *Let μ_t and $\mathcal{G}_t$ be as given in Lemma 4.5. Under condition [b], the process S_{tn} (translated sum-of-squares process) converges weakly to $W_t \equiv \mu_t - 2\mathcal{G}_t$; hence, $n^{1/3}(\widehat{\gamma}^*_{n,M} - \gamma_0)$ (i) has a non-trivial limiting distribution for all $M > 0$, which is the distribution of T_M, the minimizer over $[0, M]$ of W_t; and (ii) as $n \longrightarrow \infty$, then $M \longrightarrow \infty$, converges in distribution to T_∞, the global minimizer of W_t.*

2. *Under conditions [a] and [b], the MLE $\widehat{\gamma}_n$ converges to $\gamma_0 = 0$ no faster than $n^{-1/3}$.*

3. *As $n \longrightarrow \infty$, then $M \longrightarrow \infty$, the likelihood ratio test (deviance) statistic, $2\big[\ell_n(\widehat{\gamma}^*_{n,M}) - \ell_n(0)\big]$, converges in distribution to $-W_{T_\infty}/\sigma^2$.*

A main ingredient of Theorem 4.3 is the weak convergence of S_{tn}, so that a limiting distribution exists for its minimizer. The technical details of the idea appear in Section 4.10.2 of the Chapter Appendix. We provide the rest of the theorem proof in the next section.

4.8.6 Proof of Theorem 4.3

Assertion 1

By Lemma 4.5, the limiting process W_t is proportional to $t^3 + O_p(t^{3/2})$ (continuous), and drifts to ∞. Hence, for each $\epsilon > 0$, there is $\kappa > 0$ such that W_t attains its global minimum within $[0, \kappa]$ with probability of at least $1 - \epsilon$. Weak convergence of S_{tn} to W_t (see Section 4.10.2 of the Chapter Appendix) thus implies that, with probability tending to 1, the process S_{tn} attains its infimum over $[0, M]$ at some $Z_{n,M}$ which tends in distribution to T_∞, as $n \longrightarrow \infty$, then $M \longrightarrow \infty$. Equivalently,

$\lim_{M\to\infty} \lim_{n\to\infty} \mathcal{L}\left(n^{1/3}(\widehat{\gamma}^*_{n,M} - \gamma_0)\right)$ — the limiting distribution of the estimation error — is $\mathcal{L}\big(T_\infty\big)$.

Assertion 2

Condition [a] ensures consistency, or convergence, of $\widehat{\gamma}_n$, whether it resides within $Mn^{-1/3}$ or beyond. If beyond, its convergence rate is obviously no better than $n^{-1/3}$. If within, we proved above that $n^{-1/3}\widehat{\gamma}_n = n^{-1/3}\widehat{\gamma}^*_{n,M}$ has a non-trivial limit.

Assertion 3

The deviance statistic is

$$2\big[\ell_n(\widehat{\gamma}^*_{n,M}) - \ell_n(0)\big] = \frac{1}{\sigma^2}\big[S_n(0) - S_n(\widehat{\gamma}^*_{n,M})\big] = -\frac{1}{\sigma^2} S_{tn}\bigg|_{t=n^{1/3}\,\widehat{\gamma}^*_{n,M}} . \tag{4.44}$$

Its convergence to $-W_{T_\infty}/\sigma^2$ follows from Assertion 1. ∎

4.8.7 Final Remarks on Estimating the Kinked Truth

Maximum likelihood estimation used in modeling a broken stick by a bent cable is not only irregular, but inadequate and utterly peculiar, as we have seen via Theorem 4.3. In fact, there is another aspect to the impracticality, as highlighted by Lemma 4.6 below.

Lemma 4.6. *Label the design points in such a way that* $|x_1| \le |x_2| \le \ldots \le |x_n|$*. Then, given the regression (4.2) with model (4.40) where* $\gamma_0 = 0$*,*

1. *the sum-of-squares function* S_n *is constant everywhere on* $\big[0, |x_1|\big]$*;*
2. *if* $\varepsilon_1 > 0$*, then* $\big[0, |x_1|\big]$ *does not contain the global minimizer(s) of* $S_n(\gamma)$*.*

The lemma implies that $\widehat{\gamma}_n > |x_1|$ (> 0) happens with a probability of at least 1/2 for normally distributed errors. On this frequent event, the positive MLE points to a model that is qualitatively different from the truth. This unusual phenomenon also justifies defining $\widehat{\gamma}_n$ to be the least of all sum-of-squares minimizers, since any reasonable estimate of $\gamma_0 = 0$ should be close, if not equal, to 0 itself. The flatness in the likelihood function (negative sum of squares) at γ-values close to the origin is also evident in practical settings, such as appear in Figures 2.2 and 2.7 of Chapter 2.

We prove Lemma 4.6 in Section 4.10.1 of the Chapter Appendix.

4.9 Chapter Summary

Due to its fixed incoming and outgoing slopes, the basic bent cable is often inadequate for modeling natural phenomena. Nevertheless, we have presented in this chapter a framework for the upcoming full model theory.

First, under some simple conditions, standard large-sample maximum likelihood theory (consistency and asymptotic normality of the MLE; asymptotically χ^2 deviance statistic) applies to basic bent-cable estimation if *the truth is not kinked*. In the next chapter, we will see that, under slightly modified conditions, the standard theory is also true for the full bent-cable model.

For a kinked truth, however, the distributional properties of the one-parameter basic bent-cable MLE are highly unusual and virtually useless in practice. As a result, we will not pursue in the next chapter full bent-cable asymptotics involving an underlying broken stick.

Refer to Chapter 6 for a more in-depth summary of the results stated above.

4.10 Chapter Appendix

4.10.1 Lemma Proofs

Proof of Lemma 4.1

We begin with the first conclusion of the lemma.

Note that (4.3) to (4.5) are directly related to conditions [A] and [B] which are required for consistency of $\widehat{\boldsymbol{\theta}}_n$. Of course, a consistent parameter estimator cannot be obtained for an unidentifiable model.

Although not directly observed, (4.3) to (4.5) force the given points (w_0, z_0) and (w_1, z_1) to exhibit some properties. We will show that for any candidate basic bent cable $q_{\boldsymbol{\theta}}$ which passes through the (w_i, z_i)'s, these properties guarantee that $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, i.e. the candidate $q_{\boldsymbol{\theta}}$ is indeed the underlying cable $q_{\boldsymbol{\theta}_0}$.

Case $\gamma_0 = 0$. Note that (4.4) for $\gamma_0 = 0$ gives a w_0 value that is essentially τ_0, and so $z_0 = 0$. That is, the point $(w_0, 0)$ is observed. For the candidate cable $q_{\boldsymbol{\theta}}$ to pass through this point, it must be in its incoming linear phase at w_0. Next, the slope between (w_0, z_0) and (w_1, z_1) is observed to be exactly 1 due to (4.3) to (4.5). A slope of 1 between two points is unique to the outgoing linear phase of any basic bent cable. Thus, the (w_i, z_i)'s must both lie on the slope-one phase of $q_{\boldsymbol{\theta}}$. However, (w_0, z_0) also belongs to the slope-zero phase, forming a kink at w_0. Hence, $q_{\boldsymbol{\theta}}$ has a bend of half-width $\gamma = 0 = \gamma_0$, i.e. it has no bend but a kink at $\tau = w_0 = \tau_0$.

Case $\gamma_0 > 0$. By (4.3) to (4.5), both (w_i, z_i)'s are observed to be above the horizontal axis. Therefore, each must lie on either the bend or the slope-one phase of the candidate cable $q_{\boldsymbol{\theta}}$. However, they cannot both belong to the slope-one phase, as joining them yields a gentler slope than 1. Neither can they both lie on the bend of $q_{\boldsymbol{\theta}}$. To see this, first note that any basic cable $q_{\boldsymbol{\theta}}(x)$ has a non-decreasing slope from 0 to 1 as x increases. Therefore, the (w_i, z_i)'s being both on the bend would imply

$$q'_{\boldsymbol{\theta}}(w_0) \geq 0\,, \quad q'_{\boldsymbol{\theta}}(w_1) \leq 1\,. \tag{4.45}$$

Now, let $g(x)$ be the quadratic function through both (w_i, z_i)'s such that $g(x) = q_{\boldsymbol{\theta}}(x)$ for all $x \in [\tau - \gamma, \tau + \gamma]$. Since $q_{\boldsymbol{\theta}}$ is everywhere differentiable, $g(x)$ is forced to assume its minimum at $\tau - \gamma$. That is, $g(x) = \alpha(x - x_{\min})^2$, where $x_{\min} = \tau - \gamma$ and $\alpha > 0$ (for convexity). Now, observed are the relations

$$g(w_0) = \alpha(w_0 - x_{\min})^2 = z_0\,, \quad g(w_1) = \alpha(w_1 - x_{\min})^2 = z_1\,,$$

the latter of which gives $\alpha = z_1(w_1 - x_{\min})^{-2}$.

Define on $(-\infty, w_0]$ the function

$$\psi(x) = z_1(w_0 - x)^2 - z_0(w_1 - x)^2\,.$$

Therefore, solving $\psi(x) = 0$ yields two solutions for the value of $x_{\min}$. The properties of ψ at w_0 and w_1 determine which solution corresponds to the correct value for $x_{\min}$:

$$\psi(w_0) = -z_0(w_1 - w_0)^2 \ < \ 0 \tag{4.46}$$

$$\psi(w_1) = z_1(w_0 - w_1)^2 \ > \ 0 \tag{4.47}$$

$$\psi''(x) = 2(z_1 - z_0) \ > \ 0 \tag{4.48}$$

Using (4.46) to (4.48), one can show that the larger root of $\psi(x) = 0$ is greater than w_0. Hence, $x_{\min}$ must be the smaller root, or it will contradict the notion of (w_0, z_0) being on the bend of $q_{\boldsymbol{\theta}}$.

Now, use the unobserved relation (4.3), together with the smaller root of ψ for $x_{\min}$ to compute the value of $g'(w_1)$ $\left(= 2z_1(z_1 - x_{\min})^{-1}\right)$. The tedious algebra can be shown to reduce to $\left(q'_{\boldsymbol{\theta}}(w_1) =\right)$ $g'(w_1) > 1$, which contradicts (4.45). Although this result is not directly observed, it proves that there cannot be a candidate basic cable $q_{\boldsymbol{\theta}}$ whose bend passes through the given (w_i, z_i)'s.

Finally, the only possible candidate $q_{\boldsymbol{\theta}}$ must have (w_0, z_0) on its bend and (w_1, z_1) on its slope-one phase. Immediately, (w_1, z_1) determines that $\tau = w_1 - z_1 = \tau_0$. Next, the relations

$$z_0 = \frac{[w_0 - (w_1 - z_1) + \gamma]^2}{4\gamma} \tag{4.49}$$

$$q'_{\boldsymbol{\theta}}(w_0) \in (0, 1) \tag{4.50}$$

uniquely determine γ. In fact, the smaller solution to (4.49) yields a curvature that is too large, thus failing (4.50). Finally, since $\tau = \tau_0$ and γ is unique for these (w_i, z_i)'s, both of which have come from the underlying cable $q_{\boldsymbol{\theta}_0}$, we know γ must equal γ_0 itself.

Next, we prove the latter conclusion of the lemma.

Consider $\boldsymbol{\theta} \in \Theta_\delta^c \equiv \{\boldsymbol{\theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| > \delta\}$. (As $\gamma_0 = 0$, we only need to consider $\gamma \geq 0$.) We show

$$\boldsymbol{\theta} \in \Theta_\delta^c \implies q(w_i; \boldsymbol{\theta}) \neq z_i \text{ for } i = 0 \text{ or } 1 \,.$$

We partition Θ_δ^c by considering the three exhaustive cases of (a) $\tau < \tau_0 - \delta/2$, (b) $\tau > \tau_0 + \delta/2$, and (c) $\tau_0 - \delta/2 \leq \tau \leq \tau_0 + \delta/2$.

Case (a). Since $\tau < \tau_0 - 2\epsilon^*$, the candidate cable, $q_{\boldsymbol{\theta}}$, is everywhere above the two (w_i, z_i)'s, regardless of the value of γ. Hence, $q(w_i; \boldsymbol{\theta}) \neq z_i$ for $i = 0, 1$.

Case (b). Without loss of generality, assume $\tau_0 = 0$.

Suppose $q(w_i; \boldsymbol{\theta}) = z_i$ for $i = 0, 1$ (which we will show to be impossible for $\tau > \delta/2$). By the convexity and continuous differentiability of $q_{\boldsymbol{\theta}}$, we must have

(i) $q'_{\boldsymbol{\theta}}(x) \leq 1 \;\forall\; x$ and

(ii) $q(x; \boldsymbol{\theta}) \geq x \;\forall\; x \in [\tau - \gamma, w_1]$; hence,

(iii) $w_0 \in [-\epsilon^*, 0)$ and

(iv) $w_0 \leq \tau - \gamma < 0$.

Note that for fixed τ and x in the region of the bend, the function $q_{\boldsymbol{\theta}}$ is monotone non-decreasing as γ increases, subject to $\tau - \gamma \leq x \leq \tau + \gamma$. Thus, by (4.3), (iii), and (iv), we have

$$\begin{aligned} w_1 = q_{\boldsymbol{\theta}}(w_1) = \frac{(w_1 - \tau + \gamma)^2}{4\gamma} &\leq \frac{(w_1 - \tau + \tau + \epsilon^*)^2}{4(\tau + \epsilon^*)} = \frac{(w_1 + \epsilon^*)^2}{4(\tau + \epsilon^*)} \\ \implies\; 4w_1(\tau + \epsilon^*) &\leq w_1^2 + 2w_1\epsilon^* + (\epsilon^*)^2 \\ \implies\; 4w_1\tau &\leq w_1^2 - 2w_1\epsilon^* + (\epsilon^*)^2 = (w_1 - \epsilon^*)^2 < w_1^2 \\ \implies\; w_1 &> 4\tau \,. \end{aligned} \tag{4.51}$$

Now, by (i), we have

$$\begin{aligned} q'_{\boldsymbol{\theta}}(w_1) &= \frac{w_1 - \tau + \gamma}{2\gamma} \leq 1 \\ \implies\; w_1 &\leq \tau + \gamma \,. \end{aligned} \tag{4.52}$$

Combine (4.51) and (4.52) to see that $\gamma > 3\tau$. Combined with (iii) and (iv), this implies

$$\begin{aligned} & 3\tau < \gamma \leq \tau + \epsilon^* \\ \Longrightarrow \quad & \tau < \frac{\epsilon^*}{2} \end{aligned}$$

which contradicts the hypothesis of $\tau > \delta/2 > 2\epsilon^*$.

Case (c). Without loss of generality, assume $\tau_0 = 0$. Again, suppose $q(w_i; \boldsymbol{\theta}) = z_i$ for $i = 0, 1$ (which we will show to be impossible for $|\tau| \leq \delta/2$). For the values of τ in this case, we have $\gamma > \delta/2$ for all $\boldsymbol{\theta} \in \Theta_\delta^c$.

First, consider the case of $w_0 \in (0, \epsilon^*]$. Thus, $z_i = w_i$ for $i = 0, 1$. There are three possibilities for the candidate, $q_{\boldsymbol{\theta}}$:

(1) both (w_i, w_i)'s lie on the slope-one phase,

(2) (w_0, w_0) lies on the bend, but (w_1, w_1) lies on the slope-one phase, and

(3) both lie on the bend.

If (1) is true, then $\gamma \leq w_0 \leq \epsilon^* < \delta/4$, which contradicts $\gamma > \delta/2$. And (2) and (3) are both impossible, since they violate the convexity and continuous differentiability of $q_{\boldsymbol{\theta}}$. Thus, $w_0 \in (0, \epsilon^*]$ is impossible.

For $w_0 \in [-\epsilon^*, 0]$, the convexity and continuous differentiability of $q_{\boldsymbol{\theta}}$ imply $\tau - \gamma \geq w_0$. Now,

$$\gamma > \frac{\delta}{2} \implies \tau \geq w_0 + \frac{\delta}{2} > -\epsilon^* + 2\epsilon^* = \epsilon^* \,. \tag{4.53}$$

There are two possibilities for the candidate, $q_{\boldsymbol{\theta}}$:

(I) (w_1, w_1) on the bend, and

(II) (w_1, w_1) on the slope-one phase.

If (I) is true, then

$$\begin{aligned} & \tau + \gamma \geq w_1 \geq \delta_{11} > 2\delta \\ \Longrightarrow \quad & \gamma > 2\delta - \tau \geq 2\delta - \frac{\delta}{2} = \frac{3\delta}{2} \\ \Longrightarrow \quad & \tau - \gamma < \frac{\delta}{2} - \frac{3\delta}{2} = -\delta < -\frac{\delta}{4} < -\epsilon^* \leq w_0 \end{aligned}$$

which implies that w_0 is strictly inside the region of the bend, contradicting $z_0 = 0$. Thus, (I) is impossible.

If (II) is true, then

$$q_{\boldsymbol{\theta}}(\tau) = \frac{(\tau - \tau + \gamma)^2}{4\gamma} = \frac{\gamma}{4}\,.$$

Now, the convexity and continuous differentiability of $q_{\boldsymbol{\theta}}$ implies that $q_{\boldsymbol{\theta}}(x) \geq x$ for all $x > 0$ and that $w_0 \leq \tau - \gamma$. Therefore, $q_{\boldsymbol{\theta}}(\tau) \geq \tau$ if and only if $\gamma > 4\tau$, and so

$$\begin{aligned} & 4\tau < \gamma \leq \tau - w_0 \\ \Longrightarrow\quad & 3\tau < -w_0 \leq \epsilon^* \\ \Longrightarrow\quad & \tau < \frac{\epsilon^*}{3} \end{aligned} \tag{4.54}$$

which contradicts (4.53). Hence, (II) is impossible. ■

Proof of Lemma 4.2

Each score summand is piecewise continuously differentiable as a function of either τ or γ. Thus, $U_{\tau,i}$ can be retrieved either by integrating $V^+_{\tau\tau,i}$ over τ, or by integrating $V^+_{\gamma\tau,i}$ over γ. Hence,

$$\begin{aligned} U_{n\tau}(\tau,\gamma) - U_{n\tau}(\tau_0,\gamma) &= \int_{\tau_0}^{\tau} \sum_{i=1}^{n} V^+_{\tau\tau,i}(s,\gamma)\,ds \\ &= (\tau-\tau_0)\int_0^1 \sum_{i=1}^{n} V^+_{\tau\tau,i}\Big(\tau_0 + t(\tau-\tau_0),\gamma\Big)\,dt\,, \\ U_{n\tau}(\tau,\gamma) - U_{n\tau}(\tau,\gamma_0) &= (\gamma-\gamma_0)\int_0^1 \sum_{i=1}^{n} V^+_{\gamma\tau,i}\Big(\tau,\gamma_0 + t(\gamma-\gamma_0)\Big)\,dt\,. \end{aligned} \tag{4.55}$$

The algebra is similar for $U_{n\gamma}$. Then, since

$$U_{nj}(\boldsymbol{\theta}) = U_{nj}(\tau_0,\gamma_0) + \big[U_{nj}(\tau,\gamma_0) - U_{nj}(\tau_0,\gamma_0)\big] + \big[U_{nj}(\tau,\gamma) - U_{nj}(\tau,\gamma_0)\big]$$

for all $j = \tau, \gamma$, the lemma follows by routine algebra. ■

Proof of Lemma 4.3

First, we examine the components of $\mathbb{V}_n^+$ and $\mathbb{I}_n$. To simplify the algebra, write

$$\alpha_{1i} = \mathbf{1}\big\{x_i > \tau + \gamma\big\} \tag{4.56}$$

$$\alpha_{2i} = \frac{x_i - (\tau - \gamma)}{2\gamma}\mathbf{1}\big\{|x_i - \tau| \le \gamma\big\} \tag{4.57}$$

$$\alpha_{3i} = \frac{1}{4}\left[1 - \left|\frac{x_i - \tau}{\gamma}\right|^2\right]\mathbf{1}\big\{|x_i - \tau| \le \gamma\big\} \tag{4.58}$$

$$\alpha_{4i} = x_i - \tau \tag{4.59}$$

all of which are functions of $\boldsymbol{\theta}$ (but the argument is suppressed in the notation). Now, the partial derivatives of q become

$$\frac{\partial q_{\boldsymbol{\theta}}(x_i)}{\partial \tau} = -\alpha_{1i} - \alpha_{2i} \quad , \quad \frac{\partial q_{\boldsymbol{\theta}}(x_i)}{\partial \gamma} = \alpha_{3i}$$

and the directional derivatives for the score summands are

$$\begin{aligned}
V^+_{\tau\tau,i}(\boldsymbol{\theta}) &= -\frac{\alpha_{1i} + \alpha_{2i}^2}{\sigma^2} + \frac{1}{2\gamma\sigma^2}\Big[\varepsilon_i + d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\Big]\mathbf{1}\big\{\tau - \gamma < x_i \le \tau + \gamma\big\} \\
V^+_{\tau\gamma,i}(\boldsymbol{\theta}) &= \frac{\alpha_{2i}\,\alpha_{3i}}{\sigma^2} + \frac{\alpha_{4i}}{2\gamma^2\sigma^2}\Big[\varepsilon_i + d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\Big]\mathbf{1}\big\{\tau - \gamma < x_i \le \tau + \gamma\big\} \\
V^+_{\gamma\tau,i}(\boldsymbol{\theta}) &= \frac{\alpha_{2i}\,\alpha_{3i}}{\sigma^2} + \frac{\alpha_{4i}}{2\gamma^2\sigma^2}\Big[\varepsilon_i + d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\Big]\mathbf{1}\big\{|x_i - \tau| \le \gamma\big\} \\
V^+_{\gamma\gamma,i}(\boldsymbol{\theta}) &= -\frac{\alpha_{3i}^2}{\sigma^2} + \frac{\alpha_{4i}^2}{2\gamma^3\sigma^2}\Big[\varepsilon_i + d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\Big]\mathbf{1}\big\{|x_i - \tau| \le \gamma\big\}
\end{aligned}$$

and the Information matrix components, as functions of $\boldsymbol{\theta}$, are

$$I_{n,\tau\tau}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\sum(\alpha_{1i} + \alpha_{2i}^2) \tag{4.60}$$

$$I_{n,\tau\gamma}(\boldsymbol{\theta}) = -\frac{1}{\sigma^2}\sum \alpha_{2i}\,\alpha_{3i} = I_{n,\gamma\tau}(\boldsymbol{\theta}) \tag{4.61}$$

$$I_{n,\gamma\gamma}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\sum \alpha_{3i}^2 \tag{4.62}$$

By conditions [A] and [B], one can show that there is an N such that $n > N$ implies

$$\begin{aligned}
c_+^* + \left(\frac{\delta_1}{2\gamma_0}\right)^2 c_-^* \quad &\leq \quad \frac{\sigma^2}{n} I_{n,\tau\tau}(\boldsymbol{\theta}_0) \quad \leq \ 1 && (4.63)\\
\frac{\delta_1}{2\gamma_0}\left[1 - \left(\frac{\gamma_0 - \delta_1}{\gamma_0}\right)^2\right] c_-^* \quad &\leq \quad \frac{-4\sigma^2}{n} I_{n,\tau\gamma}(\boldsymbol{\theta}_0) \ \leq \ 1 && (4.64)\\
\left[1 - \left(\frac{\gamma_0 - \delta_1}{\gamma_0}\right)^2\right]^2 c_-^* \quad &\leq \quad \frac{16\sigma^2}{n} I_{n,\gamma\gamma}(\boldsymbol{\theta}_0) \quad \leq \ 1 && (4.65)
\end{aligned}$$

That is, for all $j, k = \tau, \gamma$, $I_{n,jk}(\boldsymbol{\theta}_0)$ is bounded between two non-trivial multiples of n for all sufficiently large n.

Now, define

$$\eta_{n,jk}(\boldsymbol{\theta}) = I_{n,jk}(\boldsymbol{\theta}) - I_{n,jk}(\boldsymbol{\theta}_0)\,, \qquad \Delta_{n,jk}(\boldsymbol{\theta}) = V^+_{n,jk}(\boldsymbol{\theta}) + I_{n,jk}(\boldsymbol{\theta})\,. \qquad (4.66)$$

Note that

$$\left|1 + \frac{V^+_{n,jk}(\boldsymbol{\theta})}{I_{n,jk}(\boldsymbol{\theta}_0)}\right| \leq \frac{\left|\eta_{n,jk}(\boldsymbol{\theta})\right| + \left|\Delta_{n,jk}(\boldsymbol{\theta})\right|}{\left|I_{n,jk}(\boldsymbol{\theta}_0)\right|}\,.$$

To show (4.12), it suffices, by (4.63) to (4.65), to show that for each pair of (j, k),

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\,\in\,\Theta_{\delta_n}} \left|\eta_{n,jk}(\boldsymbol{\theta})\right| \quad &= \quad o(n) && (4.67)\\
\sup_{\boldsymbol{\theta}\,\in\,\Theta_{\delta_n}} \left|\Delta_{n,jk}(\boldsymbol{\theta})\right| \quad &= \quad o_p(n) && (4.68)
\end{aligned}$$

There are three cases: (a) $j = k = \tau$, (b) $j = k = \gamma$, and (c) $j \neq k$.

First, for (4.67), the same logic applies to all three cases, despite somewhat different algebra. We present the details for Case (a) only.

Recall relations (4.60) to (4.62). Given any $\boldsymbol{\theta} \in \Theta_{\delta_n}$, we have

$$\sigma^2 \left|\eta_{n,\tau\tau}(\boldsymbol{\theta})\right| \leq \sum \left|\alpha_{1i}^2 - \alpha_{1i,0}^2\right| + \sum \left|\alpha_{2i}^2 - \alpha_{2i,0}^2\right| \qquad (4.69)$$

where $\alpha_{ki,0}$ $(k = 1, 2, 3)$ is the value of α_{ki} with $\boldsymbol{\theta}$ replaced by $\boldsymbol{\theta}_0$.

For the second sum in (4.69), partition the index set $\{1, \ldots, n\}$ into the four sets $\mathcal{I}_{2,1} \equiv \big\{i : |x_i - \tau| > \gamma,\, |x_i - \tau_0| > \gamma_0\big\}$, $\mathcal{I}_{2,2} \equiv \big\{i : |x_i - \tau| \leq \gamma,\, |x_i - \tau_0| \leq \gamma_0\big\}$, $\mathcal{I}_{2,3} \equiv \big\{i : |x_i - \tau| \leq \gamma,\, |x_i - \tau_0| > \gamma_0\big\}$, and $\mathcal{I}_{2,4} \equiv \big\{i : |x_i - \tau| > \gamma,\, |x_i - \tau_0| \leq \gamma_0\big\}$.

For $i \in \mathcal{I}_{2,1}$, we have $\alpha_{2i}^2 - \alpha_{2i,0}^2 = 0$.

For $i \in \mathcal{I}_{2,2}$, since $|\tau - \tau_0|, |\gamma - \gamma_0| \le \delta_n$, it can be shown that

$$\begin{aligned} \left|\alpha_{2i}^2 - \alpha_{2i,0}^2\right| &= \frac{1}{4}\left|\frac{x_i - (\tau-\gamma)}{\gamma} - \frac{x_i - (\tau_0 - \gamma_0)}{\gamma_0}\right|\left|\frac{x_i - (\tau-\gamma)}{\gamma} + \frac{x_i - (\tau_0 - \gamma_0)}{\gamma_0}\right| \\ &\le \frac{\delta_n\Big[\delta_n\big(|\tau_0| + 2\gamma_0\big) + 2\gamma_0\big(|\tau_0| + \gamma_0\big)\Big]}{\gamma_0(\gamma_0 - \delta_n)^2}. \end{aligned}$$

Denote this upper bound by ν_n, which is $\boldsymbol{\theta}$-free and has order $o(1)\big(\text{since } \delta_n = o(1)\big)$.

For $i \in \mathcal{I}_{2,3} \cup \mathcal{I}_{2,4}$, note that $\left|\alpha_{2i}^2 - \alpha_{2i,0}^2\right| \le 1$. Now, define $K_n^+ = [\tau_0 + \gamma_0 - \delta_n, \tau_0 + \gamma_0 + \delta_n]$ and $K_n^- = [\tau_0 - \gamma_0 - \delta_n, \tau_0 - \gamma_0 + \delta_n]$, which shrink to the join points $\tau_0 \pm \gamma_0$ as $n \longrightarrow \infty$. Then, by condition [C],

$$\begin{aligned} \sum_{i=1}^n \left|\alpha_{2i}^2 - \alpha_{2i,0}^2\right| &= \sum_{i \in \mathcal{I}_{2,1}} \left|\alpha_{2i}^2 - \alpha_{2i,0}^2\right| + \sum_{i \in \mathcal{I}_{2,2}} \left|\alpha_{2i}^2 - \alpha_{2i,0}^2\right| + \sum_{i \in \mathcal{I}_{2,3} \cup \mathcal{I}_{2,4}} \left|\alpha_{2i}^2 - \alpha_{2i,0}^2\right| \\ &\le 0 + n\,\nu_n + \sum_{i:\, x_i \in K_n^{\pm}} 1 \\ &\le n\,o(1) + n\,\zeta(\delta_n) \;=\; o(n) \end{aligned} \tag{4.70}$$

for all sufficiently large n. This bound is $\boldsymbol{\theta}$-free.

A similar argument applies to the first sum in (4.69), with the partition $\mathcal{I}_{1,1} \equiv \big\{i : x_i \le \tau + \gamma,\, x_i \le \tau_0 + \gamma_0\big\}$, $\mathcal{I}_{1,2} \equiv \big\{i : x_i > \tau + \gamma,\, x_i > \tau_0 + \gamma_0\big\}$, $\mathcal{I}_{1,3} \equiv \big\{i : x_i > \tau + \gamma,\, x_i \le \tau_0 + \gamma_0\big\}$, and $\mathcal{I}_{1,4} \equiv \big\{i : x_i \le \tau + \gamma,\, x_i > \tau_0 + \gamma_0\big\}$. Therefore, $\big|\eta_{n,\tau\tau}(\boldsymbol{\theta})\big|$ is uniformly bounded by $o(n)$ over Θ_{δ_n}, proving (4.67) for $j = k = \tau$.

Similarly, the algebra for Cases (b) and (c) reduces to $\big|\eta_{n,\gamma\gamma}(\boldsymbol{\theta})\big|$ and $\big|\eta_{n,\tau\gamma}(\boldsymbol{\theta})\big|$ both uniformly bounded by $o(n)$ over Θ_{δ_n}. We do not show the details, but point out that not having "accumulation" at the joints $\tau_0 \pm \gamma_0$ is crucial in keeping the matrices $\mathbb{I}_n(\boldsymbol{\theta})$ and $\mathbb{I}_n(\boldsymbol{\theta}_0)$ close everywhere on the shrinking neighborhood.

Now, to show (4.68), we need to present the details for both Cases (a) and (b), as they employ very different tactics. We leave Case (c) for the reader's verification.

Case (a) $j = k = \tau$

First, define

$$
\begin{aligned}
a_{n,\tau\tau}(\boldsymbol{\theta}) &= \frac{1}{2\gamma}\sum d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\,\mathbf{1}\{\tau-\gamma<x_i\le\tau+\gamma\} \\
B_{n,\tau\tau}(\boldsymbol{\theta}) &= \frac{1}{2\gamma}\sum \varepsilon_i\,\mathbf{1}\{\tau-\gamma<x_i\le\tau+\gamma\}
\end{aligned}
$$

so that

$$\sigma^2\,\Delta_{n,\tau\tau}(\boldsymbol{\theta}) = a_{n,\tau\tau}(\boldsymbol{\theta}) + B_{n,\tau\tau}(\boldsymbol{\theta})\,. \tag{4.71}$$

By (4.21), we have $\big|d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\big| \le \delta_n$ for all $\boldsymbol{\theta}\in\Theta_{\delta_n}$ for all i. Thus, on Θ_{δ_n},

$$\big|a_{n,\tau\tau}(\boldsymbol{\theta})\big| \le \frac{\delta_n}{2(\gamma_0-\delta_n)}\sum\mathbf{1}\{\tau-\gamma<x_i\le\tau+\gamma\} \le \frac{n\,\delta_n}{2(\gamma_0-\delta_n)}\,. \tag{4.72}$$

This upper bound is $\boldsymbol{\theta}$-free and has order $o(n)$.

Now, relabel the data so that $x_1\le x_2\le\ldots\le x_n$, and define the martingale

$$\mathcal{M}_m = \sum_{i=1}^{m}\varepsilon_i\,, \qquad 1\le m\le n\,.$$

For each $\boldsymbol{\theta}\in\Theta_{\delta_n}$, there are integers s and t, $1\le s\le t\le n$, such that $\sum_{i=1}^n\varepsilon_i\,\mathbf{1}\big\{\tau-\gamma<x_i\le\tau+\gamma\big\} = \mathcal{M}_t-\mathcal{M}_s$. Hence,

$$\big|B_{n,\tau\tau}(\boldsymbol{\theta})\big| = \frac{\big|\mathcal{M}_t-\mathcal{M}_s\big|}{2\gamma} \le \frac{\big|\mathcal{M}_t\big|+\big|\mathcal{M}_s\big|}{2\gamma} \le \frac{1}{\gamma_0-\delta_n}\left|\max_{1\le m\le n}\mathcal{M}_m\right|. \tag{4.73}$$

This upper bound is $\boldsymbol{\theta}$-free and has order $O_p(\sqrt{n})$ by the Doob-Kolmogorov inequality (see [34], p. 89, Problem 2).

Substitute (4.72) and (4.73) into (4.71) to get $\sup_{\boldsymbol{\theta}\,\in\,\Theta_{\delta_n}}\big|\Delta_{n,\tau\tau}(\boldsymbol{\theta})\big| \le o(n)+O_p(\sqrt{n}) = o_p(n)$. Thus, we have proved (4.68) for $j=k=\tau$.

Case (b) $j=k=\gamma$

Let us define

$$\omega_i(\boldsymbol{\theta})=\frac{\alpha_{4i}^2}{2\gamma^3}\mathbf{1}\{|x_i-\tau|\le\gamma\},\ a_{n,\gamma\gamma}(\boldsymbol{\theta})=\sum d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\omega_i(\boldsymbol{\theta}),\ B_{n,\gamma\gamma}(\boldsymbol{\theta})=\sum\varepsilon_i\omega_i(\boldsymbol{\theta}).$$

Thus, $\sigma^2\,\Delta_{n,\gamma\gamma}(\boldsymbol{\theta}) = a_{n,\gamma\gamma}(\boldsymbol{\theta})+B_{n,\gamma\gamma}(\boldsymbol{\theta})$. Note that $\omega_i(\boldsymbol{\theta})\le(2\gamma)^{-1}$ for all i. As we did to show (4.72), here, we can show that $\big|a_{n,\gamma\gamma}(\boldsymbol{\theta})\big| \le n\delta_n\big[2(\gamma_0-\delta_n)\big]^{-1} = o(n)$ everywhere on Θ_{δ_n}.

The martingale tactic for $B_{n,\tau\tau}$ does not apply to $B_{n,\gamma\gamma}$, because each summand of $B_{n,\gamma\gamma}$ involves ω_i, which varies with $\boldsymbol{\theta}$. Instead, define

$$h_i(\boldsymbol{\theta}) = \omega_i(\boldsymbol{\theta}) - \omega_i(\boldsymbol{\theta}_0)\ ,\ \ \Psi^d(\boldsymbol{\theta}) = \sum_i \big|h_i(\boldsymbol{\theta})\big|^2\ ,\ \ \Psi = \sum_i \big|\omega_i(\boldsymbol{\theta}_0)\big|^2\ .$$

Note that $\Psi \le n[2\gamma_0]^{-2} = O(n)$ uniformly on Θ_{δ_n}. To find an upper bound for $\Psi^d(\boldsymbol{\theta})$, again consider the four index sets $\mathcal{I}_{2,m}$, $m = 1, \ldots, 4$, defined earlier for showing (4.69). Similar to the argument there, it can be shown that

$$\big|h_i(\boldsymbol{\theta})\big|^2 \le \begin{cases} 0 & \forall\ i \in \mathcal{I}_{2,1} \\ 8\delta_n^2(\gamma_0-\delta_n)^{-4} & \forall\ i \in \mathcal{I}_{2,2} \\ (\gamma_0-\delta_n)^{-2} & \forall\ i \in \mathcal{I}_{2,3}\cup\mathcal{I}_{2,4} \end{cases} \tag{4.74}$$

for all $\boldsymbol{\theta} \in \Theta_{\delta_n}$. As the upper bounds in (4.74) are $\boldsymbol{\theta}$-free, we have

$$\sup_{\boldsymbol{\theta}\Theta_{\delta_n}} \Psi^d(\boldsymbol{\theta}) \le \frac{8\delta_n^2 n}{(\gamma_0-\delta_n)^4} + \frac{n\,\zeta(\delta_n)}{(\gamma_0-\delta_n)^2} = o(n)\,. \tag{4.75}$$

We will use the $\boldsymbol{\theta}$-free upper bounds for $\Psi(\boldsymbol{\theta})$ and $\Psi^d(\boldsymbol{\theta})$ to show that

$$\sup_{\boldsymbol{\theta}\,\in\,\Theta_{\delta_n}} \big|B_{n,\gamma\gamma}(\boldsymbol{\theta})\big| = o_p(n)\,. \tag{4.76}$$

First, given any $\epsilon > 0$, define events

$$\begin{aligned}
\mathcal{B}_n &= \left\{\sup_{\boldsymbol{\theta}\,\in\,\Theta_{\delta_n}} \Big|\sum_i \varepsilon_i\,\omega_i(\boldsymbol{\theta})\Big| \le n\epsilon\right\} \\
\mathcal{C}_n &= \left\{\Big|\sum_i \varepsilon_i\,\omega_i(\boldsymbol{\theta}_0)\Big| \le \frac{n\epsilon}{2}\right\} \\
\mathcal{D}_n &= \left\{\sup_{\boldsymbol{\theta}\,\in\,\Theta_{\delta_n}} \Big|\sum_i \varepsilon_i\,h_i(\boldsymbol{\theta})\Big| \le \frac{n\epsilon}{2}\right\}
\end{aligned}$$

Note that the event $\mathcal{C}_n \cap \mathcal{D}_n$ is a subset of $\mathcal{B}_n$. Thus, showing $P_{\boldsymbol{\theta}_0}\big\{\mathcal{C}_n^c\big\}, P_{\boldsymbol{\theta}_0}\big\{\mathcal{D}_n^c\big\} \to 0$ will show $P_{\boldsymbol{\theta}_0}\big\{\mathcal{B}_n\big\} \longrightarrow 1$, or (4.76).

For $\mathcal{C}_n^c$, the homoscedasticity of the ε_i's and Chebyshev's inequality implies

$$P_{\boldsymbol{\theta}_0}\big\{\mathcal{C}_n^c\big\} \le \frac{4\sigma^2\Psi}{\epsilon^2 n^2} \le \frac{\sigma^2}{\gamma_0^2\,\epsilon^2\,n} \longrightarrow 0\,.$$

For $\mathcal{D}_n^c$, we first note that $\sum \varepsilon_i^2 = O_p(n)$. Apply the Cauchy-Schwarz inequality and (4.75) to see that, for all $\boldsymbol{\theta} \in \Theta_{\delta_n}$,

$$\sup_{\boldsymbol{\theta} \,\in\, \Theta_{\delta_n}} \Big| \sum_i \varepsilon_i \, h_{i,m}(\boldsymbol{\theta}) \Big| \le \sqrt{\sum_i \varepsilon_i^2} \sqrt{\sup_{\boldsymbol{\theta} \,\in\, \Theta_{\delta_n}} \Psi^d(\boldsymbol{\theta}, \boldsymbol{\theta}_m)} = O_p(\sqrt{n})\; o(\sqrt{n}) = o_p(n)\,.$$

Therefore, $P_{\boldsymbol{\theta}_0}\big\{\mathcal{D}_n^c\big\} = P_{\boldsymbol{\theta}_0}\big\{o_p(n) > n\epsilon/2\big\} \longrightarrow 0$.

Having shown (4.76), we gather the bounds for $\big|a_{n,\gamma\gamma}\big|$ and $\big|B_{n,\gamma\gamma}\big|$ to see that $\big|\Delta_{n,\gamma\gamma}(\boldsymbol{\theta})\big| = o_p(n)$ uniformly over Θ_{δ_n}. Thus, we have proved (4.68) for $j = k = \gamma$.

The details for Case (c) are omitted; they closely resemble those for Case (b), but use an alternative definition for ω_i. ■

Proof of Lemma 4.5

Condition [b] ensures the convergence of the variances and covariances among the Gaussian processes $A_{tn,p}$, $p = 0, 1, 2$. For instance,

$$\mathrm{Cov}_0\big[A_{tn,0}, A_{tn,2}\big] \;=\; \mathrm{Var}_0\big[A_{tn,1}\big] \;=\; \sigma^2\, n^{\frac{2-2}{3}} \sum_1^n a_i\big(tn^{-1/3}, 2\big) \;\longrightarrow\; \sigma^2 c_2\, t^3\,.$$

The limiting variances and covariances for other values of p can be similarly obtained by applying [b]. Since R_{tn} is a linear combination of the correlated $A_{tn,p}$'s, these limits can be used to obtain

$$\mathrm{Var}_0\big[R_{tn}\big] \;\longrightarrow\; \frac{\sigma^2}{4}\left\{\frac{c_0}{4} - c_1 + \frac{3c_2}{2} - c_3 + \frac{c_4}{4}\right\} t^3\,. \tag{4.77}$$

The form of (4.77) coincides with (4.43) for $r = s = t$. Applying [b] in a similar fashion yields the general covariance structure in (4.42) and (4.43). Next, apply [b] once again to see pointwise convergence of g_{tn} to the given form of μ_t for all $t \ge 0$. Now, recall that for all $u \ge 0$ and for all i, $b_i(u) = d_{u,0}(x_i)$ is non-negative and monotone non-decreasing in u; hence, so is g_{tn}. By pointwise convergence and monotonicity, we conclude $g_{tn} \longrightarrow \mu_t$ uniformly on any closed interval in $\mathbb{R}$ (see [35], Exercise 7.17.) ■

Proof of Lemma 4.6

Assertion 1

The score function is

$$U_n(\gamma) = \frac{1}{4\sigma^2} \sum_1^n \Big[\varepsilon_i - d_{\gamma,0}(x_i)\Big] \left[1 - \frac{x_i^2}{\gamma^2}\right] \mathbf{1}\big\{|x_i| \le \gamma\big\}\,.$$

Note that as γ increases along Ω, U_n has a constant value of 0 until γ moves past $|x_1|$, which is closest to the kink of the broken stick. Equivalently, each deviation $Y_i - q(x_i;\gamma) = \varepsilon_i - d_{\gamma,0}(x_i)$ is unaffected whatsoever by varying γ over the interval $\big[0, |x_1|\big]$, leading to a constant S_n on this interval.

Assertion 2

Without loss of generality, assume $|x_2|$ is strictly greater than $|x_1|$. We show that for $\varepsilon_1 > 0$, there is $\gamma^* \in (|x_1|, |x_2|)$ such that $S_n(\gamma^*) < S_n(0)$.

It can be shown that, for $\gamma \in (|x_1|, |x_2|)$,

$$S_n^0(\gamma) \equiv S_n(\gamma) - S_n(0) = \Big\{\big[q(x_1;0) + q(x_1;\gamma)\big] - 2Y_1\Big\}\, d_{\gamma,0}(x_1)\,.$$

By the definition of $d_{\gamma,0}$, we have $S_n^0(\gamma) < 0$ if and only if

$$q(x_1;0) + q(x_1;\gamma) - 2Y_1 < 0\,. \tag{4.78}$$

Thus, it suffices to show (4.78) is true for some $\gamma^* \in (|x_1|, |x_2|)$.

First, consider the case $x_1 \le 0$. Hence, $q(x_1;0) = 0$, $q(x_1;\gamma) = (\gamma - |x_1|)^2/(4\gamma)$, and $Y_1 = \varepsilon_1 > 0$. That is, the LHS of (4.78) is $h(\gamma)/(4\gamma)$, where $h(\gamma) = \gamma^2 - 2\big(4Y_1 + |x_1|\big)\gamma + x_1^2$. Now, the concave function h has roots $r_\pm = \big(4Y_1 + |x_1|\big) \pm \sqrt{4Y_1}\sqrt{4Y_1 + 2|x_1|} > 0$. Therefore, for all $\gamma \in (r_-, r_+)$, we have $h(\gamma) < 0$, i.e. we have (4.78). As $r_- < \big(4Y_1 + |x_2|\big) - \sqrt{4Y_1}\sqrt{4Y_1} = |x_2|$, and $r_+ > \big(0 + |x_1|\big) + 0 = |x_1|$, we can take γ^* to be any value in the non-empty set $(r_-, r_+) \subset (|x_1|, |x_2|)$ so that (4.78) holds for $\gamma = \gamma^*$.

It can be similarly verified that the result is also true for $x_1 > 0$. ∎

4.10.2 Weak Convergence of the Translated Sum-of-Squares

(Refer to Section 4.8.3 and Lemma 4.5 for the definitions of the quantities below.)

Assertion 1 of Theorem 4.3 involves the weak convergence of the translated sum-of-squares process, S_{tn}, to the limiting process, W_t, over the t-interval of $[0, M]$ for any $M > 0$. To show this, it suffices to show uniform convergence of g_{tn} to μ_t, and weak convergence of R_{tn} to $\mathcal{G}_t$. The former convergence is true by Lemma 4.5. For the latter, we apply Theorem 15.6 in [36], which requires (a) convergence of finite dimensional distributions (fdd's) of R_{tn} to $\mathcal{G}_t$, and (b) tightness of the sequence R_{tn}.

For (a), fix $t_1, \ldots, t_k$ where $0 \le t_1 < \ldots < t_k \le M$ and $k < \infty$. Now, covergence of fdd's is the convergence in distribution of the vector $\left[R_{t_1 n}, \ldots, R_{t_k n}\right]$ to the vector $\left[\mathcal{G}_{t_1}, \ldots, \mathcal{G}_{t_k}\right]$. For normality of the ε_i's, we have exact normality of R_{tn}, and hence, of $\mathcal{G}_t$. Thus, convergence of fdd's is equivalent to convergence in covariance structure, the latter of which is true by Lemma 4.5.

For (b), fix t, t_1, and t_2 where $0 \le t_1 \le t \le t_2 \le M$. We show that

$$E_0 \left\{ \left| R_{tn} - R_{t_1 n} \right|^2 \left| R_{t_2 n} - R_{tn} \right|^2 \right\} \le \left| F(t_2) - F(t_1) \right|^4 \qquad \forall\ n \ge 1 \tag{4.79}$$

for some continuous and non-decreasing function F on $[0, M]$. For $u_2 \ge u_1 \ge 0$, define

$$\begin{aligned} \xi_{1i}(u_1, u_2) &= \left[\frac{(u_2 - |x_i|)^2}{4u_2} - \frac{(u_1 - |x_i|)^2}{4u_1} \right] \mathbf{1}\left\{ |x_i| \le u_1 \right\} \\ \xi_{2i}(u_1, u_2) &= \frac{(u_2 - |x_i|)^2}{4u_2} \mathbf{1}\left\{ u_1 < |x_i| \le u_2 \right\} \end{aligned}$$

so that $b_i(u_2) - b_i(u_1) = \xi_{1i}(u_1, u_2) + \xi_{2i}(u_1, u_2)$. It can be shown that, for each i, $0 \le \xi_{1i}(u_1, u_2), \xi_{2i}(u_1, u_2) \le (u_2 - u_1)/4$. As the two ξ's cannot be simultaneously non-zero, then, for all i,

$$0 \le b_i(u_2) - b_i(u_1) \le \frac{u_2 - u_1}{4} \mathbf{1}\left\{ |x_i| \le u_2 \right\} = \frac{u_2 - u_1}{4} a_i(u_2, 0) . \tag{4.80}$$

Now, define shorthands

$$\begin{aligned}
\eta_i &= b_i\big(t\,n^{-1/3}\big) - b_i\big(t_1\,n^{-1/3}\big) \\
\kappa_i &= b_i\big(t_2\,n^{-1/3}\big) - b_i\big(t\,n^{-1/3}\big) \\
\nu_n &= E_0\Big\{\big|R_{tn} - R_{t_1 n}\big|^2\,\big|R_{t_2 n} - R_{tn}\big|^2\Big\} \;=\; E_0\left\{\left|\sum_1^n \varepsilon_i\,\eta_i\right|^2\left|\sum_1^n \varepsilon_i\,\kappa_i\right|^2\right\} \\
&= \sum_i\sum_j\sum_k\sum_l E_0\big[\varepsilon_i\,\varepsilon_j\,\varepsilon_k\,\varepsilon_l\big]\eta_i\,\eta_j\,\kappa_k\,\kappa_l
\end{aligned}$$

By independence, $E_0\big[\varepsilon_i\,\varepsilon_j\,\varepsilon_k\,\varepsilon_l\big]$ is non-zero only when (1) $i = j = k = l$, in which case $E_0\big[\varepsilon_i\,\varepsilon_j\,\varepsilon_k\,\varepsilon_l\big] = 3\sigma^4$, the fourth moment of a $N(0,\sigma^2)$ random variable; and (2) $i = j \neq k = l$, or $i = k \neq j = l$, or $i = l \neq j = k$, in which case $E_0\big[\varepsilon_i\,\varepsilon_j\,\varepsilon_k\,\varepsilon_l\big] = \sigma^4$. Now, apply (4.80) to see that

$$\begin{aligned}
\nu_n &= 3\sigma^4\sum_i \eta_i^2\kappa_i^2 + \sigma^4\sum_{i\neq k}\sum \eta_i^2\kappa_k^2 + 2\sigma^4\sum_{i\neq j}\sum \eta_i\,\eta_j\,\kappa_i\,\kappa_j \\
&\leq \left(\frac{t_2 - t_1}{4n^{1/3}}\right)^4\left\{3\sigma^4\sum a_i\big(t_2\,n^{-1/3},0\big) + 3\sigma^4\Big[\sum a_i\big(t_2\,n^{-1/3},0\big)\Big]^2\right\} \\
&\leq 3\left(\frac{t_2 - t_1}{4/\sigma}\right)^4\left\{n^{-2/3}\sum a_i\big(t_2\,n^{-1/3},0\big) + \Big[n^{-2/3}\sum a_i\big(t_2\,n^{-1/3},0\big)\Big]^2\right\}.
\end{aligned}$$

The existence of a limit from condition [b] implies that there exists $K \in (0,\infty)$ such that $0 \leq n^{-2/3}\sum a_i(t_2\,n^{-1/3},0) \leq K$ for all $n \geq 1$. Finally, take K^* to be $(\sigma/4)\big[3K(1+K)\big]^{1/4}$ and $F(s)$ to be sK^*; thus, F is continuous and monotone increasing. Hence, $\nu_n \leq \big[(t_2 - t_1)K^*\big]^4 = \big|F(t_2) - F(t_1)\big|^4$ for all $n \geq 1$. Thus, we have proved (4.79).

Chapter 5

The Full Bent Cable

In Chapter 2, we saw that sensible applications of bent-cable regression require the model to have free slopes for the two linear phases.

Altogether, the full bent cable is a five-parameter model. In addition to the two free slopes, it involves a half-width and center for the bend region, and an overall "location" of the graph on the (x, y)-plane. If we denote the regression function by f, then one parametrization of f, as a function of the covariate x, is

$$f_{\boldsymbol{\theta}}(x) \equiv f(x; \boldsymbol{\theta}) = \beta_0 + \beta_1 x + \beta_2 q(x; \boldsymbol{\theta}), \quad \text{where} \quad \boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \tau, \gamma), \qquad (5.1)$$

and q is the basic bent cable model from Chapter 4. This parametrization gives the following characteristics to the full bent cable:

- the incoming linear phase has a slope of β_1 and a y-intercept of β_0;
- the outgoing linear phase has a slope of $\beta_1 + \beta_2$ and a y-intercept of $\beta_0 - \beta_2\tau$; and
- the middle bend is that of the basic bent cable, q, but scaled by a factor of β_2 (this scaling preserves the center, τ, and the half-width, γ, of the bend).

Note that for fixed values of the other parameters, β_0 can be regarded as a description of the vertical location of the graph of f.

Given a sequence of n covariate values, $\{x_i\}_{i=1}^{n}$, our regression model is now

$$Y_i = f(x_i\,;\boldsymbol{\theta}_0) + \varepsilon_i \tag{5.2}$$

where $\boldsymbol{\theta}_0 = (b_0, b_1, b_2, \tau_0, \gamma_0)$ is the underlying model parameter, and the ε_i's are i.i.d. random errors with mean 0 and variance σ^2. From now on, we consider estimating $\boldsymbol{\theta}_0$ by $\widehat{\boldsymbol{\theta}}_n$ via least squares (instead of maximum likelihood as in Chapter 4). If σ^2 is unknown, we estimate it by the minimized error mean-square,

$$\widehat{\sigma^2_n} = \overline{S}_n(\widehat{\boldsymbol{\theta}}_n) = \frac{1}{n}\sum_i \left|Y_i - f(x_i;\widehat{\boldsymbol{\theta}}_n)\right|^2 . \tag{5.3}$$

In the case of normally distributed errors, least-squares estimation of $\boldsymbol{\theta}_0$ and variance estimation via (5.3) are equivalent to maximum likelihood estimation.

As before, denote the regression domain by $\mathcal{X} \subset \mathbb{R}$, and the parameter space, $\Omega \subset \mathbb{R}^5$. In Section 5.2, we will discuss the form of Ω, and its role in the idea of consistency.

5.1 Results

Due to the impractical asymptotics in the case of $\gamma_0 = 0$ (see Section 4.8), we only consider the case of a bend region of non-trivial width for the true cable, $f_{\boldsymbol{\theta}_0}$. Under some design conditions (see Section 5.2.1), this regression problem (with $\gamma_0 > 0$) is regular, i.e. the usual results of consistency and asymptotic normality are valid for the LSE, $\widehat{\boldsymbol{\theta}}_n$, as well as the asymptotic χ^2 distribution for the deviance statistic (or the log-likelihood ratio statistic when the ε_i's are normally distributed). Furthermore, in the case of an unknown error variance, the minimized error mean-square, $\widehat{\sigma^2_n}$, is consistent.

A more in-depth summary of the results is found in Chapter 6, together with a practical paraphrase of the design conditions of Section 5.2.1 below.

5.2 Parameter Space and Design Conditions

Besides ensuring that the error sum-of-squares function, S_n, is asymptotically regular, our design conditions are posed also to guarantee identifiability for the regression problem. These conditions can be regarded as required characteristics of the dataset, $\{x_1, \ldots, x_n\}$.

In fact, the characteristics of both the dataset and the parameter space intertwine in the notion of regularity. For example, imposing compactness on Ω immediately rules out a $\widehat{\boldsymbol{\theta}}_n$ that is wildly distant from the true parameter, $\boldsymbol{\theta}_0$. If compactness is also imposed on $\mathcal{X}$, then the continuous S_n is automatically bounded uniformly over Ω for all n, making the argument for a consistent $\widehat{\boldsymbol{\theta}}_n$ relatively simple.

In Section 5.6.1 of the Chapter Appendix, we show that an entirely open Ω is deemed problematic in arguing for consistency in our five-parameter regression problem. However, as there is sometimes a trade-off between practicality and rigor,[1] we do not wish to impose compactness conditions merely as a technical device that allows for an overly-simplified proof of desirable estimation asymptotics. Instead, we will consider the very practical open regression domain, $\mathcal{X} = \mathbb{R}$, but relax the compactness assumptions on those parameter values whose unboundedness, given our choice of $\mathcal{X}$, does not hinder the proving of consistency for $\widehat{\boldsymbol{\theta}}_n$.

[1]Depending on the kind of compactness conditions being imposed, practicality would be sacrificed, to a certain degree, for simplification of rigor. For example, we have seen the restrictions placed on the search for candidate join points in the quadratic-quadratic model of [28] and [29] (see Chapter 3, page 27). In some situations, it may appear unreasonable to the investigator that one is not allowed to consider a candidate joint at, say, one of the five x-values that form the so-called basic design. In practice, one is generally willing to tolerate compactness conditions in the simple form of boundedness, especially when such bounds yield pleasant asymptotics. From a mathematical standpoint, however, this "localization" of the least-squares estimation problem has its disadvantages. In particular, there are no closed-form expressions for $\widehat{\boldsymbol{\theta}}_n$ due to non-linearity. Thus, if the remaining space in $\mathbb{R}^5$ that exists beyond the imposed bounds is simply ignored when studying the statistical properties of $\widehat{\boldsymbol{\theta}}_n$, then one is never sure how likely it is for the localized least-squares problem to behave differently from the global one (i.e. one in which $\widehat{\boldsymbol{\theta}}_n$ could possibly exist anywhere in $\mathbb{R}^5$). On the other hand, by considering an unbounded parameter space, one could assess quantitatively the impact of this localization on the asymptotics. In this thesis, we establish, for instance, that the impact is non-existent for at least the β_1-parameter, provided that the imposed bound is reasonably large. (See Section 5.6.1 of the Chapter Appendix for more details.) This way, we are assured that localizing the β_1-estimation problem (if one chooses to do so) is not merely a convenient device, but a mathematically sound practice with concrete probabilistic evidence to support it. While bounding Ω is solely a mathematical issue, having an unbounded $\mathcal{X}$ gives the investigator the freedom to place design points over vast ranges.

5.2.1 Design Conditions for Full Bent-Cable Regression

For least-squares estimation of the model parameters b_0, b_1, b_2, τ_0, and γ_0 of a full bent cable, $f_{\boldsymbol{\theta}_0}$, we consider the regression domain, $\mathcal{X} = \mathbb{R}$, and the parameter space,

$$\begin{aligned} \Omega &= [-M_1, M_1] \times \mathbb{R} \times [-M_3, -\epsilon_3] \cup [\epsilon_3, M_3] \times [-M_4, M_4] \times [0, \infty) \qquad (5.4) \\ \text{or} \quad \Omega &= \mathbb{R}^2 \times [-M_3, M_3] \times [-M_4, M_4] \times [0, M_5] \qquad (5.5) \end{aligned}$$

where the M_j's are some large but finite positive constants, and ϵ_3, a tiny positive constant. Note that bounds are not imposed on the space for β_1, nor are they for β_0 and γ simultaneously. In Section 5.6.1 of the Chapter Appendix, we will discuss in detail the rationale behind our choices of Ω. In a practical setting, one of them may be preferable depending on the investigator's prior understanding of the underlying bent-cable structure.

Recall the definitions of $c_-(\cdot)$, $c_+(\cdot)$, and $\zeta(\cdot)$ from Chapter 4, page 32. In addition, define

$$\begin{aligned} c_=(\delta) &= \liminf_{n \longrightarrow \infty} \frac{1}{n} \sum \mathbf{1}\{x_i \le \tau_0 - \gamma_0 - \delta\}, \qquad \delta > 0 \\ c_\ell(\delta_1, \delta_2) &= \liminf_{n \longrightarrow \infty} \frac{1}{n} \sum \mathbf{1}\{x_i \in [\tau_0 - \gamma_0 - \delta_2, \tau_0 - \gamma_0 - \delta_1]\}, \qquad \delta_2 > \delta_1 > 0 \\ c_r(\delta_1, \delta_2) &= \liminf_{n \longrightarrow \infty} \frac{1}{n} \sum \mathbf{1}\{x_i \in [\tau_0 + \gamma_0 + \delta_1, \tau_0 + \gamma_0 + \delta_2]\}, \qquad \delta_2 > \delta_1 > 0 \\ \nu(\delta, p) &= \limsup_{n \longrightarrow \infty} \frac{1}{n} \sum |x_i|^p \, \mathbf{1}\{|x_i| > \delta\}, \qquad \delta > 0 \quad \text{and} \quad p = 0, 1, 2, \ldots \end{aligned}$$

The following design conditions provide regularity for full bent-cable regression:

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

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

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

[C] $\forall\ \xi_n \downarrow 0$, $\zeta(\xi_n) \longrightarrow 0$.

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

[E] $\exists\ a > 0$ (large but finite) such that $\kappa_p \equiv \nu(a,p)\ <\ \infty\ \ \forall\ p = 1, 2.$

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

Not all conditions are needed for all lemmas and theorems stated in Section 5.4. For example, condition [F] is required only when we do not assume normally distributed random errors.

While conditions [A], [C], and [D] are unchanged from Chapter 4 (the basic case), we have modified condition [B], and relabeled it [B′]. It now requires that each linear phase contain two detached x-intervals, and that there exist a non-trivial fraction of data in each of these intervals. Therefore, conditions [A] and [B′] require altogether five detached regions containing non-trivial fractions of data: one in the bend, and two in each linear phase. (See Figure 5.1.) The regions from condition [B′] produce consistent intercept and slope estimators for each linear phase. Then, as we have seen in the basic case, the middle region from condition [A], together with consistent slope estimators for the linear phases, produce consistent estimators for the center and width parameters of the bend. We will refer to [A] and [B′] jointly as the "$\{2, 1, 2\}$-configuration for $f_{\boldsymbol{\theta}_0}$."

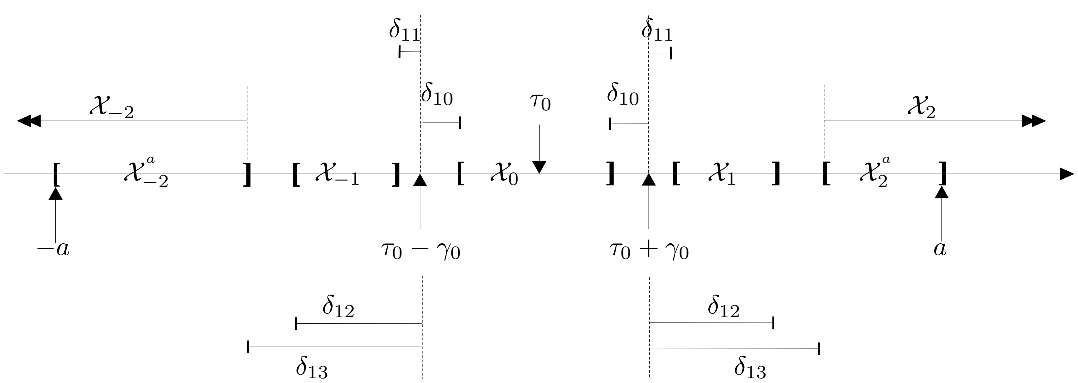

Figure 5.1: The $\mathcal{X}_i$ regions are required to contain non-trivial fractions of data to ensure regular asymptotics for the bent-cable estimation problem. We refer to this design as the "2-1-2"-configuration for the underlying bent cable. (Refer to Section 5.3 for the definitions of the $\mathcal{X}^a_i$'s.)

The rationale behind conditions [C] and [D] can be found in Chapter 4, Section 4.4.

The additional conditions for the full bent cable are [E] [2] and [F]. Now, condition [E] prevents the average absolute or squared covariate value from blowing up. As the terms $x_i\varepsilon_i$'s play a role in the error-sum-of-squares gradient (or simply, the gradient) function, such a condition is required to ensure that this function and its covariance matrix are asymptotically well-behaved. Finally, condition [F] slightly strengthens [E]. It requires the largest data point in absolute value to grow more slowly than the square root of the sample size. Recall that asymptotic normality of the LSE comes from an asymptotically normal gradient function. However, the latter would be difficult to achieve if the furthest covariate value puts too much weight on an ε_i that is non-normally distributed. (This is precisely the idea of the Lindeberg condition.) In practice, conditions [E] and [F] can be satisfied if the data are, say, restricted within a compact set, or, generated from any probability distribution with a finite variance.

---

[2]**Technical remarks on [E]:**

1. It suffices to have $\kappa_2 < \infty$. (Assume without loss of generality $a \geq 1$, so that $\kappa_1 = \limsup_{n\to\infty} n^{-1}\sum_{|x_i|>a}|x_i| \leq \limsup_{n\to\infty} n^{-1}\sum_{|x_i|>a} x_i^2 = \kappa_2$.)

2. Note that $\kappa_p < \infty$ is equivalent to $\limsup_{n\to\infty} n^{-1}\sum_1^n |x_i|^p < \infty$. (For all sequences $\{u_n\}$ and $\{v_n\}$, we have $\limsup_{n\to\infty}(u_n + v_n) \leq \limsup_{n\to\infty} u_n + \limsup_{n\to\infty} v_n$. Now, $n^{-1}\sum_{|x_i|>a}|x_i|^p \leq n^{-1}\sum_1^n |x_i|^p \leq a^p + n^{-1}\sum_{|x_i|>a}|x_i|^p$. Take the limit supremum of this inequality to get $\kappa_p \leq \limsup_{n\to\infty} n^{-1}\sum_1^n |x_i|^p \leq a^p + \kappa_p$.)

3. By Remark 2, we can take a of [E] to be any arbitrary positive constant.

4. Note that we can take $\kappa_0 \equiv \nu(a, 0)$ to be 1 (the maximum possible fraction of data outside of $[-a, a]$), and that $\kappa_p > a^p$ for all $p = 1, 2$.

5.3 Notation

Throughout the remainder of this thesis, notation from the basic case (Chapter 4) will be retained, except for the following modifications and additions:

$$
\begin{aligned}
d_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(x) &= f(x;\boldsymbol{\theta}_1) - f(x;\boldsymbol{\theta}_2) \\
d_i(\boldsymbol{\theta}) &= d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i) \\
W_i(\boldsymbol{\theta}) &= \varepsilon_i + d_i(\boldsymbol{\theta}) \\
S_n(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \left|W_i(\boldsymbol{\theta})\right|^2 \\
\ell_{n,\sigma^2}(\boldsymbol{\theta}) \;=\; \ell_n(\boldsymbol{\theta};\sigma^2) &= -\frac{1}{2}\left\{\frac{1}{\sigma^2}S_n(\boldsymbol{\theta}) + n\ln\sigma^2 + \ln(2\pi)\right\} \\
f_i(\boldsymbol{\theta}) &= f_{\boldsymbol{\theta}}(x_i) \;=\; f(x_i;\,\boldsymbol{\theta}) \\
\boldsymbol{\theta} \;=\; (\theta_1,\ldots,\theta_5) &= (\beta_0,\beta_1,\beta_2,\tau,\gamma) \\
\boldsymbol{\theta}_0 \;=\; (\theta_{01},\ldots,\theta_{05}) &= (b_0,b_1,b_2,\tau_0,\gamma_0) \\
\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}) &= \boldsymbol{\nabla}\ell_{n,\sigma^2}(\boldsymbol{\theta}) \qquad \text{where } \boldsymbol{\nabla} \text{ is taken with respect to } \boldsymbol{\theta} \\
\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}) &= \mathrm{Cov}_{\boldsymbol{\theta}_0}\left[\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta})\right] \\
\mathbb{V}_{n,\sigma^2}(\boldsymbol{\theta}) &= \boldsymbol{\nabla}\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}) \qquad \text{wherever defined} \\
\mathcal{X}_0 &= [\tau_0-\gamma_0+\delta_{10},\tau_0+\gamma_0-\delta_{10}] \\
\mathcal{X}_{-1} &= [\tau_0-\gamma_0-\delta_{12},\tau_0-\gamma_0-\delta_{11}] \\
\mathcal{X}_1 &= [\tau_0+\gamma_0+\delta_{11},\tau_0+\gamma_0+\delta_{12}] \\
\mathcal{X}_{-2}^{a} &= [-a,\tau_0-\gamma_0-\delta_{13}] \\
\mathcal{X}_{-2} &= (-\infty,a)\cup\mathcal{X}_{-2}^{a} \\
\mathcal{X}_{2}^{a} &= [\tau_0-\gamma_0+\delta_{13},a] \\
\mathcal{X}_{2} &= \mathcal{X}_{-2}^{a}\cup(a,\infty) \\
c^* &= \min\{c_0,c_{\pm1},c_{\pm2}\}
\end{aligned}
$$

where the δ_{1j}'s and a are from conditions [A], [B′], and [E].[3] Note that

$$
c_i = \liminf_{n\longrightarrow\infty}\frac{1}{n}\sum \mathbf{1}\left\{x_i\in\mathcal{X}_i\right\} \qquad \text{for all} \quad i=0,\pm1,\pm2
$$

[3]See page 76, Footnote 2, Remarks 2 and 3.

and is each strictly positive. Thus, $c^* > 0$.

To emphasize the dependence on σ^2, we have made σ^2 an extra subscript to ℓ, $\boldsymbol{U}$, $\mathbb{I}$, and $\mathbb{V}$.

When normal random errors were assumed in Chapter 4, $\boldsymbol{U}_n$ was taken to be the score function, i.e. the gradient of the likelihood function, ℓ_n. Except for a proportionality constant, $\boldsymbol{U}_{n,\sigma^2}$ here is also essentially the gradient of the error sum-of-squares, S_n. Thus, we remain interested in the properties of $\boldsymbol{U}_{n,\sigma^2}$ even in a general least-squares problem where the i.i.d. ε_i's may not follow the normal distribution. In this context, ℓ_{n,σ^2} should no longer be regarded as the log-likelihood function, but merely as a label for the anti-derivative of $\boldsymbol{U}_{n,\sigma^2}$.

In practice, we often need to estimate an unknown σ^2. As we will see in Section 5.4 below, an estimated σ^2 affects our regression problem through the distribution of $\boldsymbol{U}_{n,\sigma^2}$.

5.4 Formal Statements of Results

In this section, we present, in the form of lemmas and theorems, our results of model identifiability, consistency and asymptotic normality of the LSE's, and asymptotic properties of the deviance statistic.

Lemma 5.1 (Identifiability). *Given five points $\{(w_i, z_i)\}_{i=-2}^{2}$ such that*

$$\left.\begin{array}{rcl} z_i &=& f(w_i\,;\boldsymbol{\theta}_0) \\ w_i &\in& \mathcal{X}_i \end{array}\right\} \;\forall\;\; i = 0, \pm 1, \pm 2\,, \tag{5.6}$$

then, $$f(w_i\,;\boldsymbol{\theta}) = z_i \;\forall\; i = 0, \pm 1, \pm 2 \quad\Longrightarrow\quad \boldsymbol{\theta} = \boldsymbol{\theta}_0\,.$$

As in the basic case, identifiability of the model is directly related to conditions [A] and [B′] which contribute to the consistency of $\widehat{\boldsymbol{\theta}}_n$. To prove that $f_{\boldsymbol{\theta}_0}$ is identifiable by the (w_i, z_i)'s in a $\{2, 1, 2\}$-configuration, we consider all twenty-one possible five-point configurations for the candidate cable, $f_{\boldsymbol{\theta}}$. We will show that the given (w_i, z_i)'s cannot be in any configuration other than a $\{2, 1, 2\}$ if they were to lie on $f_{\boldsymbol{\theta}}$. We will further argue that, in order for the (w_i, z_i)'s to be $\{2, 1, 2\}$ on both $f_{\boldsymbol{\theta}_0}$ and $f_{\boldsymbol{\theta}}$, the true and candidate cables must indeed coincide everywhere. Again, convexity of the model and the smoothness constraints force the given (w_i, z_i)'s to belong uniquely to $f_{\boldsymbol{\theta}_0}$. The tedious proof is provided in Section 5.6.4 of the Chapter Appendix.

However, the $\{2,1,2\}$-configuration for the underlying $f_{\boldsymbol{\theta}_0}$ is not unique in leading to an identifiable bent cable. For example, a $\{1,3,1\}$-configuration — three distinct w_i's in the bend, and one w_i in each of the linear phases — also yields an identifiable $f_{\boldsymbol{\theta}_0}$. (The proof is outside the scope of this thesis.) From a practical point of view, a $\{2,1,2\}$-configuration may be more appealing when a bent-cable experiment is conducted as an alternative to the traditional broken-stick experiment. This is because the gradual transition region (i.e. the bend) could be small when a sharp change point is traditionally assumed for the regression function. It may then be infeasible to locate three detached regions within a small bend when the location of the bend itself is still in question.

In the following theorems, the parameter space is taken to be Ω as given by either (5.4) or (5.5).

Theorem 5.1 (Consistency). *Take conditions [A], [B′], and [E]. Then, the LSE, $\widehat{\boldsymbol{\theta}}_n$, and the minimized error mean-square, $\widehat{\sigma_n^2}$, are consistent estimators of $\boldsymbol{\theta}_0$ and σ^2, respectively.*

Corollary 5.1. *Under conditions [A], [B′], and [E], there exists a decreasing sequence $\xi_n \downarrow 0$ such that $P_{\boldsymbol{\theta}_0,\sigma^2}\{|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0|, |\widehat{\sigma_n^2} - \sigma^2| \le \xi_n\} \longrightarrow 1$ as $n \longrightarrow \infty$.*

The same logic behind Consistency Corollary 4.1 applies to Corollary 5.1 here.

Proving consistency for $\widehat{\boldsymbol{\theta}}_n$ involves the same three ideas for Theorem 4.1 of the basic case. They are, namely, (i) asymptotic boundedness of $\widehat{\boldsymbol{\theta}}_n$ in probability; (ii) nontrivial asymptotic curvature of the expected error mean-square function, H_n, around $\boldsymbol{\theta}_0$; and (iii) uniform convergence of the error mean-square function, $\overline{S}_n$, to H_n, in probability, over the bounded domain from (i). For the full bent cable here, $\overline{S}_n$ and H_n are no longer surfaces, but hypersurfaces over the five-dimensional Ω. Nonetheless, we can recycle virtually all the mathematical details from the proof of Theorem 4.1, with slight modifications due to the absence of condition [E] and presence of the normality assumption in the basic case. Note that whether σ^2 is estimated does not affect the consistency of $\widehat{\boldsymbol{\theta}}_n$, since $\widehat{\boldsymbol{\theta}}_n$ minimizes a function that is σ^2-free. The details appear in Section 5.5.

For proving consistency of $\widehat{\sigma_n^2}$, an idea analogous to (i) above is not necessary because of its closed form, as given by (5.3). Its consistency has the following heuristic. Consider n large. From (iii) above, $\overline{S}_n$ is, in probability, uniformly close to H_n over some bounded neighborhood of $\boldsymbol{\theta}_0$. By Lemma 5.1, we know that $H_n(\boldsymbol{\theta}_0)$ is the unique minimum value of H_n over Ω. Thus, the minimized error mean-square, $\widehat{\sigma_n^2} = \overline{S}_n(\widehat{\boldsymbol{\theta}}_n)$, is close to $H_n(\boldsymbol{\theta}_0) = \sigma^2$, in probability. The corresponding mathematical argument appears in Section 5.5.

Theorem 5.2 (Asymptotic Normality). *Assume $\gamma_0 > 0$. Under conditions [A], [B′], [C], [D], and [E],*

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

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

[3.] *condition [F] can replace the normality assumption for the ε_i's in [2.] above;*

[4.] *Assertions [1.] to [3.] above are true even if σ^2 in the expression of $\mathbb{I}_{n,\sigma^2}$ is replaced by $\widehat{\sigma_n^2}$.*

The discussion which follows Theorem 4.2 in Section 4.5 (pp. 35–36) applies to Assertions [1.] and [2.] of Theorem 5.2, except for what is now a five-dimensional parameter space, Θ_{ξ_n} (where ξ_n is from Corollary 5.1); a log-likelihood hypersurface, ℓ_{n,σ^2}, above the Θ_{ξ_n}-hyperplane; and a set of $2n$ "hyper-rays," $R_{i\pm}$'s, along which the Hessian of ℓ_{n,σ^2} is undefined.

In Assertion [3.], normality of the ε_i's is removed. Therefore, least-squares estimation is no longer equivalent to the method of maximum likelihood. However, the error sum-of-squares function, S_n, has a gradient that is essentially $\boldsymbol{U}_{n,\sigma^2}$, and a Hessian that is essentially $\mathbb{V}_{n,\sigma^2}$. Hence, the LSE, $\widehat{\boldsymbol{\theta}}_n$, remains to be a solution of $\boldsymbol{U}_{n,\sigma^2}$, and the Taylor expansion of $\boldsymbol{U}_{n,\sigma^2}$ (which involves $\mathbb{V}_{n,\sigma^2}$) is unaffected by removing normality of the ε_i's. Furthermore, by Lemma 5.5 in Section 5.5.3, $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)$ is asymptotically normal, leading to an asymptotically normal $\widehat{\boldsymbol{\theta}}_n$ through the Taylor-type expansion of $\boldsymbol{U}_{n,\sigma^2}\big(\widehat{\boldsymbol{\theta}}_n\big)$.

Finally, as the value of $\widehat{\boldsymbol{\theta}}_n$ does not depend on σ^2, Assertions [1.] to [3.] are affected by an estimated σ^2 only through the formula of $\mathbb{I}_{n,\sigma^2}$. Then, Assertion [4.] of Theorem 5.2 is true due to the consistency of $\widehat{\sigma^2_n}$.

The detailed proof of Theorem 5.2 is found in Section 5.5.

Theorem 5.3 (χ^2-limit). *Assume $\gamma_0 > 0$. Under conditions [A], [B′], [C], [D], [E], and [F], each deviance statistic below has a limiting χ^2 distribution, with its degree-of-freedom in parentheses:*

- $G_{n,\sigma^2} = 2\big[\ell_n(\widehat{\boldsymbol{\theta}}_n;\sigma^2) - \ell_n(\boldsymbol{\theta}_0;\sigma^2)\big]$ *(df=5), in the case of a known σ^2;*
- $G_n = 2\Big[\ell_n\Big(\widehat{\boldsymbol{\theta}}_n;\widehat{\sigma^2_n}\Big) - \ell_n\Big(\boldsymbol{\theta}_0;\widehat{\sigma^2_n}\Big)\Big]$ *(df=5), in the case of an unknown σ^2 estimated by $\widehat{\sigma^2_n}$; and*
- $D_n = n\Big[\ln\Big(\widehat{\sigma^2_n}^{*}\Big) - \ln\Big(\widehat{\sigma^2_n}'\Big)\Big]$ *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 $\widehat{\sigma^2_n}^{*} = \overline{S}_n(\widehat{\boldsymbol{\theta}}^{*}_n)$, $\widehat{\sigma^2_n}' = \overline{S}_n(\widehat{\boldsymbol{\theta}}'_n)$.*

Recall the basic case in which the χ^2-limit of the likelihood ratio statistic resulted from applying the Taylor-type expansion of the score, $\boldsymbol{U}_n$, to a (true) one-term Taylor expansion of ℓ_n. Note that if we assume normal ε_i's here, then the three deviance statistics of Theorem 5.3 are also likelihood ratio statistics, in which case the result for G_{n,σ^2} is merely an extension of Assertion (5.) of Theorem 4.2 (p. 35). However, as we have argued for Theorem 5.2 above, condition [F] can replace the normality assumption for the ε_i's without affecting the asymptotic properties of the quantities involved in the Taylor-type expansion of $\boldsymbol{U}_{n,\sigma^2}$. Further generalizations, based on a consistent $\widehat{\sigma^2_n}$, to the case of an unknown σ^2 and the testing of full bent-cable hypotheses yield the latter two conclusions of the theorem. The mathematical details are in Section 5.5.

5.5 Theorem Proofs

In this section, we provide the mathematical details of all the theorems presented in the chapter. Lemma proofs appear in Section 5.6.4 of the Chapter Appendix.

With the framework already provided through the basic case in Chapter 4, we avoid redundancy by not showing full details for all the proofs of this chapter. Any notational modifications and ingredients absent in the previous proofs will be carefully explained. However, in order to follow the mathematical logic presented below, it is important for the reader to have a clear idea of the corresponding proofs in Chapter 4.

5.5.1 Proof of Theorem 5.1

To prove consistency of $\widehat{\boldsymbol{\theta}}_n$, we apply the framework from the proof of Theorem 4.1 in Section 4.7. That is, we establish relations (4.13), (4.19), and (4.20) (pp. 39, 41, 43) for the five-parameter case here. Let us relabel these three relations, with necessary notational modifications, as follows:

$$\exists\ M^* \in (0,\infty)\ \text{ s.t. }\ P_{\boldsymbol{\theta}_0}\left\{\widehat{\boldsymbol{\theta}}_n \notin \Theta^*\right\} \longrightarrow 0 \qquad (5.7)$$
$$\Theta^* = \left\{\boldsymbol{\theta} : \left|\boldsymbol{\theta} - \boldsymbol{\theta}_0\right| \le M^*\right\}$$

$$\forall\ \delta \in (0, M^*]\ \ \exists\ N, \epsilon_1 > 0\ \text{ s.t. }\ n > N_1 \Longrightarrow \inf_{\boldsymbol{\theta}\,\in\, D_\delta} T_n(\boldsymbol{\theta}) \ge \epsilon_1 \qquad (5.8)$$
$$D_\delta = \left\{\boldsymbol{\theta} : \delta \le |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \le M^*\right\}$$

$$P_{\boldsymbol{\theta}_0}\left\{\sup_{\boldsymbol{\theta}\,\in\,\Theta^*}\left|\overline{S}_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta})\right| \le \epsilon\right\} \longrightarrow 1\ \ \forall\ \epsilon > 0 \qquad (5.9)$$

First, we find M^* for (5.7). To this end, we prove that the $\widehat{\theta}_{j,n}$'s ($j = 1, \ldots, 5$) — the components of $\widehat{\boldsymbol{\theta}}_n$ — are asymptotically bounded in probability. That is, for any $\epsilon > 0$, there exists N such that for all $n > N$, $P_{\boldsymbol{\theta}_0}\left\{\left|\widehat{\theta}_{j,n} - \theta_{0j}\right| \le M_j \text{ for all } j = 1, \ldots, 5\right\} > 1 - \epsilon$ for some M_j's which do not depend on ϵ. Then, M^* can be chosen accordingly to establish (5.7). As the boundedness in probability of the $\widehat{\theta}_{j,n}$'s is closely related to the form of Ω in (5.4) and (5.5), we leave it for the reader to verify that, not only does the discussion in Section 5.6.1 of the Chapter Appendix provide the intuition for formulae (5.4) and (5.5), but it also establishes the boundedness in probability of the $\widehat{\theta}_{j,n}$'s, thus proving (5.7).

Next, we show that the hypersurface H_n is not asymptotically flat by establishing relation (5.8). Note that our choice of δ here is slightly different from that for (4.19) of the basic case. Also note that D_δ is now a five-dimensional sphere whose "core" of radius δ has been punctured out.

Since D_δ is compact, there is an $a > 0$ so large such that, for all $\boldsymbol{\theta} \in D_\delta$, (i) $f_{\boldsymbol{\theta}}$ and $f_{\boldsymbol{\theta}_0}$ are both in their linear phases over $[a, \infty)$ and $(-\infty, -a]$, and (ii) they either entirely coincide or do not intersect at all over these intervals. Using this a (see page 76, Footnote 2, Remark 3), we define

$$\begin{aligned}
T^*(w_{-2}, w_{-1}, w_0, w_1, w_2, \boldsymbol{\theta}) &= \sum_{i=-2}^{2} \left| d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(w_j) \right|^2 , \\
C_{\delta,a} &= \left\{ \begin{array}{c} (w_{-2}, w_{-1}, w_0, w_1, w_2, \boldsymbol{\theta}) : \\ w_i \in \left\{ \begin{array}{lll} \mathcal{X}_i & \forall & i = 0, \pm 1 \\ \mathcal{X}_i^a & \forall & i = \pm 2 \end{array} \right. \\ \boldsymbol{\theta} \in D_\delta \end{array} \right\} , \\
C_{\delta,*} &= \lim_{K \longrightarrow \infty} C_{\delta,K} .
\end{aligned}$$

Note that $C_{\delta,*}$ is $C_{\delta,a}$ whose $\mathcal{X}_i^a$'s are replaced by $\mathcal{X}_i$ for all $i = \pm 2$.

By continuity of T^* and compactness of $C_{\delta,a}$, T^* attains its infimum, $\eta = \inf_{C_\delta} T^*$, on $C_{\delta,a}$. This infimum is strictly positive by Lemma 5.1. By the choice of a, the gap $|d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x)|$ is non-decreasing in x as x increases over $[a, \infty)$ or decreases over $(-\infty, -a]$. Thus, for any $\boldsymbol{\theta} \in D_\delta$, the function T^* is non-decreasing as w_2 increases over $[a, \infty)$ given fixed $w_i \in \mathcal{X}_i$ for $i = 0, \pm 1, -2$, and it is non-decreasing as w_{-2} decreases over $(-\infty, -a]$ given fixed $w_i \in \mathcal{X}_i$ for $i = 0, \pm 1, +2$. Hence,

$$\inf_{C_{\delta,*}} T^* = \lim_{K \longrightarrow \infty} \left\{ \inf_{C_{\delta,K}} T^* \right\} \geq \inf_{C_{\delta,a}} T^* = \eta > 0 .$$

Now, relabel and group the data $\{x_1, \ldots, x_n\}$ into the following six sets:

$$\begin{aligned} \mathcal{A}_i = \{s_{i1}, s_{i2}, \ldots, s_{ia_i}\} &= \left\{x_j : x_j \in \mathcal{X}_i\right\}, \quad i = 0, \pm 1, \pm 2 \\ \mathcal{B} = \{t_1, t_2, \ldots, t_b\} &= \left\{x_j : x_j \notin \bigcup_{i=-2}^{2} \mathcal{A}_i\right\} \end{aligned}$$

where the total cardinality is $b + \sum_{i=-2}^{2} a_i = n$. For any $\boldsymbol{\theta} \in D_\delta$, we consider T^* totaled over all the elements of the $\mathcal{A}_i$'s, as follows. By conditions [A] and [B′], there is an N_1 such that $a_i \geq nc^*$ for all $i = 0, \pm 1, \pm 2$, where c^* is from Section 5.3. Let $m = \min_i a_i$. Thus, $(m/n) \geq c^*$. Hence, for all $n > N_1$ and $\boldsymbol{\theta} \in D_\delta$, we have

$$T_n(\boldsymbol{\theta}) \;\geq\; \frac{1}{n}\sum_{j=1}^{m} T^*(s_{-2j},\, s_{-1j},\, s_{0j},\, s_{1j},\, s_{2j},\, \boldsymbol{\theta}) \;\geq\; \frac{1}{n}\sum_{j=1}^{m} \inf_{C_{\delta,*}} T^* \;\geq\; \eta c^*\,.$$

Take $\epsilon_1 = \eta c^* > 0$ to establish (5.8).

To complete the proof of consistency for $\widehat{\boldsymbol{\theta}}_n$, we establish (5.9). We adapt relations (4.22) to (4.25) (pp. 43–44) from the basic case to our five-parameter problem here.

Let us recapitulate these four relations. First, (4.22) asserts a certain uniform upper bound on $|d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)|$ for $i = 1, \ldots, n$ and $\boldsymbol{\theta} \in \Theta^*$. Similarly, (4.23) asserts a certain uniform upper bound on $|d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)|$ for $i = 1, \ldots, n$ and $\boldsymbol{\theta} \in B_k\ \forall\ k$, where B_k is the k-th subcover of Θ^* centered at $\boldsymbol{\theta}_k \in \Theta^*$, and has radius δ', $k = 1, \ldots, R$.[4] Both of these bounds are then utilized to establish relations (4.24) and (4.25), which assert uniform bounds on $\left|\overline{S}_n(\boldsymbol{\theta}) - \overline{S}_n(\boldsymbol{\theta}_k)\right|$ (in probability) and $\left|H_n(\boldsymbol{\theta}_k) - H_n(\boldsymbol{\theta})\right|$, respectively. All these bounds can be made arbitrarily small via an appropriately chosen δ' for any given ϵ in (4.20) (i.e. the basic-case version of (5.9)). As Chebyshev's inequality implies that $\overline{S}_n(\boldsymbol{\theta}_k)$ converges to $H_n(\boldsymbol{\theta}_k)$ pointwise (in probability) for all k, applying the triangle inequality establishes (4.20).

Here, we slightly modify the bounds of (4.22) and (4.23). In Section 5.6.2 of the Chapter Appendix, we prove

- relation (5.35), which, in this context, asserts that

$$\sup_{\boldsymbol{\theta}\,\in\,\Theta^*} |d_i(\boldsymbol{\theta})| \;\leq\; 3M^*\left[a\,\mathbf{1}\{|x_i| \leq a\} + |x_i|\,\mathbf{1}\{|x_i| > a\}\right], \quad \text{and}$$

[4] In fact, we took $\delta'/2$ instead of δ' as the radius of each B_k in the basic case. For algebraic convenience in the context of full bent-cable regression here, we now take δ' as the radius.

- relation (5.36), which, in this context, asserts that

$$\sup_{\boldsymbol{\theta}\,\in\,\Theta^*\,\cap\,B_k} \left|d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)\right| \;\leq\; 3\delta'\big[a\,\mathbf{1}\{|x_i|\leq a\} + |x_i|\,\mathbf{1}\{|x_i|>a\}\big] \qquad \forall\; k$$

  where each B_k has positive radius δ' that is chosen to give an appropriate bound for the event in (5.9): for any given $\epsilon > 0$, we take δ' to be such that $6\delta'\kappa_2(\mu^* + \sigma + 12M^*) < \epsilon$, where κ_2 is from condition [E], and $\mu^* = E(|\varepsilon_i|)$ for all i.

We are now ready to apply these modified bounds to establish bounds on $\left|\overline{S}_n(\boldsymbol{\theta}) - \overline{S}_n(\boldsymbol{\theta}_k)\right|$ and $\left|H_n(\boldsymbol{\theta}_k) - H_n(\boldsymbol{\theta})\right|$. To do so, we additionally apply condition [E], and the Cauchy-Schwarz inequality, as follows. According to condition [E], take N such that for all $n > N$, we have

$$\begin{aligned}
\left|H_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta}_k)\right| &\leq \frac{1}{n}\sum\Big(\left|d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)\right| + \left|d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x_i)\right|\Big)\left|d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)\right| \\
&\leq (3a\delta' + 3aM^*)(3a\delta') + 9\delta'(\delta' + M^*)\left\{\frac{1}{n}\sum_{|x_i|>a} x_i^2\right\} \\
&\leq 9\delta'(\delta' + M^*)(a^2 + \kappa_2) \;<\; 18\delta'\kappa_2(\delta' + M^*)\,, && (5.10) \\
\left|\overline{S}_n(\boldsymbol{\theta}) - \overline{S}_n(\boldsymbol{\theta}_k)\right| &\leq 2\left|\frac{1}{n}\sum_1^n \varepsilon_i\, d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)\right| + \left|H_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta}_k)\right|, && (5.11) \\
\left|\frac{1}{n}\sum_1^n \varepsilon_i\, d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)\right| &\leq \frac{1}{n}\sum_{|x_i|\leq a}\left|\varepsilon_i\right|\left|d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)\right| + \left|\sum_{|x_i|>a}\frac{\varepsilon_i\, d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)}{n}\right| \\
&< \frac{3a\delta'}{n}\sum_1^n|\varepsilon_i| + \sqrt{\frac{1}{n}\sum_1^n\varepsilon_i^2}\sqrt{\frac{1}{n}\sum_{|x_i|>a}\left|d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)\right|^2} \\
&\leq \frac{3a\delta'}{n}\sum_1^n|\varepsilon_i| + 3\delta'\sqrt{\kappa_2}\sqrt{\frac{1}{n}\sum_1^n\varepsilon_i^2} \\
&\xrightarrow{\text{P}} 3\delta'(a\mu^* + \sigma\sqrt{\kappa_2}) \\
&< 3\delta'\sqrt{\kappa_2}(\mu^* + \sigma) \;<\; 3\delta'\kappa_2(\mu^* + \sigma)\,. && (5.12)
\end{aligned}$$

Before applying the triangle inequality, we must verify the pointwise convergence, in probability, of $\overline{S}_n(\boldsymbol{\theta}_k)$ to $H_n(\boldsymbol{\theta}_k)$ for all k. Since some x_i's might be unbounded, we need to apply the uniform bound on $|d_i(\boldsymbol{\theta})|$ and condition [E] to Chebyshev's inequality to get:

$$\begin{aligned}
\forall\ \epsilon > 0, \quad P_{\boldsymbol{\theta}_0}\left\{\frac{1}{n}\left|\sum \varepsilon_i\, d_i(\boldsymbol{\theta}_k)\right| > \epsilon\right\} &\leq \frac{\sigma^2}{\epsilon^2 n}\left\{\frac{1}{n}\sum_{|x_i|\leq a}\left|d_i(\boldsymbol{\theta}_k)\right|^2 + \frac{1}{n}\sum_{|x_i|>a}\left|d_i(\boldsymbol{\theta}_k)\right|^2\right\} \\
&\leq \frac{9\sigma^2(M^*)^2}{\epsilon^2 n}\left\{a^2 + \frac{1}{n}\sum_{|x_i|>a} x_i^2\right\} \quad \forall\ \boldsymbol{\theta}_k \in \Theta^* \\
&= O\left(\frac{1}{n}\right)\left\{a^2 + O(1)\right\} \\
\therefore \quad \forall\ k, \quad \overline{S}_n(\boldsymbol{\theta}_k) - H_n(\boldsymbol{\theta}_k) &= \frac{1}{n}\sum \varepsilon_i^2 + \frac{2}{n}\sum \varepsilon_i\, d_i(\boldsymbol{\theta}_k) - \sigma^2 \xrightarrow{\mathrm{P}} 0\,.
\end{aligned}$$

Finally, note that Θ^* is larger than all subcovers, B_k's, and so $M^* > \delta'$. Thus, the choice of δ' (p. 85), relations (5.10) to (5.12), pointwise convergence (in probability) of $\overline{S}_n$ to H_n, and the triangle inequality establish (5.9).

For consistency of $\widehat{\sigma_n^2}$, we show

$$P_{\boldsymbol{\theta}_0,\,\sigma^2}\left\{\left|\overline{S}_n(\widehat{\boldsymbol{\theta}}_n) - \sigma^2\right| \leq \epsilon\right\} \longrightarrow 1 \quad \forall\ \epsilon > 0\,. \tag{5.13}$$

Given $\epsilon > 0$, take $0 < \delta \leq \min\left\{M^*, (1/5)\sqrt{\epsilon/\kappa_2}\right\}$ and N such that, by condition [E], the uniform bound on $|d_i|$, and relation (5.9), $n > N$ implies

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\,\in\,\Theta_\delta}\left|\overline{S}_n(\boldsymbol{\theta}) - \sigma^2\right| &= \sup_{\boldsymbol{\theta}\,\in\,\Theta_\delta}\left|\overline{S}_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta}) + T_n(\boldsymbol{\theta})\right| \\
&\leq \sup_{\boldsymbol{\theta}\,\in\,\Theta_\delta}\left|\overline{S}_n(\boldsymbol{\theta}) - H_n(\boldsymbol{\theta})\right| + \sup_{\boldsymbol{\theta}\,\in\,\Theta_\delta}\frac{1}{n}\sum_1^n\left|d_i(\boldsymbol{\theta})\right|^2 \\
&\overset{\mathrm{P}}{\leq} 9\delta^2\left\{a^2 + \frac{1}{n}\sum_{|x_i|>a} x_i^2\right\} \leq 9\delta^2(a^2+\kappa_2) < 18\delta^2\kappa_2 < \epsilon\,.
\end{aligned}$$

As consistency of $\widehat{\boldsymbol{\theta}}_n$ implies $\widehat{\boldsymbol{\theta}}_n \overset{\mathrm{P}}{\in} \Theta_\delta$, we have established (5.13). ■

5.5.2 Important Quantities in Theorems 5.2 and 5.3

We precede the detailed proofs of Theorems 5.2 and 5.3 by presenting some notation and auxiliary lemmas specific to the theorems.

As in the basic case, the gradient of the error sum-of-squares function and the covariance of this gradient play integral roles in the idea of asymptotic normality. Keeping the notation from Chapter 4, but using the modifications presented in Section 5.3, we have the following relations associated with $\boldsymbol{U}_{n,\sigma^2}$:

$$\begin{aligned}
\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}) &= -\frac{1}{2\sigma^2}\boldsymbol{\nabla} S_n(\boldsymbol{\theta}) \;=\; \boldsymbol{\nabla}\ell_{n,\sigma^2}(\boldsymbol{\theta}) \\
&= \frac{1}{\sigma^2}\sum W_i(\boldsymbol{\theta})\,\boldsymbol{\nabla} f_i(\boldsymbol{\theta}) \\
\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}) &= \frac{1}{\sigma^2}\sum \big[\boldsymbol{\nabla} f_i(\boldsymbol{\theta})\big]\big[\boldsymbol{\nabla} f_i(\boldsymbol{\theta})\big]^T \\
\text{where} \qquad \nabla_j\, f_i &= \frac{\partial f_i}{\partial \theta_j}, \qquad j = 1,\ldots,5 \\
\frac{\partial f_i(\boldsymbol{\theta})}{\partial \beta_0} = 1 \;\; \forall\, i \quad &, \quad \frac{\partial f_i(\boldsymbol{\theta})}{\partial \beta_1} = x_i \\
\frac{\partial f_i(\boldsymbol{\theta})}{\partial \beta_2} = \alpha_{1i}\,\alpha_{4i} + \gamma\,\alpha_{2i}^2 \quad &, \quad \frac{\partial f_i(\boldsymbol{\theta})}{\partial \tau} = -\beta_2\big\{\alpha_{1i} + \alpha_{2i}\big\} \quad , \quad \frac{\partial f_i(\boldsymbol{\theta})}{\partial \gamma} = \beta_2\,\alpha_{3i}
\end{aligned}$$

where, as in (4.56) to (4.59) from the basic case,

$$\alpha_{1i} = \mathbf{1}\big\{x_i > \tau + \gamma\big\}, \qquad \alpha_{2i} = \frac{x_i - (\tau - \gamma)}{2\gamma}\,\mathbf{1}\big\{|x_i - \tau| \le \gamma\big\},$$

$$\alpha_{3i} = \frac{1}{4}\left[1 - \left|\frac{x_i - \tau}{\gamma}\right|^2\right]\mathbf{1}\big\{|x_i - \tau| \le \gamma\big\}, \quad \text{and} \quad \alpha_{4i} = x_i - \tau\,.$$

Another important quantity associated with $\boldsymbol{U}_{n,\sigma^2}$ is the directional Hessian matrix, $\mathbb{V}^{+}_{n,\sigma^2}$. Its definition is analogous to that of (4.10) from the basic case (p. 37). Recall that its expected value is $-\mathbb{I}_{n,\sigma^2}$, so that we can simplify the elaborate form of $\mathbb{V}^{+}_{n,\sigma^2}$ by writing it in terms of $\mathbb{I}_{n,\sigma^2}$. To do so, we first need some notation for the (j,k)-th element of the i-th summand of $\mathbb{I}_{n,\sigma^2}$, with all dependence on $\boldsymbol{\theta}$ and σ^2 suppressed:

$$\begin{aligned}
I_{jk,i} \;\equiv\; I_{jk,i}(\boldsymbol{\theta}) \;=\; I_{kj,i}(\boldsymbol{\theta}) &= \frac{1}{\sigma^2}\big[\nabla_j\, f_i(\boldsymbol{\theta})\big]\big[\nabla_k\, f_i(\boldsymbol{\theta})\big] \quad \forall\; j,k = \beta_0,\beta_1,\beta_2,\tau,\gamma \\
I_{\beta_0\beta_0,\,i} = \frac{1}{\sigma^2}\;,\quad I_{\beta_0\beta_1,\,i} &= \frac{x_i}{\sigma^2}\;,\quad I_{\beta_0\beta_2,\,i} = \frac{1}{\sigma^2}\big\{\alpha_{1i}\,\alpha_{4i} + \gamma\,\alpha_{2i}^2\big\}\;, \\
& I_{\beta_0\tau,\,i} = -\frac{\beta_2}{\sigma^2}\big\{\alpha_{1i} + \alpha_{2i}\big\}\;,\quad I_{\beta_0\gamma,\,i} = \frac{\beta_2\,\alpha_{3i}}{\sigma^2} \\
I_{\beta_1 k,\,i} &= x_i\, I_{\beta_0 k,\,i} \qquad \forall\; k = \beta_0,\beta_1,\beta_2,\tau,\gamma
\end{aligned}$$

$$I_{\beta_2\beta_2,\,i} \;=\; \frac{1}{\sigma^2}\big\{\alpha_{1i}\,\alpha_{4i}^2+\gamma^2\alpha_{2i}^4\big\} \quad , \quad I_{\beta_2\tau,\,i} \;=\; -\frac{\beta_2}{\sigma^2}\big\{\alpha_{1i}\,\alpha_{4i}+\gamma\,\alpha_{2i}^3\big\}\;,$$

$$I_{\beta_2\gamma,\,i} \;=\; \frac{\beta_2\,\gamma\,\alpha_{2i}^2\,\alpha_{3i}}{\sigma^2}$$

$$I_{\tau\tau,\,i} \;=\; \frac{\beta_2^2}{\sigma^2}\big\{\alpha_{1i}+\alpha_{2i}^2\big\}\;,\quad I_{\tau\gamma,\,i} \;=\; \frac{-\beta_2^2\,\alpha_{2i}\,\alpha_{3i}}{\sigma^2}$$

$$I_{\gamma\gamma,\,i} \;=\; \frac{\beta_2^2\,\alpha_{3i}^2}{\sigma^2}$$

Now, we can write the (j,k)-th element of the i-th summand of $\mathbb{V}^+_{n,\sigma^2}$, again with $\boldsymbol{\theta}$ and σ^2 suppressed, as follows:

$$\begin{aligned}
V^+_{jk,i} \;=\; V^+_{kj,i} \;&=\; -I_{kj,i} \;=\; -I_{jk,i} \qquad\qquad (5.14)\\
&\forall \;\; j=\beta_0,\beta_1,\beta_2,\tau,\gamma \quad k=\beta_0,\beta_1 \quad \text{and } j=k=\beta_2\\
V^+_{\beta_2\tau,i} \;=\; V^+_{\tau\beta_2,i} \;&=\; -I_{\beta_2\tau,i}-\frac{(\alpha_{1i}+\alpha_{2i})W_i}{\sigma^2}\\
V^+_{\beta_2\gamma,i} \;=\; V^+_{\gamma\beta_2,i} \;&=\; -I_{\beta_2\gamma,i}+\frac{\alpha_{3i}\,W_i}{\sigma^2}\\
V^+_{\tau\tau,i} \;&=\; -I_{\tau\tau,i}+\frac{\beta_2\,W_i}{2\gamma\sigma^2}\,\mathbf{1}\big\{\tau-\gamma<x_i\le\tau+\gamma\big\}\\
V^+_{\tau\gamma,i} \;&=\; -I_{\tau\gamma,i}+\frac{\beta_2\,\alpha_{4i}\,W_i}{2\gamma^2\sigma^2}\,\mathbf{1}\big\{\tau-\gamma<x_i\le\tau+\gamma\big\}\\
V^+_{\gamma\tau,i} \;&=\; -I_{\tau\gamma,i}+\frac{\beta_2\,\alpha_{4i}\,W_i}{2\gamma^2\sigma^2}\,\mathbf{1}\big\{|x_i-\tau|\le\gamma\big\}\\
V^+_{\gamma\gamma,i} \;&=\; -I_{\gamma\gamma,i}+\frac{\beta_2\,\alpha_{4i}^2\,W_i}{2\gamma^3\sigma^2}\,\mathbf{1}\big\{|x_i-\tau|\le\gamma\big\}
\end{aligned}$$

Despite an undefined Hessian at a $\boldsymbol{\theta}$ along any hyper-ray, $R_{i\pm}$ (see page 37, equation (4.9)), the directional Hessian, $\mathbb{V}^+_{n,\sigma^2}$, is well defined over the entire parameter space, Ω. Note that (5.14) reflects the non-randomness of the first two rows and columns of $\mathbb{V}^+_{n,\sigma^2}$, as well as its (β_2,β_2)-th element. This is a result of the partial linearity of the bent-cable model, $f_{\boldsymbol{\theta}}$.

Recall the somewhat complex arguments in the integrand that appears in Lemma 4.2 (Taylor-type expansion of $\boldsymbol{U}_n$ on page 38) from the basic case. This complexity is due to a discontinuous directional Hessian, and is amplified by the five-parameter regression problem here. Section 5.6.3 of the Chapter Appendix is devoted to providing definitions to reduce the elaborate form of the multi-parameter Taylor-type expansion that appears as Lemma 5.2 in the next section.

5.5.3 Auxiliary Lemmas for Proving Theorem 5.2

As in the basic case, the framework for the asymptotic normal and χ^2 distributions includes a Taylor-type expansion of $\boldsymbol{U}_{n,\sigma^2}$; uniform convergence, in probability, of $-\mathbb{V}^+_{n,\sigma^2}$ to $\mathbb{I}_{n,\sigma^2}$ over the ξ_n-neighborhood of $\boldsymbol{\theta}_0$ (from Corollary 5.1); and convexity of $\overline{S}_n$, in probability, over Θ_{ξ_n}. Also, we add to this framework the asymptotic normality of $\boldsymbol{U}_{n,\sigma^2}$ to compensate for not assuming normally distributed ε_i's.

Refer to Section 5.6.4 of the Chapter Appendix for the proofs of Lemmas 5.2 to 5.5 below.

Lemma 5.2. *For all $\boldsymbol{\theta} \in \Omega$, we have*

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

where $\mathbb{V}^*_{n,\sigma^2}$ is given by (5.39) in the Chapter Appendix.

Lemma 5.3. *Given are conditions [A], [B′], [C], [D], and [E], and a sequence $\delta_n \downarrow 0$. As $n \longrightarrow \infty$, the negative directional Hessian, $-\mathbb{V}^+_{n,\sigma^2}$, covergences component-wise to $\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)$, in probability, uniformly over Θ_{δ_n}, in the following manner:*

$$\sup_{\boldsymbol{\theta} \,\in\, \Theta_{\delta_n}} \frac{1}{n} \left| I_{n,jk}(\boldsymbol{\theta}_0) + V^+_{n,jk}(\boldsymbol{\theta}) \right| \stackrel{\mathrm{P}}{\longrightarrow} 0 \qquad \forall\ j,k = \beta_0, \beta_1, \beta_2, \tau, \gamma$$

where $V^+_{n,jk}$ denotes the (j,k)-th component of $\mathbb{V}^+_{n,\sigma^2}$, and similarly for $\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)$. In fact, we have

$$\sup_{\boldsymbol{\theta} \,\in\, \Theta_{\delta_n}} \left| \eta_{n,jk}(\boldsymbol{\theta}) \right| = o(n)\,, \tag{5.15}$$

$$\sup_{\boldsymbol{\theta} \,\in\, \Theta_{\delta_n}} \left| \Delta_{n,jk}(\boldsymbol{\theta}) \right| = o_p(n)\,. \tag{5.16}$$

where, as in (4.66) of Chapter 4, $\eta_{n,jk}(\boldsymbol{\theta}) = I_{n,jk}(\boldsymbol{\theta}) - I_{n,jk}(\boldsymbol{\theta}_0)$ and $\Delta_{n,jk}(\boldsymbol{\theta}) = V^+_{n,jk}(\boldsymbol{\theta}) + I_{n,jk}(\boldsymbol{\theta})$.

Note that the quantity in Lemma 5.3 that converges in probability to 0 is slightly different from that in Lemma 4.3 for the basic case (p. 39). Also note that relations (5.15) and (5.16) together imply the main convergence result of the lemma.

Lemma 5.4. *Given are* $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Omega$ *and* n. *Define* $h_{n,\sigma^2}(t) = \ell_{n,\sigma^2}\big(\boldsymbol{\theta}_1 + t(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)\big)$ *for* $t \in [0,1]$. *Then,*

$$h''_{n,\sigma^2}(t) = (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)^T \Big[\mathbb{V}^+_{n,\sigma^2}\big(\boldsymbol{\theta}_1 + t(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)\big)\Big](\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1) \tag{5.17}$$

for all but isolated values of t.

Note that (5.17) is not automatically true by the chain rule. For example, if both $\boldsymbol{\theta}$'s lie on some hyper-ray $R_{i\pm}$ along which the Hessian of ℓ_{n,σ^2} is undefined, then the chain rule is inapplicable.

Lemma 5.5. *Assume that* $\varepsilon_1, \ldots, \varepsilon_n$ *are i.i.d. zero-mean random errors with constant finite variance,* σ^2. *Under conditions [A], [B′], [E], and [F], the Lindeberg-Feller Central Limit Theorem applies, and* $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)$ *is asymptotically normal. More precisely, for all fixed non-zero* $\boldsymbol{w} \in \mathbb{R}^5$, *we have*

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

The key ingredient of the lemma is that under condition [F], 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 Assertion [1.] of Theorem 5.2).

We now provide the proofs of Theorems 5.2 and 5.3.

5.5.4 Proof of Theorem 5.2

Assertion [1.]

To prove the first half of this assertion, we wish to show that there exists an N such that for all $n > N$, the matrix $n^{-1}\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0) = (n\sigma^2)^{-1}\sum_{i=1}^n \big[\boldsymbol{\nabla} f_i\big|_{\boldsymbol{\theta_0}}\big]\big[\boldsymbol{\nabla} f_i\big|_{\boldsymbol{\theta_0}}\big]^T$ is positive definite. (Recall from Section 5.3 that f_i refers to the full bent-cable function, evaluated at the i-th data point, x_i.) Similar to the proof of Assertion (1.) of Theorem 4.2, we establish a slightly stronger assertion, namely,

$$\exists\ N, K, \text{ and } \eta^* > 0 \text{ s.t.} \qquad n > N \implies \mu_{n,j} \in [n\eta^*, nK]\ \forall\ j = 1, \ldots, 5 \tag{5.18}$$

where $\mu_{n,j}$'s are the five eigenvalues of $\mathbb{I}_{n,\sigma^2} \equiv \mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)$.

First, let us establish the asymptotic finiteness of each $n^{-1}\mu_{n,j}$. Equivalently, we argue that each component of $n^{-1}\mathbb{I}_{n,\sigma^2}$ is asymptotically bounded. (This is necessary since we have yet to establish relations that are analogous to (4.63) to (4.65) on page 63 of the basic case.) In fact, we only need to consider the second row (and column) of $n^{-1}\mathbb{I}_{n,\sigma^2}$, since, from Section 5.5.2, we see that all the other elements are functions of α_{ji}'s only, which are bounded. Now, each second row element has the form $n^{-1}\sum_i x_i^p z_i$ for some bounded z_i's and some $p = 1, 2$. Thus, the element is bounded if the average absolute or squared covariate value is bounded. This is true for all sufficiently large n due to condition [E]. Hence, there is a K such that $\max_{1\le j\le 5}\big\{n^{-1}\mu_{n,j}\big\} \le K$ for all sufficiently large n.

It remains to show that $\mu_{n,*} \equiv \min_{1\le j\le 5}\big\{n^{-1}\mu_{n,j}\big\}$ is asymptotically bounded above 0 to establish (5.18). We can find $\mu_{n,*}$ explicitly by minimizing $\boldsymbol{v}^T\big[\mathbb{I}_{n,\sigma^2}\big]\boldsymbol{v}$ subject to $\boldsymbol{v}$ having unit length. To this end, let

$$\rho(\boldsymbol{v},\lambda) = \boldsymbol{v}^T\big[\mathbb{I}_{n,\sigma^2}\big]\boldsymbol{v} - \lambda(\boldsymbol{v}^T\boldsymbol{v}-1) \qquad \forall\ \boldsymbol{v}\in\mathbb{R}^5 \text{ and } \lambda\in\mathbb{R}$$

and solve this Lagrange minimization problem by first solving "$\boldsymbol{\nabla}\rho(\boldsymbol{v},\lambda) = \boldsymbol{0}$." Note that

$$\boldsymbol{\nabla}\rho(\boldsymbol{v},\lambda) = \begin{bmatrix} 2\big([\mathbb{I}_{n,\sigma^2}]\boldsymbol{v} - \lambda\boldsymbol{v}\big) \\ \boldsymbol{v}^T\boldsymbol{v} - 1 \end{bmatrix}.$$

Thus, $\boldsymbol{\nabla}\rho(\boldsymbol{v},\lambda) = \boldsymbol{0}$ if and only if $[\mathbb{I}_{n,\sigma^2}]\boldsymbol{v} = \lambda\boldsymbol{v}$, where $\boldsymbol{v}$ has unit length, i.e. if and only if λ is an eigenvalue of $\mathbb{I}_{n,\sigma^2}$, whose corresponding unit-length eigenvector is $\boldsymbol{v}$. Denote by $\boldsymbol{w}_n$ the unit-length eigenvector that corresponds to $\mu_{n,*}$. Now, if we take $\lambda = \mu_{n,*}$ and $\boldsymbol{v} = \boldsymbol{w}_n$, then $\boldsymbol{v}^T\big[\mathbb{I}_{n,\sigma^2}\big]\boldsymbol{v} = \boldsymbol{w}_n^T\mu_{n,*}\boldsymbol{w}_n = \mu_{n,*}$, hence minimizing $\boldsymbol{v}^T\big[\mathbb{I}_{n,\sigma^2}\big]\boldsymbol{v}$. We can now write

$$\mu_{n,*} = \inf_{\boldsymbol{v}:|\boldsymbol{v}|=1} \boldsymbol{v}^T\big[\mathbb{I}_{n,\sigma^2}\big]\boldsymbol{v} = \inf_{\boldsymbol{v}:|\boldsymbol{v}|=1} \frac{1}{\sigma^2}\sum_{i=1}^n \left(\boldsymbol{v}\cdot\boldsymbol{\nabla} f_i\Big|_{\boldsymbol{\theta}_0}\right)^2. \tag{5.19}$$

To establish (5.18), we use (5.19) and employ a compactness tactic similar to that in the proof of Theorem 5.1.

Take $a > 0$ to be very large but finite (see page 76, Footnote 2, Remark 3), and define

$$\begin{aligned} G(z_{-2}, z_{-1}, z_0, z_1, z_2, \boldsymbol{v}) &= \sum_{i=-2}^{2} \left(\boldsymbol{v} \boldsymbol{.} \boldsymbol{\nabla} f(z_i;\, \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}\,=\,\boldsymbol{\theta}_0} \right)^2 \\ &= \sum_{i=-2}^{2} \sum_{r=1}^{5} v_r^2 \left(\nabla_r\, f(z_i;\, \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}\,=\,\boldsymbol{\theta}_0} \right)^2 \\ &\qquad \text{where } \boldsymbol{\nabla} \text{ is taken with respect to } \boldsymbol{\theta}\,, \\ C_a &= \left\{ \begin{array}{c} (z_{-2}, z_{-1}, z_0, z_1, z_2, \boldsymbol{v}) : \\ z_i \in \left\{ \begin{array}{ll} \mathcal{X}_i & \forall \quad i = 0, \pm 1 \\ \mathcal{X}_i^a & \forall \quad i = \pm 2 \end{array} \right. \\ \boldsymbol{v} \in \mathbb{R}^5, \ |\boldsymbol{v}| = 1 \end{array} \right\}, \\ C_* &= \lim_{K \longrightarrow \infty} C_K \quad \text{(i.e. all } \mathcal{X}_i^a\text{'s of } C_a \text{ are replaced by } \mathcal{X}_i\text{'s)}\,. \end{aligned}$$

Since C_a is compact and G is continuous, G attains on C_a its infimum, $\eta_1 = \inf_{C_a} G \geq 0$. Now, recall the formulae for the $\nabla_r\, f$'s in Section 5.5.2. Note that at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, the quantity $\left|\nabla_r\, f(z;\, \boldsymbol{\theta})\right|$ for each r is non-decreasing in z as z increases over $[a, \infty)$, or as z decreases over $(-\infty, -a]$. Thus, $\inf_{C_*} G \geq \inf_{C_a} G = \eta_1 \geq 0$.

Take from the proof of Theorem 5.1 the sets $\mathcal{A}_i$'s and $\mathcal{B}$ (p. 84), and the integers N_1 and m. Then,

$$\begin{aligned} n > N_1 \implies \mu_{n,*} = \inf_{\boldsymbol{v}:|\boldsymbol{v}|=1} \boldsymbol{v}^T \left[\mathbb{I}_{n,\sigma^2}\right] \boldsymbol{v} \ &\geq \ \inf_{\boldsymbol{v}:|\boldsymbol{v}|=1} \frac{1}{\sigma^2} \sum_{i=-2}^{2} \sum_{j=1}^{m} \left(\boldsymbol{v} \boldsymbol{.} \boldsymbol{\nabla} f(s_{ij};\, \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}\,=\,\boldsymbol{\theta}_0} \right)^2 \\ &\geq \ \frac{1}{\sigma^2} \sum_{j=1}^{m} \inf_{\boldsymbol{v}:|\boldsymbol{v}|=1} G(s_{-2j}, s_{-1j}, s_{0j}, s_{1j}, s_{2j}, \boldsymbol{v}) \\ &\geq \ \frac{1}{\sigma^2} \sum_{j=1}^{m} \inf_{C_*} G \ = \ \frac{m\eta_1}{\sigma^2} \ \geq \ \frac{nc^*\eta_1}{\sigma^2} \ \geq \ 0\,. \end{aligned}$$

Take $N = N_1$ and $\eta^* = c^*\eta_1/\sigma^2$. Since $c^* > 0$ and σ^2 is finite, (5.18) can be established by showing that $\eta_1 \neq 0$. We do so by an argument of contradiction, as follows.

Suppose $\eta_1 = 0$. Denote by $(r_{-2}, r_{-1}, r_0, r_1, r_2, \boldsymbol{t})$ the minimizer of G over C_a. By the definition of η_1, we have

$$\boldsymbol{t} \boldsymbol{.} \boldsymbol{\nabla} f(r_i;\, \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}\,=\,\boldsymbol{\theta}_0} = \eta_1 = 0 \qquad \forall\ i = 0, \pm 1, \pm 2\,. \tag{5.20}$$

Now, by Section 5.5.2, we know

$$\nabla f(r_i;\,\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}\,=\,\boldsymbol{\theta}_0} \;=\; \begin{cases} (1,\, r_i,\, 0, 0, 0) & \text{for} \quad i = -2, -1 \\ (1,\, r_i,\, q_i,\, b_2 q'_{1i},\, b_2 q'_{2i}) & \text{for} \quad i = 0 \\ (1,\, r_i,\, r_i - \tau_0,\, -b_2,\, 0) & \text{for} \quad i = 1, 2 \end{cases}$$

$$\text{where} \qquad q_i \;=\; \frac{(r_i - \tau_0 + \gamma_0)^2}{4\gamma_0}\,, \quad q'_{1i} \;=\; \frac{\partial q_i}{\partial \tau_0}\,, \quad q'_{2i} \;=\; \frac{\partial q_i}{\partial \gamma_0}\,.$$

Note that q_0, q'_{10}, and q'_{20} are all non-zero by the definition of C_a. Substitute the cases for $i = -2, -1$ into (5.20) to get $t_1 + t_2 r_{-2} = t_1 + t_2 r_{-1} = 0$. Since $r_{-1} \neq r_{-2}$, we know $t_1 = t_2 = 0$. Together with the cases for $i = 1, 2$, we have $t_3(r_1 - \tau_0) - t_4 b_2 = t_3(r_2 - \tau_0) - t_4 b_2 = 0$. Since $r_1 \neq r_2$ and $b_2 \neq 0$, we have $t_3 = t_4 = 0$. Finally, apply these to the case for $i = 0$ to see that $t_5\, b_2\, q'_{20} = 0$. Since $b_2, q'_{20} \neq 0$, we have $t_5 = 0$ also. Assembling, we have $\boldsymbol{t} = \mathbf{0}$, contradicting the definition of C_a. Hence, $\eta_1 \neq 0$, and we have established (5.18).

Finally, we can apply (5.18), (5.15), and consistency of $\widehat{\boldsymbol{\theta}}_n$ (Theorem 5.1) to see that $n^{-1}\mathbb{I}_{n,\sigma^2}\big(\widehat{\boldsymbol{\theta}}_n\big)$ is asymptotically positive definite in probability.

Assertion [2.]

The argument for proving Assertions (2.) and (4.) of Theorem 4.2 from the basic case directly applies to full bent-cable regression, as follows.

First, a uniformly negative definite $\mathbb{V}^+_{n,\sigma^2}$ over the ever-decreasing Θ_{ξ_n}, in probability (which can be asserted by a five-parameter version of (4.29) on page 45), follows from Assertion [1.] above and Lemma 5.3 (uniform component-wise convergence over Θ_{ξ_n} of $-\mathbb{V}_{n,\sigma^2}$ to $\mathbb{I}_{n,\sigma^2}$). Now, by Lemma 5.4, the "chain-rule formula" for h''_{n,σ^2} in (5.17) is true (for any given n) for all $t \notin \mathcal{N}$, where $\mathcal{N} \in [0, 1]$ is the set of isolated values which correspond to the intersections of the hyper-rays, $R_{k\pm}$'s. (See Section 5.6.4, page 114, equation (5.47) for the explicit formulae for the elements of $\mathcal{N}$.) On the highly likely event that $\mathbb{V}^+_{n,\sigma^2}$ is negative definite, we therefore have $h''_{n,\sigma^2}(t) < 0$ except for $t \in \mathcal{N}$. This allows us to once again apply Lemma 4.4 (concavity via a negative value of h''_{n,σ^2}) to conclude an asymptotically concave ℓ_{n,σ^2} uniformly over Θ_{ξ_n}, in probability. Of course, a concave ℓ_{n,σ^2} is equivalent to a convex S_n. Thus, an asymptotically unique LSE $\big($p. 46, relation (4.32)$\big)$ follows from Lemma 4.4 and Corollary 5.1 (consistency).

Next, apply Lemma 5.2 (Taylor-type expansion of $\boldsymbol{U}_{n,\sigma^2}$) to yield a five-parameter version of relation (4.33), which essentially stipulates that $\boldsymbol{\theta}$ is asymptotically linear in $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}) - \boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)$ uniformly over Θ_{ξ_n}, in probability (p. 46). At $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_n$ (which is the asymptotically unique LSE within Θ_{ξ_n}), this therefore further stipulates that $\widehat{\boldsymbol{\theta}}_n$ is asymptotically linear (in probability) in $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)$, which is normally distributed. The final rigor for Assertion [2.] can be obtained by applying Lemma 5.3, Assertion [1.] above, exact normality of $U_{n,\sigma^2}(\boldsymbol{\theta}_0)$, and Slutsky's Theorem $\big($when $\widehat{\boldsymbol{\theta}}_n$ replaces $\boldsymbol{\theta}_0$ as the argument of $\mathbb{I}_{n,\sigma^2}$ in the five-parameter version of (4.33)$\big)$.

Assertion [3.]

Instead of being exactly normal, $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)$ is asymptotically normal under condition [F] by Lemma 5.5. Substitute this into the argument for Assertion [2.] above. As all other ingredients are unaffected by non-normally distributed ε_i's, the conclusion of an asymptotically normal $\widehat{\boldsymbol{\theta}}_n$ remains valid.

Assertion [4.]

The matrix $\mathbb{I}_{n,\sigma^2}$ is the only quantity of the Theorem that explicitly depends on σ^2 (division by σ^2). Hence, Assertion [4.] follows from consistency of $\widehat{\sigma_n^2}$ and Slutsky's Theorem. ■

5.5.5 Proof of Theorem 5.3

Recall from the basic case the equations (4.34) to (4.39) in the proof of Theorem 4.2, Assertion (5.). They involve algebraic manipulations of what is essentially a two-term Taylor-type expansion of ℓ_n about $\boldsymbol{\theta}_0$ (via a one-term Taylor-type expansion of $\boldsymbol{U}_n$) (pp. 47–48). The result is that the deviance statistic can be written as an inner product of two standard normal random vectors plus an asymptotically negligible term, so that the statistic has an asymptotic χ^2 distribution. Note that the χ^2-limit there is valid as long as $\boldsymbol{U}_n(\boldsymbol{\theta}_0)$ is asymptotically normal. Thus, by Lemma 5.5, the result from the basic case is easily extended to showing a χ^2_5-limit for G_{n,σ^2} in our five-parameter problem here, by mere modifications of notation, as follows: in the algebra, replace ℓ_n by ℓ_{n,σ^2}, $\boldsymbol{U}_n$ by $\boldsymbol{U}_{n,\sigma^2}$, $\mathbb{V}^*_n$ by $\mathbb{V}^*_{n,\sigma^2}$, and $\mathbb{I}_n$ by $\mathbb{I}_{n,\sigma^2}$.

The same χ^2_5-limit applies to G_n $\left(\text{i.e. } \sigma^2 \text{ estimated by } \widehat{\sigma^2_n}\right)$, due to the consistency of $\widehat{\sigma^2_n}$ (by Theorem 5.1) and Slutsky's Theorem.

Now, if the ε_i's were normal, then $\widehat{\boldsymbol{\theta}}^*_n$ and $\widehat{\sigma^2_n}^*$ would be the MLE's under H^*, and $\widehat{\boldsymbol{\theta}}'_n$ and $\widehat{\sigma^2_n}'$, under H', in which case the quantity

$$\begin{aligned} D_n &= 2\Big[\ell_n\Big(\widehat{\boldsymbol{\theta}}'_n; \widehat{\sigma^2_n}'\Big) - \ell_n\Big(\widehat{\boldsymbol{\theta}}^*_n; \widehat{\sigma^2_n}^*\Big)\Big] \\ &= 2\left[-\frac{1}{2}\left\{\frac{S_n\big(\widehat{\boldsymbol{\theta}}'_n\big)}{\overline{S}_n\big(\widehat{\boldsymbol{\theta}}'_n\big)} + n\ln \overline{S}_n\big(\widehat{\boldsymbol{\theta}}'_n\big)\right\} + \frac{1}{2}\left\{\frac{S_n\big(\widehat{\boldsymbol{\theta}}^*_n\big)}{\overline{S}_n\big(\widehat{\boldsymbol{\theta}}^*_n\big)} + n\ln \overline{S}_n\big(\widehat{\boldsymbol{\theta}}^*_n\big)\right\}\right] \\ &= n\ln\frac{\overline{S}_n\big(\widehat{\boldsymbol{\theta}}^*_n\big)}{\overline{S}_n\big(\widehat{\boldsymbol{\theta}}'_n\big)} \;=\; n\ln\frac{\widehat{\sigma^2_n}^*}{\widehat{\sigma^2_n}'} \end{aligned}$$

would be the likelihood ratio test statistic. In general, we replace the normality assumption by condition [F] and thus do not consider maximum likelihood.

Now, define $D_{n,\sigma^2} = G'_{n,\sigma^2} - G^*_{n,\sigma^2}$, where

$$G'_{n,\sigma^2} = 2\big[\ell_n\big(\widehat{\boldsymbol{\theta}}'_n; \sigma^2\big) - \ell_n\big(\boldsymbol{\theta}^*; \sigma^2\big)\big]\ , \quad G^*_{n,\sigma^2} = 2\big[\ell_n\big(\widehat{\boldsymbol{\theta}}^*_n; \sigma^2\big) - \ell_n\big(\boldsymbol{\theta}^*; \sigma^2\big)\big]\ ,$$

and $\boldsymbol{\theta}^*$ is the null parameter value (but q of its five components are estimated under H^*). Note that D_n and D_{n,σ^2} are identical, except for the two $\widehat{\sigma^2_n}$'s in D_n. By Theorem 5.1, both of the sets $\left(\widehat{\boldsymbol{\theta}}^*_n, \widehat{\sigma^2_n}^*\right)$ and $\left(\widehat{\boldsymbol{\theta}}'_n, \widehat{\sigma^2_n}'\right)$ are consistent under the same null hypothesis. Thus, the limiting null distribution for both D_n and D_{n,σ^2} are identical, and hence, it suffices to consider the latter.

In what follows, the logical arguments are standard for proving a χ^2 limiting distribution for a quadratic form of an almost-normal variate. Nevertheless, we expose the full details to illustrate that, under the conditions of Section 5.2.1, the irregularity (discontinuous Hessian) of our estimation problem does not interfere with these standard asymptotics.

First, extend relations (4.33), (4.37), and (4.38) from the basic case (pp. 46, 48) to the q-dimensional estimation problem under H^*. For the reader's convenience, we restate them below, with suitable notational modifications (and the consideration of only those $\boldsymbol{\theta}$'s whose q components of interest reside in the ever-decreasing ξ_n-neighborhood of the q unknown elements of $\boldsymbol{\theta}^*$):

- Relation (4.33) becomes

$$\boldsymbol{U}^q_{n,\sigma^2}(\boldsymbol{\theta}) = \boldsymbol{U}^q_{n,\sigma^2}(\boldsymbol{\theta}^*) + \frac{1}{n}\Big[-\mathbb{I}^q_{n,\sigma^2}(\boldsymbol{\theta}^*) + \mathbb{Z}_n\Big]\, n(\boldsymbol{\theta} - \boldsymbol{\theta}^*) \tag{5.21}$$

  where $\boldsymbol{U}^q_{n,\sigma^2}$ is obtained by differentiating ℓ_{n,σ^2} with respect to the q unknown parameters, and thus $\mathbb{I}^q_{n,\sigma^2} = \mathrm{Cov}\big[\boldsymbol{U}^q_{n,\sigma^2}\big]$ is a $q \times q$ submatrix of $\mathbb{I}_{n,\sigma^2}$ in the five-parameter problem; and $\mathbb{Z}_n$ is the generic label for what is essentially a 5×5 asymptotically-negligible random matrix.[5]

- Relation (4.37) becomes

$$\big(\widehat{\boldsymbol{\theta}}^*_n - \boldsymbol{\theta}^*\big)^T \Big[\boldsymbol{U}^q_{n,\sigma^2}(\boldsymbol{\theta}^*) - \boldsymbol{U}^{q\star}_{n,\sigma^2}\big(\widehat{\boldsymbol{\theta}}^*_n\big)\Big] \;=\; o_p(1) \tag{5.22}$$

$$\text{where} \qquad \boldsymbol{U}^{q\star}_{n,\sigma^2}(\boldsymbol{\theta}) = \frac{1}{n}\big[\mathbb{I}^q_n(\boldsymbol{\theta}^*) + \mathbb{Z}_n\big]\, n(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\,.$$

  (The superscript '$\star$' for $\boldsymbol{U}^q_{n,\sigma^2}$ is not to be confused with '$*$' which is used here to denote quantities associated with H^*, the null hypothesis.)

- Finally, (4.38) becomes

$$G^*_{n,\sigma^2} = \big(\widehat{\boldsymbol{\theta}}^*_n - \boldsymbol{\theta}^*\big)^T \boldsymbol{U}^q_{n,\sigma^2}(\boldsymbol{\theta}^*) + \big(\widehat{\boldsymbol{\theta}}^*_n - \boldsymbol{\theta}^*\big)^T \Big[\boldsymbol{U}^q_{n,\sigma^2}(\boldsymbol{\theta}^*) - \boldsymbol{U}^{q\star}_{n,\sigma^2}\big(\widehat{\boldsymbol{\theta}}^*_n\big)\Big]\,. \tag{5.23}$$

Now, assemble (5.21) to (5.23) while noting that the LHS of (5.21) is 0 if $\boldsymbol{\theta}$ is replaced by $\widehat{\boldsymbol{\theta}}^*_n$ in the equation. Then, we have

$$G^*_{n,\sigma^2} = \big[\boldsymbol{U}^q_{n,\sigma^2}(\boldsymbol{\theta}^*)\big]^T \big[\mathbb{I}^q_{n,\sigma^2}(\boldsymbol{\theta}^*)\big]^{-1} \boldsymbol{U}^q_{n,\sigma^2}(\boldsymbol{\theta}^*) + o_p(1)\,.$$

Similarly, for the p-dimensional estimation problem of H', we have

$$\begin{aligned} G'_{n,\sigma^2} &= \big[\boldsymbol{U}^p_{n,\sigma^2}(\boldsymbol{\theta}')\big]^T \big[\mathbb{I}^p_{n,\sigma^2}(\boldsymbol{\theta}')\big]^{-1} \boldsymbol{U}^p_{n,\sigma^2}(\boldsymbol{\theta}') + o_p(1) \\ &\overset{H^*}{=} \big[\boldsymbol{U}^p_{n,\sigma^2}(\boldsymbol{\theta}^*)\big]^T \big[\mathbb{I}^p_{n,\sigma^2}(\boldsymbol{\theta}^*)\big]^{-1} \boldsymbol{U}^p_{n,\sigma^2}(\boldsymbol{\theta}^*) + o_p(1) \end{aligned}$$

[5] See page 47 for more details on $\mathbb{Z}_n$ and its generic usage in the algebra.

where the symbol '$\stackrel{H^*}{=}$' denotes equating the LHS and RHS under the null hypothesis, H^*.

From now on, we drop $\boldsymbol{\theta}^*$ from the notation in order to reduce clutter in the display. After some standard matrix manipulation, we obtain

$$D_{n,\sigma^2} \stackrel{H^*}{=} \left[\boldsymbol{U}^p_{n,\sigma^2}\right]^T \left(\left[\mathbb{I}^p_{n,\sigma^2}\right]^{-1} - \mathbb{J}_n \right) \boldsymbol{U}^p_{n,\sigma^2} + o_p(1) \tag{5.24}$$

where $\mathbb{J}_n$ is a $p \times p$ matrix of 0's, except that q^2 of its components are filled accordingly by those of $\left[\mathbb{I}^q_{n,\sigma^2}\right]^{-1}$. We can rescale the quadratic form of (5.24) to get

$$\begin{aligned} D_{n,\sigma^2} &\stackrel{H^*}{=} \boldsymbol{W}_n^T \, \mathbb{Q}_n \, \boldsymbol{W}_n + o_p(1) \\ \text{where} \qquad \boldsymbol{W}_n &= \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right]^{-\frac{1}{2}} \left[\frac{1}{\sqrt{n}}\boldsymbol{U}^p_{n,\sigma^2}\right] \\ \text{and} \qquad \mathbb{Q}_n &= \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right]^{\frac{1}{2}} \left(\left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right]^{-1} - n\mathbb{J}_n \right) \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right]^{\frac{1}{2}} . \end{aligned}$$

Note that by Lemma 5.5, $\boldsymbol{W}_n$ is asymptotically standard normal under H^*. By Assertion [1.] of Theorem 5.2, all large n yield positive definite $n^{-1}\,\mathbb{I}^p_{n,\sigma^2}$ and $n^{-1}\,\mathbb{I}^q_{n,\sigma^2}$ matrices, and hence, a $\mathbb{Q}_n$ matrix of rank $p-q$. Moreover,

$$\begin{aligned} \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right]^{\frac{1}{2}} \mathbb{Q}_n \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right]^{\frac{1}{2}} &= \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right] \left(\left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right]^{-1} - n\mathbb{J}_n \right) \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right] \\ &= \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right] \left(\mathbb{K} - \mathbb{J}_n\mathbb{I}^p_{n,\sigma^2} \right) \end{aligned} \tag{5.25}$$

where $\mathbb{K}$ is the $p \times p$ identity matrix. More algebra will reduce $\mathbb{K} - \mathbb{J}_n\mathbb{I}^p_{n,\sigma^2}$ to a diagonal matrix, whose diagonal elements are each either 1 or 0. Hence, $\mathbb{K} - \mathbb{J}_n\mathbb{I}^p_{n,\sigma^2}$ is idempotent. Then, the RHS of (5.25) becomes

$$\begin{aligned} & \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right] \left(\mathbb{K} - \mathbb{J}_n\mathbb{I}^p_{n,\sigma^2} \right) \left(\mathbb{K} - \mathbb{J}_n\mathbb{I}^p_{n,\sigma^2} \right) \\ = \ & \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right] \left(\left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right]^{-1} - n\mathbb{J}_n \right) \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right] \left(\left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right]^{-1} - n\mathbb{J}_n \right) \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right] \\ = \ & \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right]^{\frac{1}{2}} \mathbb{Q}_n\mathbb{Q}_n \left[\frac{1}{n}\mathbb{I}^p_{n,\sigma^2}\right]^{\frac{1}{2}} . \end{aligned}$$

Therefore, $\mathbb{Q}_n$ is also idempotent.

Finally, apply the Fisher-Cochran Theorem to conclude that D_{n,σ^2} is asymptotically χ^2-distributed under H^*, with $p - q$ degrees of freedom. ■

5.6 Chapter Appendix

5.6.1 More on the Choice of Ω

Recall the basic bent-cable theory from Chapter 4. There, we searched for (τ_0, γ_0) over the parameter space $(-\infty, M] \times [0, \infty)$. The upper bound, M, was imposed on the τ-space to enable simple logic in the proof of consistency (Theorem 4.1).

The first part of the proof involved showing that for all sufficiently large samples, the LSE is highly unlikely to reside outside of a compact subset of the parameter space. We chose a tactic which employs the following central idea. As a parameter value, say θ_j $(j = 1, \ldots, 5)$, tends to $\pm\infty$, the gap between the true and candidate cables, $|d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x)|$, tends to ∞ *uniformly* over (1) some suitably chosen compact subset of the x-domain, denoted by $\mathcal{X}'_j$, and (2) the domains of all the remaining parameters, θ_k's, $k \neq j$. Denote the vector of these θ_k's by $\boldsymbol{\phi}_j$, and its domain by Φ_j. The choice of $\mathcal{X}'_j$ is made in such a way that it contains a non-trivial fraction of data for all large n. Then, as $\theta_j \longrightarrow \pm\infty$, those x_i's in $\mathcal{X}'_j$ are guaranteed to capture the ever-increasing gaps, leading to an $\overline{S}_n$ function that increases, in probability, *uniformly* over Φ_j. This uniformity ensures that while $\boldsymbol{\phi}_j$ is being varied over Φ_j, the LSE of the true value of θ_j is always highly likely to reside inside the closed interval $[-M_j, M_j]$ for some positive (and possibly huge) M_j. The rigor of this argument was previously presented through equations (4.14) to (4.18) on page 40. In particular, the crucial element that yields $\lim_{n\to\infty} P_{\boldsymbol{\theta}} \left\{ \widehat{\theta}_{j,n} \in [-M_j, M_j] \right\} = 1$ is

$$\exists \;\; N \;\; \text{s.t.} \;\; n > N \Longrightarrow \inf_{\substack{|\theta_j| \,>\, M_j \\ \boldsymbol{\phi}_j \,\in\, \Phi_j}} T_n(\boldsymbol{\theta}) \geq \frac{1}{n} \sum_i |d_i(\boldsymbol{\theta})|^2 \, \mathbf{1}\{x_i \in \mathcal{X}'_j\} \geq 6\sigma^2 \,. \tag{5.26}$$

Of course, the existence of such an N is due to the given design conditions.

This uniformity argument can be applied sequentially to each unknown parameter, with one possible modification. Once uniformity over x and $\boldsymbol{\phi}_j$ is proved for $\theta_j \longrightarrow \pm\infty$, one may wish to consider $\theta_{j'} \longrightarrow \pm\infty$ while restricting, without loss of generality, θ_j to vary over the above interval, $[-M_j, M_j]$, instead of the entire $\mathbb{R}$. (Of course, x and the new $\boldsymbol{\phi}_{j'}$ would be varied as before.)

However, in the basic case, it was not obvious that uniformity over the γ-domain of $[0,\infty)$ could be achieved as $\tau \longrightarrow +\infty$. In fact, $\partial d_i/\partial\tau \longrightarrow 0$ for all i and γ as $\tau \longrightarrow +\infty$. (See equation (5.28) below.) Thus, we employed a hard upper bound, M, for the τ-space in order to achieve an asymptotically bounded $\widehat{\tau}_n$.

Proving that $\widehat{\boldsymbol{\theta}}_n$ is asymptotically bounded becomes a much more complex exercise for the five-parameter full bent cable. There seems to be no obvious order of parameters in which one should consider the uniformity argument discussed above. The easiest tactic to simplify the rigor would be to restrict attention to an entirely compact $\boldsymbol{\theta}$-domain. However, we have already mentioned in Section 5.2 the possible trade-off between rigor and practicality. Thus, we wish to impose as few hard bounds as possible on the individual parameters' domains. Noting that $\boldsymbol{\nabla} d_i = \boldsymbol{\nabla} f_i$, we examine the gradient formulae from Section 5.5.2 for clues to recognizing those domains on which to impose hard bounds.

The Bad Guys

Of the five parameters, β_2 and τ are the potential "troublemakers" when left entirely unbounded. We discuss the respective parameters' domains below.

The β_2-space

Potential difficulties are immediate when β_2 is allowed to range over the entire real line. Indeed, we have suggested $[-M_3,-\epsilon_3]\cup[\epsilon_3,M_3]$ in (5.4) and $[-M_3,M_3]$ in (5.5) as the domain of β_2. The rationale is as follows. First,

$$\lim_{\beta_2\to 0}\frac{\partial d_i(\boldsymbol{\theta})}{\partial\tau} = \lim_{\beta_2\to 0}\frac{\partial d_i(\boldsymbol{\theta})}{\partial\gamma} = 0 \qquad \forall\ \boldsymbol{\theta}\,.$$

In fact, as $\beta_2 \longrightarrow 0$, the candidate $f_{\boldsymbol{\theta}}$ tends to a "degenerate" bent cable, i.e. a straight line with slope β_1 and y-intercept β_0, and with neither τ nor γ playing any role in its functional form. Thus, we impose a hard lower bound, ϵ_2, on the range of $|\beta_2|$ for (5.4), where γ is allowed to range anywhere along the non-negative real line. Furthermore,

$$\frac{\partial d_i(\boldsymbol{\theta})}{\partial\beta_2} = q(x_i;\,\boldsymbol{\theta}) = 0 \quad\Longleftrightarrow\quad x_i \le \tau-\gamma\,. \tag{5.27}$$

Depending on the value of $\boldsymbol{\theta}$, it is possible that data accumulation occurs only below $\tau - \gamma$, thus merely capturing the $f_{\boldsymbol{\theta}}$-$f_{\boldsymbol{\theta}_0}$ gap that is constant even as $\beta_2 \longrightarrow \pm\infty$. And by symmetry of the class of bent cables,[6] data accumulation beyond $\tau + \gamma$ would yield the same difficulty. However, there would be no obstacles in proving asymptotic boundedness for $\widehat{\beta}_{2,n}$ if we were to impose a hard upper bound on the range of $|\beta_2|$ for both (5.4) and (5.5).

The τ-space

Here, we examine the τ-derivative of d_i. As a basic bent cable is in the class of full bent cables, we can use this special case to demonstrate a need for bounds on τ. That is, we take $\boldsymbol{\theta}_0 = (0, 0, 1, \tau_0, \gamma_0)$ and $\boldsymbol{\theta} = (0, 0, 1, \tau, \gamma)$. Then,

$$\frac{\partial d_i(\boldsymbol{\theta})}{\partial \tau} = -\left\{ \frac{x_i - (\tau - \gamma)}{2\gamma} \mathbf{1}\big\{|x_i - \tau| \le \gamma\big\} + \mathbf{1}\big\{x_i > \tau + \gamma\big\} \right\} \tag{5.28}$$

which equals 0 for all $x_i < \tau - \gamma$. Now, we see that as $\tau \longrightarrow +\infty$, this partial derivative becomes 0 for any given γ and x_i. Since there may be insufficient data accumulation towards $+\infty$ to capture an increasing $|d_i|$ as $\tau \longrightarrow +\infty$, we impose a hard upper bound, M_4, on the τ-space.

This upper τ-bound can also be explained by visualizing the graph of a candidate basic cable sliding towards infinity away from the true basic cable. As $\tau \longrightarrow +\infty$, an ever increasing γ will "bring back" the candidate to the vicinity of the true cable. Thus, it is not obvious that $|d_i|$ uniformly increases over any x- or γ-domain as $\tau \longrightarrow +\infty$.

We now come to the question of having a hard lower bound on τ. The symmetry argument for (5.27) also applies here, although it does not apply to the class of basic bent cables in Chapter 4. (The mirror image of a basic cable about the y-axis is not a basic cable, but it is in the class of full bent cables.) Hence, we suggest $[-M_4, M_4]$ as a domain for τ.

[6]Denote the class of all bent-cable models by $\mathbf{C}$. Define $\mathbf{C}^* = \{\text{reflection of } f \text{ about the } y\text{-axis} : f \in \mathbf{C}\}$. Obviously, $\mathbf{C} \equiv \mathbf{C}^*$.

The Other Guys

While there is no obvious order in which to consider the remaining three parameters, we consider β_1 prior to β_0 or γ. In particular, we prove below that $P_{\boldsymbol{\theta}_0}\{\widehat{\boldsymbol{\theta}}_n$ is bounded$\} \longrightarrow 1$ can be achieved while: (a) the β_1-space is left unbounded, so long as the β_2-space is compact (regardless of the β_0-, τ-, and γ-spaces); (b) given (a), γ is allowed to vary over the non-negative real line, so long as the β_0- and τ-spaces are also compact, and β_2 is bounded away from 0; (c) given (a), the β_0-space is left unbounded, so long as the τ- and γ-spaces are also compact.

In fact, (a) and (b) together explain our choice of Ω in (5.4), while combining (a) and (c) explains that in (5.5).

Assertions (a) to (c) may be expressed mathematically by the following case-specific adaptations of (5.26) for $j = 2, 5, 1$:

- $j = 2$ (β_1):

$$\forall\ M_3 \in (0,\infty) \qquad \exists\ M_2, N \in (0,\infty) \text{ such that} \tag{5.29}$$
$$n > N \Longrightarrow \inf_{\substack{|\beta_1| > M_2; \\ |\beta_2| \le M_3}} T_n(\boldsymbol{\theta}) \ \ge\ 6\sigma^2$$

That is, we take $\beta_1 \longrightarrow \pm\infty$, and vary $\boldsymbol{\phi}_2 = (\beta_0, \beta_2, \tau, \gamma)$ over $\Phi_2 = \mathbb{R} \times [-M_3, M_3] \times \mathbb{R} \times [0,\infty)$.

- $j = 5$ (γ):

$$\forall\ M_1, M_2, M_3, M_4 \in (0,\infty) \text{ and } \epsilon_3 \in (0, M_3)\ \ \exists\ M_5, N \in (0,\infty) \text{ such that} \tag{5.30}$$
$$n > N \Longrightarrow \inf_{\substack{\gamma > M_5; \\ |\beta_0| \le M_1;\ |\beta_1| \le M_2; \\ \epsilon_3 \le |\beta_2| \le M_3;\ |\tau| \le M_4}} T_n(\boldsymbol{\theta}) \ \ge\ 6\sigma^2$$

That is, we take $\gamma \longrightarrow \infty$, and vary $\boldsymbol{\phi}_5 = (\beta_0, \beta_1, \beta_2, \tau)$ over $\Phi_5 = [-M_1, M_1] \times [-M_2, M_2] \times [-M_3, -\epsilon_3] \cup [\epsilon_3, M_3] \times [-M_4, M_4]$.

- $j = 1$ (β_0):

$$\forall\ M_2, M_3, M_4, M_5 \in (0,\infty) \qquad \exists\ M_1, N \in (0,\infty) \text{ such that} \tag{5.31}$$
$$n > N \Longrightarrow \inf_{\substack{|\beta_0| > M_1; \\ |\beta_1| \le M_2;\ |\beta_2| \le M_3; \\ |\tau| \le M_4;\ \gamma \le M_5}} T_n(\boldsymbol{\theta}) \ \ge\ 6\sigma^2$$

That is, we take $\beta_0 \longrightarrow \pm\infty$, and vary $\boldsymbol{\phi}_1 = (\beta_1, \beta_2, \tau, \gamma)$ over $\Phi_1 = [-M_2, M_2] \times [-M_3, M_3] \times [-M_4, M_4] \times [-M_5, M_5]$.

The reader may recall that $T_n(\boldsymbol{\theta}) \geq n^{-1} \sum_1^n |d_i(\boldsymbol{\theta})|^2 \, \mathbf{1}\{x_i \in \mathcal{X}_j'\}$ for all $\mathcal{X}_j' \subset \mathbb{R}$, and therefore the proofs of (5.29) to (5.31) involve choosing a suitable $\mathcal{X}_j'$ for each j.

We now proceed to the details of the three cases.

The β_1-space

To prove (5.29), first note that the slope of a bent cable is everywhere bounded between β_1 and $\beta_1 + \beta_2$. As we are considering bounded values of β_2, all sufficiently large values of β_1 lead to an $f_{\boldsymbol{\theta}}'$ that is everywhere larger than $|b_1| + |b_2| \geq \sup_x f_{\boldsymbol{\theta}_0}'(x)$, where b_1 and b_2 are the slopes associated with the true cable, $f_{\boldsymbol{\theta}_0}$. Consequently, there is an M_2^* such that for all $\beta_1 \geq M_2^*$, $f_{\boldsymbol{\theta}}$ is ensured to intersect $f_{\boldsymbol{\theta}_0}$ only once, regardless of τ, γ, and β_0 (which alters the location of the graph of $f_{\boldsymbol{\theta}}$ over the (x, y)-plane). For a given $\boldsymbol{\theta}$ whose $\beta_1 \in [M_2^*, \infty)$ and $(\beta_0, \beta_2, \tau, \gamma) = \boldsymbol{\phi}_2 \in \Phi_2 = \mathbb{R} \times [-M_3, M_3] \times \mathbb{R} \times [0, \infty)$, denote by $x_{\boldsymbol{\theta}}^*$ the sole point of intersection of $f_{\boldsymbol{\theta}}$ and $f_{\boldsymbol{\theta}_0}$. Note that $d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x_{\boldsymbol{\theta}}^*) = 0$. Then, as long as $\mathcal{X}_2'$ is detached from $x_{\boldsymbol{\theta}}^*$, the x_i's contained in it will capture the ever-increasing $|d_i|$'s as β_1 increases over $[M_2^*, \infty)$. Among the $\mathcal{X}_i$'s (see Section 5.3), at least one of them is detached from $x_{\boldsymbol{\theta}}^*$ for any given $\boldsymbol{\theta}$. Thus, as the slope of $f_{\boldsymbol{\theta}}$ increases due to an ever-increasing β_1, there is always an $\mathcal{X}_i$ that captures the increasing difference between $f_{\boldsymbol{\theta}}'$ and $f_{\boldsymbol{\theta}_0}'$, resulting in a large T_n, which is the average of the $|d_i|^2$'s.

More rigorously, take, say, $\delta_0 = (1/4) \times \min\{\text{width}(\mathcal{X}_0), \text{width}(\mathcal{X}_{\pm 1})\}$ and $M_2 > \min\{M_2^*, (\sigma/\delta_0)\sqrt{18/c^*} + M_3 + |b_1| + |b_2|\}$, recalling the definition of c^* from Section 5.3. Now, for any $\boldsymbol{\theta}$, we have $d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}'(x) = f_{\boldsymbol{\theta}_0}'(x) - f_{\boldsymbol{\theta}}'(x)$, which is negative for all $\beta_1 > M_2$ and $\boldsymbol{\phi}_2 \in \Phi_2$. Therefore, for all $\boldsymbol{\phi}_2 \in \Phi_2$,

$$\beta_1 > M_2 \implies \quad -d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}'(x) \;\geq\; M_2 - M_3 - |b_1| - |b_2| \;>\; \frac{\sigma}{\delta_0}\sqrt{\frac{18}{c^*}} \quad \forall \; x \, .$$

Hence, by a one-term Taylor expansion, we have

$$d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x) - 0 \;=\; d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x) - d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}(x_{\boldsymbol{\theta}}^*) \;=\; (x - x_{\boldsymbol{\theta}}^*) \int_0^1 d_{\boldsymbol{\theta}_0, \boldsymbol{\theta}}'\big(x_{\boldsymbol{\theta}}^* + t(x - x_{\boldsymbol{\theta}}^*)\big) \, dt$$

so that for all $\boldsymbol{\theta}$ whose $\beta_1 > M_2$ and $\boldsymbol{\phi}_2 \in \Phi_2$, we have

$$\left|d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x)\right|^2 \;>\; \delta_0^2 \int_0^1 \frac{18\sigma^2}{c^* \delta_0^2}\, dt \;=\; \frac{18\sigma^2}{c^*} \quad \forall\; x \in \mathcal{X}_2'(\boldsymbol{\theta}) \equiv \left\{x : |x - x_{\boldsymbol{\theta}}^*| > \delta_0\right\}\,.$$

By the choice of δ_0, we see that for all such $\boldsymbol{\theta}$, $\mathcal{X}_2'(\boldsymbol{\theta})$ contains at least one $\mathcal{X}_i$ that is detached from $x_{\boldsymbol{\theta}}^*$ (p. 75, Figure 5.1). Finally, choose N based on conditions [A] and [B′] such that, for all $n > N$, we have $n^{-1} \sum \mathbf{1}\left\{x_i \in \mathcal{X}_i\right\} > c^*$ for all $i = 0, \pm 1, \pm 2$. Thus, for all $n > N$, we also have $n^{-1} \sum_k \mathbf{1}\left\{x_k \in \mathcal{X}_2'(\boldsymbol{\theta})\right\} \geq c^*$ for all $\boldsymbol{\theta}$ whose $\beta_1 > M_2$ and $\boldsymbol{\phi}_2 \in \Phi_2$. Hence,

$$\begin{aligned} n > N \Longrightarrow \qquad \inf_{\beta_1 > M_2} T_n(\boldsymbol{\theta}) \;&\geq\; \inf_{\beta_1 > M_2} \frac{1}{n} \sum_k \left|d_k(\boldsymbol{\theta})\right|^2 \mathbf{1}\left\{x_k \in \mathcal{X}_2'(\boldsymbol{\theta})\right\} \\ &\geq\; \left(\frac{18\sigma^2}{c^*}\right) \frac{1}{n} \sum_k \mathbf{1}\left\{x_k \in \mathcal{X}_2'(\boldsymbol{\theta})\right\} \\ &\geq\; 6\sigma^2 \quad \forall \quad \boldsymbol{\phi}_2 \in \Phi_2\,. \end{aligned}$$

This establishes (5.29) for $\beta_1 \longrightarrow +\infty$. Of course, a similar argument applies to $\beta_1 \longrightarrow -\infty$. Hence, we have shown that $\widehat{\beta}_{1,n}$ is asymptotically bounded even if we search for b_1 over the entire real line.

The β_0- and γ-spaces

The last two parameter spaces to consider are those of $\beta_0 \in \mathbb{R}$ and $\gamma \in [0, \infty)$.

First, consider varying β_0 only. Ignoring the cable's bend region, we are left with a broken-stick "frame" for $f_{\boldsymbol{\theta}}$. Thus, varying β_0 while keeping other parameters fixed moves this frame structure in the y-direction.

Next, consider varying γ only. Depending on the values of β_1 and β_2, $f_{\boldsymbol{\theta}}$ may open upwards or downwards. If upwards, a fixed β_0 but increasing γ moves the entire $f_{\boldsymbol{\theta}}$ cable towards ∞. Thus, a large but negative β_0 puts the frame structure of an upward opening $f_{\boldsymbol{\theta}}$ near $-\infty$, while a large γ can pull the entire $f_{\boldsymbol{\theta}}$ back up towards the vicinity of $f_{\boldsymbol{\theta}_0}$. Similarly, a hugely positive β_0 and a huge γ nullify each other's effect on a downward-opening $f_{\boldsymbol{\theta}}$.

Therefore, it is apparent that the $|d_i|$'s do not necessarily diverge to ∞ as $|\beta_0|, \gamma \longrightarrow \infty$. To avoid difficulties in proving their asymptotic boundedness, we search over a compact space for one of these two parameters.

First, we establish (5.30) for the γ-space, as follows. We impose on $|\beta_0|$ a hard upper bound, M_1, and vary γ over $[0, \infty)$. Since β_1 is asymptotically bounded (in probability) by M_2, we may vary $\boldsymbol{\phi}_5 = (\beta_0, \beta_1, \beta_2, \tau)$ over $\Phi_5 = [-M_1, M_1] \times [-M_2, M_2] \times [-M_3, -\epsilon_3] \cup [\epsilon_3, M_3] \times [-M_4, M_4]$. As $\gamma \longrightarrow \infty$, $\partial d_i(\boldsymbol{\theta})/\partial\gamma \longrightarrow \beta_2/4$ for all i and for all $\boldsymbol{\phi}_5 \in \Phi_5$. Since the absolute value of this limit is bounded below by $\epsilon_3/4$ (strictly above 0), all $|d_i|$'s diverge to ∞ uniformly over Φ_5 as $\gamma \longrightarrow \infty$. Take $\mathcal{X}_5'$ to be, say, $\mathcal{X}_0$, and we have thus established relation (5.30).

Now, we establish (5.31) by reversing the roles of β_0 and γ above. That is, let β_0 vary over $\mathbb{R}$, and $\boldsymbol{\phi}_1 = (\beta_1, \beta_2, \tau, \gamma)$ over $\Phi_1 = [-M_2, M_2] \times [-M_3, M_3] \times [-M_4, M_4] \times [0, M_5]$. For all i and $\boldsymbol{\theta}$, we have $\partial d_i(\boldsymbol{\theta})/\partial\beta_0 = 1$, which is strictly above 0. Thus, as $\beta_0 \longrightarrow \pm\infty$, all $|d_i|$'s diverge to ∞ uniformly over Φ_5, where uniformity is a result of the compactness of Φ_1. Finally, take, say, $\mathcal{X}_0$ as $\mathcal{X}_1'$ to establish (5.31).

Remarks

The ultimately practical parameter space to consider for $\boldsymbol{\theta}$ is $\mathbb{R}^5$. Of course, our choices of Ω are far from being omnibus, possibly due to the parametrization we have chosen for $f_{\boldsymbol{\theta}}$. An open question is then:

> *What parametrization of the full bent-cable model, if any, would allow us to consider entirely unbounded regression and parameter domains?*

We do not attempt to answer this question here.

5.6.2 Some Bounds on $d_{\boldsymbol{\theta}_0,\boldsymbol{\theta}}(x)$

In this section, we provide some useful bounds for the distance between the candidate and underlying full bent cables. These bounds are used in the various proofs that appear in Section 5.5.

Since $d_i(\boldsymbol{\theta}_0) = 0$ for all i, a one-term Taylor expansion around $\boldsymbol{\theta}_0$ gives

$$d_i(\boldsymbol{\theta}) \;=\; d_i(\boldsymbol{\theta}) - d_i(\boldsymbol{\theta}_0) \;=\; (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \int_0^1 \boldsymbol{\nabla} d_i\big(\boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\big)\, dt \qquad (5.32)$$

$$\text{where} \qquad \boldsymbol{\nabla} d_i(\boldsymbol{\theta}) \;=\; \boldsymbol{\nabla} f_i(\boldsymbol{\theta}) \;=\; \left(\frac{\partial f_{\boldsymbol{\theta}}(x_i)}{\partial \beta_0}, \frac{\partial f_{\boldsymbol{\theta}}(x_i)}{\partial \beta_1}, \frac{\partial f_{\boldsymbol{\theta}}(x_i)}{\partial \beta_2}, \frac{\partial f_{\boldsymbol{\theta}}(x_i)}{\partial \tau}, \frac{\partial f_{\boldsymbol{\theta}}(x_i)}{\partial \gamma} \right)$$

$$\text{and} \qquad \boldsymbol{\nabla} f_i\big(\boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\big) \;=\; \boldsymbol{\nabla} f_i(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta} \,:=\, \boldsymbol{\theta}_0 \,+\, t(\boldsymbol{\theta} \,-\, \boldsymbol{\theta}_0)} \qquad \forall\, t \in [0,1],\ i = 1, \ldots, n\,.$$

By the Cauchy-Schwarz inequality, we have

$$\big|d_i(\boldsymbol{\theta})\big| \;\le\; \big|\boldsymbol{\theta} - \boldsymbol{\theta}_0\big| \int_0^1 \Big|\boldsymbol{\nabla} f_i\big(\boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\big)\Big|\, dt \qquad \forall\, i = 1, \ldots, n\,.$$

For each i, we wish to find a uniform bound for $|d_i(\boldsymbol{\theta})|$ as $\boldsymbol{\theta}$ is varied over $\Theta_r = \big\{\boldsymbol{\theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \le r\big\}$ for some positive radius, r. Given such $\boldsymbol{\theta}$ and $t \in [0,1]$, we have $\boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \in \Theta_r$ also. Hence, with an appropriately chosen B_i, we have

$$\big|\boldsymbol{\nabla} f_i(\boldsymbol{\theta})\big| \le B_i \quad \forall\, \boldsymbol{\theta} \in \Theta_r \quad \Longrightarrow \quad \sup_{\boldsymbol{\theta} \,\in\, \Theta_r} \big|d_i(\boldsymbol{\theta})\big| \le r\, B_i\,. \tag{5.33}$$

From Section 5.5.2 (pp. 86–89), $\boldsymbol{\theta} \in \Theta_r$ gives

$$\begin{aligned}
\big|\boldsymbol{\nabla} f_i(\boldsymbol{\theta})\big|^2 \;&=\; \sum_{j=1}^{5} \left(\frac{\partial f_i(\boldsymbol{\theta})}{\partial \theta_j} \right)^2 \\
&\le\; 1 + x_i^2 + \max\big\{ \big(|x_i| + |\tau|\big)^2, \gamma^2 \big\} + \beta_2^2 + \frac{\beta_2^2}{16} \\
&<\; 1 + x_i^2 + \max\big\{ \big(|x_i| + |\tau_0| + r\big)^2, (\gamma_0 + r)^2 \big\} + 2\big(|b_2| + r\big)^2\,.
\end{aligned}$$

This upper bound depends on r. However, the relevant values of r in Section 5.5 are all fixed and finite. Thus, for condition [E], we can assume without loss of generality, that $a \ge \max_j \big\{|\theta_j| + r, 1\big\}$ for all these values of r (see page 76, Footnote 2, Remark 3). Hence, we can take

$$B_i \;=\; \begin{cases} 3a & \text{if} \quad |x_i| \le a \\ 3|x_i| & \text{if} \quad |x_i| > a \end{cases} \tag{5.34}$$

in (5.33) to get

$$\sup_{\boldsymbol{\theta} \,\in\, \Theta_r} \big|d_i(\boldsymbol{\theta})\big| \;\le\; \begin{cases} 3ar & \text{if} \quad |x_i| \le a \\ 3|x_i| r & \text{if} \quad |x_i| > a \end{cases}\,. \tag{5.35}$$

Another quantity in Section 5.5 which we wish to bound is $d_{\boldsymbol{\theta}_k, \boldsymbol{\theta}}$, $k = 1, \ldots, R$, where both $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_k$ are in Θ_r, and $|\boldsymbol{\theta} - \boldsymbol{\theta}_k| \le \delta'$ for some small $\delta' > 0$. For any given

k, we expand, as we did in (5.32), $d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)$ about $\boldsymbol{\theta}_k$ and apply the Cauchy-Schwarz inequality to get

$$\left|d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)\right| \leq |\boldsymbol{\theta}-\boldsymbol{\theta}_k| \int_0^1 \left|\boldsymbol{\nabla} f\big(x_i;\, \boldsymbol{\theta}_k + t(\boldsymbol{\theta}-\boldsymbol{\theta}_k)\big)\right| dt \qquad \forall\, i, k\,.$$

Note that $\boldsymbol{\theta}_k + t(\boldsymbol{\theta}-\boldsymbol{\theta}_k) \in \Theta_r$ for all $\boldsymbol{\theta}, \boldsymbol{\theta}_k \in \Theta_r$ and $t \in [0,1]$. Thus, for all i, the B_i in (5.34) is an upper bound for $\boldsymbol{\nabla} f\big(x_i;\, \boldsymbol{\theta}_k + t(\boldsymbol{\theta}-\boldsymbol{\theta}_k)\big)$, uniformly for $\boldsymbol{\theta}, t$, and k. Hence, for all k such that $\boldsymbol{\theta}_k \in \Theta_r$, we have

$$\sup_{\boldsymbol{\theta}\,\in\,\Theta_r\,:\,|\boldsymbol{\theta}-\boldsymbol{\theta}_k|\,\leq\,\delta'} \left|d_{\boldsymbol{\theta}_k,\boldsymbol{\theta}}(x_i)\right| \leq \begin{cases} 3a\delta' & \text{if} \quad |x_i| \leq a \\ 3|x_i|\delta' & \text{if} \quad |x_i| > a \end{cases} . \tag{5.36}$$

5.6.3 Some Notation for the Taylor-type Expansion

Let $\boldsymbol{\theta}_{(-j)} = (\theta_1,\, \theta_2,\, \ldots,\, \theta_j,\, \theta_{0,j+1},\, \ldots,\, \theta_{05})$ for $j = 1, \ldots, 5$, and take $\boldsymbol{\theta}_{(-0)} = \boldsymbol{\theta}_0$. That is, the first j elements of $\boldsymbol{\theta}_{(-j)}$ are from the parameter candidate, $\boldsymbol{\theta}$, while the last $5-j$ elements are from the true parameter, $\boldsymbol{\theta}_0$. For example, $\boldsymbol{\theta}_{(-2)} = (\beta_0, \beta_1, b_2, \tau_0, \gamma_0)$. Note that for each j, the vectors $\boldsymbol{\theta}_{(-j)}$ and $\boldsymbol{\theta}_{(-(j-1))}$ differ only by their j-th arguments, which are θ_j and θ_{0j}, respectively.

Now, suppress the dependence on σ^2 in the notation, and write $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}) = \big[U_{n\beta_0}(\boldsymbol{\theta}),\, U_{n\beta_1}(\boldsymbol{\theta}),\, \ldots,\, U_{n\gamma}(\boldsymbol{\theta})\big]$. Then, the k-th element of the $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta})$ vector is

$$U_{nk}(\boldsymbol{\theta}) = U_{nk}(\boldsymbol{\theta}_0) + \sum_{j=1}^{5} \Big[U_{nk}\big(\boldsymbol{\theta}_{(-j)}\big) - U_{nk}\big(\boldsymbol{\theta}_{(-(j-1))}\big) \Big] \qquad \forall\ k = \beta_0, \beta_1, \beta_2, \tau, \gamma \tag{5.37}$$

because the summands cancel.

Next, define the vector-valued function $\boldsymbol{\psi}(\cdot,\cdot,\cdot)$ as

$$\boldsymbol{\psi}(\boldsymbol{\theta}, t, j) = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_{j-1} \\ \theta_{0j} + t(\theta_j - \theta_{0j}) \\ \theta_{0,j+1} \\ \vdots \\ \theta_{05} \end{bmatrix} \qquad \forall\ t \in [0,1] \quad j = 1, \ldots, 5\,, \tag{5.38}$$

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

$$\boldsymbol{V}^+_{n,j\boldsymbol{\cdot}}(\boldsymbol{\theta}) = \begin{bmatrix} V^+_{n,\,j\beta_0}(\boldsymbol{\theta}) \\ V^+_{n,\,j\beta_1}(\boldsymbol{\theta}) \\ \vdots \\ V^+_{n,\,j\gamma}(\boldsymbol{\theta}) \end{bmatrix}^T \qquad \text{where } V^+_{n,jk}(\boldsymbol{\theta}) = \sum_{i=1}^{n} V^+_{jk,i}(\boldsymbol{\theta}) \qquad \forall\ k = \beta_0, \beta_1, \beta_2, \tau, \gamma\,.$$

Again, note the suppressed dependence on σ^2 in the notation.

Using (5.38), we stack the row vectors, $\boldsymbol{V}^+_{n,j\boldsymbol{\cdot}}\big(\boldsymbol{\psi}(\boldsymbol{\theta}, t, j)\big)$'s, to form the $\mathbb{V}^*_{n,\sigma^2}$ matrix in the Taylor-type expansion of Lemma 5.2. That is,

$$\mathbb{V}^*_{n,\sigma^2}(\boldsymbol{\theta}, t) = \begin{bmatrix} \boldsymbol{V}^+_{n,\,\beta_0\boldsymbol{\cdot}}\big(\boldsymbol{\psi}(\boldsymbol{\theta}, t, 1)\big) \\ \boldsymbol{V}^+_{n,\,\beta_1\boldsymbol{\cdot}}\big(\boldsymbol{\psi}(\boldsymbol{\theta}, t, 2)\big) \\ \cdots \\ \boldsymbol{V}^+_{n,\,\gamma\boldsymbol{\cdot}}\big(\boldsymbol{\psi}(\boldsymbol{\theta}, t, 5)\big) \end{bmatrix}. \tag{5.39}$$

5.6.4 Lemma Proofs

With the framework already provided through the basic case in Chapter 4, we avoid redundancy by not showing full details for all the proofs of this chapter. Any notational modifications and ingredients absent in the previous proofs will be carefully explained. However, in order to follow the mathematical logic presented below, it is important for the reader to have a clear idea of the corresponding proofs in Chapter 4.

Proof of Lemma 5.1

Without loss of generality, take $b_0 = b_1 = 0$, $b_2 = 1$, and $\tau_0 = 0$. That is, $f_{\boldsymbol{\theta}_0}$ is a basic bent cable centered at 0. We consider a candidate full cable, $f_{\boldsymbol{\theta}}$, which may go through the given (w_i, z_i)'s in any of twenty-one possible five-point configurations.

Excluding the $\{2, 1, 2\}$-configuration itself, let us partition the remaining twenty configurations for $f_{\boldsymbol{\theta}}$ into the following sets:

$$
\begin{aligned}
\text{Set } \mathcal{A} &= \big\{\{5,0,0\},\{4,0,1\},\{4,1,0\},\{3,2,0\},\{3,1,1\},\{3,0,2\}\big\} \\
\text{Set } \mathcal{B} &= \big\{\{1,3,1\},\{0,5,0\},\{0,4,1\}\big\} \\
\text{Set } \mathcal{C} &= \big\{\{2,3,0\},\{2,2,1\}\big\} \\
\text{Set } \mathcal{D} &= \big\{\text{collection of all mirror configurations of Sets } \mathcal{A} \text{ to } \mathcal{C} : \text{“mirror”=}y\text{-axis}\big\} \\
&= \big\{\{0,0,5\},\{1,0,4\},\ldots,\{2,0,3\},\{1,0,4\},\{0,3,2\},\{1,2,2\}\big\}
\end{aligned}
$$

We show that for the given (w_i, z_i)'s to lie on $f_{\boldsymbol{\theta}}$, it is impossible that they be in a configuration of any of Sets $\mathcal{A}$ to $\mathcal{D}$.

First, each member of Set $\mathcal{A}$ has at least three neighboring w_i's in the incoming linear phase of $f_{\boldsymbol{\theta}}$. Of course, collinearity of more than two consecutive w_i's contradicts the true $\{2,1,2\}$-configuration generated by $f_{\boldsymbol{\theta}_0}$. This proves that Set $\mathcal{A}$ is impossible for the candidate cable.

Next, a Set $\mathcal{B}$ configuration features at least three neighboring w_i's in the bend of $f_{\boldsymbol{\theta}}$, i.e. $\tau - \gamma < w_{-1} < -\gamma_0 < w_0 < \gamma_0 < w_1 < \tau + \gamma$. As (w_{-2}, z_{-2}) and (w_{-1}, z_{-1}) lie on the slope-0 phase of $f_{\boldsymbol{\theta}_0}$, convexity of $f_{\boldsymbol{\theta}}$ forces its slope at w_{-1} to be non-negative. Similarly, (w_1, z_1) and (w_2, z_2) form a slope of 1, thus forcing the slope of $f_{\boldsymbol{\theta}}$ at w_1 to be no more than 1. Therefore,

$$f'_{\boldsymbol{\theta}}(w_{-1}) \geq 0 \ , \qquad f'_{\boldsymbol{\theta}}(w_1) \leq 1 \, . \tag{5.40}$$

To show that Set $\mathcal{B}$ is impossible for $f_{\boldsymbol{\theta}}$, we will show that

$$f'_{\boldsymbol{\theta}}(w_1) - f'_{\boldsymbol{\theta}}(w_{-1}) > 1 \tag{5.41}$$

so that at least one of the inequalities in (5.40) is invalid.

Let $g(w) = \alpha_0 + \alpha_1 w + \alpha_2 w^2$ be the unique quadratic through (w_i, z_i) for $i = 0, \pm 1$. The secants of g between pairs of neighboring w_i's have slopes:

$$\frac{g(w_{i+1}) - g(w_i)}{w_{i+1} - w_i} = \frac{\alpha_1(w_{i+1} - w_i) + \alpha_2(w_{i+1}^2 - w_i^2)}{w_{i+1} - w_i} = \alpha_1 + \alpha_2(w_i + w_{i+1}) \, .$$

In particular, for $i = 0, \pm 1$, we have

$$
\begin{aligned}
\alpha_1 + \alpha_2(w_{-1} - w_0) &= \frac{z_0 - z_{-1}}{w_0 - w_{-1}} = \frac{z_0}{w_0 - w_{-1}} \\
\alpha_1 + \alpha_2(w_0 - w_1) &= \frac{z_1 - z_0}{w_1 - w_0} = \frac{w_1 - z_0}{w_1 - w_0}
\end{aligned}
$$

where $z_0 = (w_0 - \tau_0 + \gamma_0)^2/(4\gamma_0) = (w_0 + \gamma_0)^2/(4\gamma_0)$. Therefore,

$$\begin{aligned}
f'_{\boldsymbol{\theta}}(w_1) - f'_{\boldsymbol{\theta}}(w_{-1}) &= g'(w_1) - g'(w_{-1}) \\
&= 2\alpha_2(w_1 - w_{-1}) \;=\; 2\big[\alpha_2(w_0 + w_1) - \alpha_2(w_{-1} + w_0)\big] \\
&= 2\,\psi(w_0) \\
\text{where} \qquad \psi(x) &= \frac{w_1 - z_0}{w_1 - x} - \frac{z_0}{x - w_{-1}}\,, \qquad x \in [-\gamma_0, \gamma_0]\,.
\end{aligned}$$

Thus, showing (5.41) is equivalent to showing $\psi(w_0) > 1/2$.

One can show that the first derivative, ψ', has roots

$$r_1 = \frac{\gamma_0(w_{-1} + w_1)}{2\gamma_0 - (w_1 - w_{-1})} \qquad \text{and} \qquad r_2 = \frac{\gamma_0(w_1 - w_{-1}) + 2w_{-1}\,w_1}{w_{-1} + w_1}\,.$$

That is, $\psi(r_1)$ and $\psi(r_2)$ are local optima. After some lengthy algebra, we have

$$\psi''(r_1) = \frac{-[2\gamma_0 - (w_1 - w_{-1})]^4}{2(w_1 - w_{-1})^3\,(|w_{-1}| - \gamma_0)\,(x_1 - \gamma_0)\,\gamma_0} < 0 \tag{5.42}$$

$$\psi''(r_2) = \frac{(w_{-1} + w_1)^4}{2(w_1 - w_{-1})^3\,(|w_{-1}| - \gamma_0)\,(w_1 - \gamma_0)\,\gamma_0} > 0 \tag{5.43}$$

$$\psi(r_2) = \frac{|w_{-1}|\,w_1}{(w_1 - w_{-1})\gamma_0} > \frac{\gamma_0^2}{(2\gamma_0)\gamma_0} = \frac{1}{2} \tag{5.44}$$

$$\psi(-\gamma_0) = \frac{1}{1 + \gamma_0/w_1} > \frac{1}{1+1} = \frac{1}{2} \tag{5.45}$$

$$\psi(\gamma_0) = \frac{|w_{-1}|}{(|w_{-1}| + \gamma_0)} > \frac{1}{2} \tag{5.46}$$

From (5.42) and (5.43), we see that r_1 is a local maximizer, while r_2 is a local minimizer. Then, (5.44) to (5.46) imply $\psi(x) > 1/2$ for all $x \in [-\gamma_0, \gamma_0]$. In particular,

$$w_0 \in (-\gamma_0, \gamma_0) \implies \psi(w_0) > \frac{1}{2}\,.$$

This establishes (5.41). Hence, Set $\mathcal{B}$ is impossible for $f_{\boldsymbol{\theta}}$.

To see that Set $\mathcal{C}$ is also impossible for $f_{\boldsymbol{\theta}}$, apply the basic-case argument (from Chapter 4) which begins with relation (4.45) on page 57, and stretches to the full paragraph on page 58 before (4.49), except with q replaced by f. The entire argument there is valid for a Set $\mathcal{C}$ configuration here, since $w_0 < \gamma_0 < w_1$.

Finally, a symmetry argument shows that Set $\mathcal{D}$ is also impossible for $f_{\boldsymbol{\theta}}$. More specifically, denote the mirror image of $f_{\boldsymbol{\theta}}$ by $f^*_{\boldsymbol{\theta}}$ for any given $\boldsymbol{\theta}$. Then, the class of all $f_{\boldsymbol{\theta}_0}$ cables with a $\{2,1,2\}$-configuration is equivalent to the class of all $f^*_{\boldsymbol{\theta}_0}$ cables with a $\{2,1,2\}$-configuration. Now, consider $f_{\boldsymbol{\theta}}$ with an $\{s,t,u\}$-configuration from any of Sets $\mathcal{A}$ to $\mathcal{C}$, satisfying $f_{\boldsymbol{\theta}}(w_i) = f_{\boldsymbol{\theta}_0}(w_i)$ for all $i = 0, \pm 1, \pm 2$. Thus, the mirror cable, $f^*_{\boldsymbol{\theta}}$, has a $\{u,t,s\}$-configuration from Set $\mathcal{D}$, and it satisfies $f^*_{\boldsymbol{\theta}}(w^*_i) = f^*_{\boldsymbol{\theta}_0}(w^*_i)$ $(w^*_i = -w_i)$ for all i. By an equivalence class argument, the fact that $\{s,t,u\}$ is impossible for $f_{\boldsymbol{\theta}}$ implies that $\{u,t,s\}$ is impossible for $f^*_{\boldsymbol{\theta}}$. In other words, Set $\mathcal{D}$ is impossible for any candidate cable when the given (w_i, z_i)'s from the true cable are in a $\{2,1,2\}$-configuration.

Having ruled out all other configurations, we can easily argue for the uniqueness of a full cable with a $\{2,1,2\}$-configuration. The points (w_{-2}, z_{-2}) and (w_{-1}, z_{-1}) form a 0 slope for the incoming phase of $f_{\boldsymbol{\theta}}$, forcing $\beta_0 = b_0 = 0 = b_1 = \beta_1$. Similarly, $\beta_2 = b_2 = 1$ in order for the outgoing phase of $f_{\boldsymbol{\theta}}$ to go through (w_1, z_1) and (w_2, z_2). The two linear phases intersect at $\tau = \tau_0 = 0$. Lastly, the relations

$$\frac{(w_0+\gamma)^2}{4\gamma} = f_{\boldsymbol{\theta}}(w_0) = f_{\boldsymbol{\theta}_0}(w_0) = \frac{(w_0+\gamma_0)^2}{4\gamma_0}$$

and

$$-\gamma,\ -\gamma_0 < x_0 < \gamma_0,\ \gamma$$

uniquely determine that $\gamma = \gamma_0$. ■

Proof of Lemma 5.2

Similar to (4.55) from the basic case on page 61, we integrate $V^+_{n,jk}$ over θ_j to retrieve U_{nk}. Recall relation (5.38) on page 107. Then, for each k, summing over j gives

$$\begin{aligned}\sum_{j=1}^{5}\Big[U_{nk}\big(\boldsymbol{\theta}_{(-j)}\big) - U_{nk}\big(\boldsymbol{\theta}_{(-(j-1))}\big)\Big] &= \sum_{j=1}^{5}(\theta_j-\theta_{0j})\int_0^1 V^+_{n,jk}\big(\boldsymbol{\psi}(\boldsymbol{\theta},t,j)\big)\,dt \\ &= \left(\int_0^1 \begin{bmatrix} V^+_{n,\beta_0 k}\big(\boldsymbol{\psi}(\boldsymbol{\theta},t,1)\big) \\ V^+_{n,\beta_1 k}\big(\boldsymbol{\psi}(\boldsymbol{\theta},t,2)\big) \\ V^+_{n,\beta_2 k}\big(\boldsymbol{\psi}(\boldsymbol{\theta},t,3)\big) \\ V^+_{n,\tau k}\big(\boldsymbol{\psi}(\boldsymbol{\theta},t,4)\big) \\ V^+_{n,\gamma k}\big(\boldsymbol{\psi}(\boldsymbol{\theta},t,5)\big) \end{bmatrix} dt\right)^T (\boldsymbol{\theta}-\boldsymbol{\theta}_0)\,.\end{aligned}$$

Substitute this into (5.37), then stack the U_{nk}'s to establish the lemma. ■

Proof of Lemma 5.3

To prove the lemma, it suffices to prove relations (5.15) and (5.16), i.e. we prove that $\eta_{n,jk}(\boldsymbol{\theta})$ $\big(= I_{n,jk}(\boldsymbol{\theta}) - I_{n,jk}(\boldsymbol{\theta}_0)\big)$ and $\Delta_{n,jk}(\boldsymbol{\theta})$ $\big(= V^{+}_{n,jk}(\boldsymbol{\theta}) + I_{n,jk}(\boldsymbol{\theta})\big)$ have uniform sizes $o(n)$ and $o_p(n)$, respectively, over Θ_{δ_n} for any $\delta_n \downarrow 0$. To this end, we extend the proof of Lemma 4.3 (pp. 62–67) from the basic case to the five-parameter case here.

From Section 5.5.2 (pp. 86–89), we see that some (j,k)-pairs are irrelevant to (5.15) or (5.16). In particular, for $j,k = \beta_0, \beta_1$, the matrix $\mathbb{I}_{n,jk}$ is $\boldsymbol{\theta}$-free, and hence, $\eta_{n,jk}(\boldsymbol{\theta}) \equiv 0$ for all $\boldsymbol{\theta}$. Furthermore, by (5.14) there, $\Delta_{n,jk}(\boldsymbol{\theta}) \equiv 0$ for all $\boldsymbol{\theta}$ for those (j,k)'s which correspond to the first two rows and columns, as well as the (β_2,β_2)-th element, of $\mathbb{I}_{n,jk}$.

For the remaining (j,k)-pairs, we first consider the $\eta_{n,jk}$'s and write

$$\sigma^2\, I_{jk,i}(\boldsymbol{\theta}) = (-1)^p\, \beta_2^{p_0} \big\{ z\,(\alpha_{1i}\, \alpha_{4i}^{p_1}) + \gamma^{p_2}\, \alpha_{2i}^{p_3}\, \alpha_{3i}^{p_4} \big\}\, x_i^{p_5}$$

where $p, p_5, z \in \{0,1\}$, $p_0, p_1, p_2, p_4 \in \{0,1,2\}$, and $p_3 \in \{0,1,2,3,4\}$. In the case of $z = 0$, we have

$$\begin{aligned} \sigma^2\, I_{jk,i}(\boldsymbol{\theta}) &= x_i^{p_5}\, g_i(\boldsymbol{\theta})\, \mathbf{1}\big\{|x_i - \tau| \le \gamma\big\} \\ \text{where} \qquad g_i(\boldsymbol{\theta}) &= (-1)^p\, \beta_2^{p_0}\, \gamma^{p_2} \left[\frac{x_i - (\tau - \gamma)}{2\gamma}\right]^{p_3} \left[\frac{1}{4}\left(1 - \left|\frac{x_i - \tau}{\gamma}\right|^2\right)\right]^{p_4}. \end{aligned}$$

Note that each g_i is infinitely-many-times differentiable with respect to $\boldsymbol{\theta}$. Since Θ_{δ_n} is compact and ever-decreasing in size, there is $M_n = O(1)$ such that $|g_i(\boldsymbol{\theta})|, |\boldsymbol{\nabla}\, g_i(\boldsymbol{\theta})| \le M_n$ uniformly for $\boldsymbol{\theta} \in \Theta_{\delta_n}$. Hence, the Cauchy-Schwarz inequality applied to a one-term Taylor expansion of g_i about $\boldsymbol{\theta}_0$ yields

$$\sup_{\boldsymbol{\theta}\,\in\,\Theta_{\delta_n}} \big|g_i(\boldsymbol{\theta}) - g_i(\boldsymbol{\theta}_0)\big| \le \sup_{\boldsymbol{\theta}\,\in\,\Theta_{\delta_n}} \big|\boldsymbol{\theta} - \boldsymbol{\theta}_0\big| \int_0^1 \Big|\boldsymbol{\nabla}\, g_i\big(\boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\big)\Big|\, dt \le \delta_n M_n = o(1)\,.$$

For all $\boldsymbol{\theta} \in \Theta_{\delta_n}$, apply this upper bound to get

$$
\begin{aligned}
\sigma^2\big|\eta_{n,jk}(\boldsymbol{\theta})\big| \;&=\; \sigma^2\sum_i\big|I_{jk,i}(\boldsymbol{\theta})-I_{jk,i}(\boldsymbol{\theta}_0)\big| \\
=\;& \sum_i\Big|x_i^{p_5}\Big|\,\Big|g_i(\boldsymbol{\theta})\,\mathbf{1}\big\{|x_i-\tau|\le\gamma\big\}-g_i(\boldsymbol{\theta}_0)\,\mathbf{1}\big\{|x_i-\tau_0|\le\gamma_0\big\}\Big| \\
\le\;& \sum_i\big|x_i\big|^{p_5}\Big[\big|g_i(\boldsymbol{\theta})-g_i(\boldsymbol{\theta}_0)\big|\,\mathbf{1}\big\{|x_i-\tau|\le\gamma,|x_i-\tau_0|\le\gamma_0\big\}+ \\
& \big|g_i(\boldsymbol{\theta})\big|\,\mathbf{1}\big\{|x_i-\tau|\le\gamma,|x_i-\tau_0|>\gamma_0\big\}+\big|g_i(\boldsymbol{\theta}_0)\big|\,\mathbf{1}\big\{|x_i-\tau|>\gamma,|x_i-\tau_0|\le\gamma_0\big\}\Big] \\
\le\;& \delta_nM_n\left(\sum_{|x_i|\le a}\big|x_i\big|^{p_5}+\sum_{|x_i|>a}\big|x_i\big|^{p_5}\right)+M_n\left(\sum_{\substack{x_i\,\in\,K_n^{\pm}\\ |x_i|\,\le\,a}}\big|x_i\big|^{p_5}+\sum_{\substack{x_i\,\in\,K_n^{\pm}\\ |x_i|\,>\,a}}\big|x_i\big|^{p_5}\right)
\end{aligned}
$$

where $K_n^+=[\tau_0+\gamma_0-\delta_n,\tau_0+\gamma_0+\delta_n]$ and $K_n^-=[\tau_0-\gamma_0-\delta_n,\tau_0-\gamma_0+\delta_n]$ (as in the proof of Lemma 4.3 from the Chapter-4 Appendix), and a, from condition [E]. Without loss of generality, assume $a>|\tau_0|+\gamma_0+1$ (see page 76, Footnote 2, Remark 3). Now, based on condition [E], choose N such that for all $n>N$, we have $n^{-1}\sum_{|x_i|>a}|x_i|^{p_5}\le\kappa_{p_5}$ for $p_5=0,1$. Then, by condition [C], there is an $N_1\ge N$ such that for all $n>N_1$, the intersection $K_n^{\pm}\cap\{x:|x|>a\}$ is the empty set. Hence, $n>N_1$ implies

$$
\begin{aligned}
\sup_{\boldsymbol{\theta}\,\in\,\Theta_{\delta_n}}\sigma^2\big|\eta_{n,jk}(\boldsymbol{\theta})\big| \;&\le\; n\Big\{\delta_nM_n\big[a^{p_5}+\kappa_{p_5}\big]+M_n\big[a^{p_5}\,\zeta(\delta_n)+0\big]\Big\} \\
&=\; n\big\{o(1)\,O(1)+O(1)\,o(1)\big\} \;=\; o(n)\,.
\end{aligned}
$$

We have thus established (5.15) for $z=0$. The algebra is similar for $z=1$, and is omitted here.[7]

It remains to consider the $\Delta_{n,jk}$'s for all $(j,k)=(\beta_2,\tau),(\beta_2,\gamma),(\tau,\tau),(\tau,\gamma),(\gamma,\tau),(\gamma,\gamma)$.

First, write $\sigma^2\,\Delta_{n,\,jk}(\boldsymbol{\theta})=a_{n,jk}+B_{n,jk}(\boldsymbol{\theta})$, where $a_{n,jk}(\boldsymbol{\theta})=\sum\omega_i(\boldsymbol{\theta})\,d_i(\boldsymbol{\theta})$ and $B_{n,jk}(\boldsymbol{\theta})=\sum\omega_i(\boldsymbol{\theta})\,\varepsilon_i$ for some weights ω_i's. Each ω_i is a function of the α_{ki}'s, $k=1,\ldots,4$. Due to the indicator(s), each $\omega_i(\boldsymbol{\theta})$ has a uniform bound of order

[7] For $z=1$, it impossible that $p_1=2$ and $p_5=1$ simultaneously. Thus, each x_i has a total power of p_1+p_5 that does not exceed 2, and condition [E] suffices to control the uniform size of $\eta_{n,jk}$.

$O(1)$ over Θ_{δ_n}. Then, conditions [C] and [E] can be applied, together with (5.35) of Section 5.6.2, to show that $a_{n,jk}(\boldsymbol{\theta}) = o(n)$ and $B_{n,jk}(\boldsymbol{\theta}) = o_p(n)$ uniformly over Θ_{δ_n}. In particular, depending on the (j,k)-pair, either the martingale tactic or the chaining argument from the proof of Lemma 4.3 of the basic case can be applied here to show $B_{n,jk}(\boldsymbol{\theta}) = o_p(n)$. As the tedious algebra is routine for all (j,k)-pairs, we will only discuss the case of $(j,k) = (\beta_2, \tau)$ below, but omit the others.

By (5.35),

$$\sup_{\boldsymbol{\theta} \,\in\, \Theta_{\delta_n}} |d_i(\boldsymbol{\theta})| \le 3\delta_n \Big[a\, \mathbf{1}\big\{|x_i| \le a\big\} + |x_i|\, \mathbf{1}\big\{|x_i| > a\big\} \Big] \qquad \forall\ i = 1, \ldots, n\,.$$

Now, $a_{n,\,\beta_2\tau}(\boldsymbol{\theta}) = -\sum(\alpha_{1i} + \alpha_{2i})\, d_i(\boldsymbol{\theta})$ and $B_{n,\,\beta_2\tau}(\boldsymbol{\theta}) = -\sum(\alpha_{1i} + \alpha_{2i})\, \varepsilon_i$. Note that some d_i's may not be uniformly bounded over Θ_{δ_n} due to possibly unbounded x_i's. Nonetheless, since $|\alpha_{1i}|, |\alpha_{2i}| \le 1$ for all $\boldsymbol{\theta}$, then by condition [E], there is N such that

$$n > N \implies \sup_{\boldsymbol{\theta} \,\in\, \Theta_{\delta_n}} \big| a_{n,\,\beta_2\tau}(\boldsymbol{\theta}) \big| \le 6n\delta_n(a + \kappa_1) = o(n)\,.$$

Furthermore, when condition [E] is included in a chaining argument similar to that for (4.76) of the basic case (p. 66), one can see that $B_{n,\,\beta_2\tau}(\boldsymbol{\theta})$ also has order $o_p(n)$ uniformly over Θ_{δ_n}. (We omit the tedious algebra.) This establishes (5.16) for $(j,k) = (\beta_2, \tau)$. The remaining (j,k)-pairs can be examined in a similar fashion to produce a uniform bound of order $o_p(n)$ for $\Delta_{n,jk}$. ■

Proof of Lemma 5.4

First, consider $\boldsymbol{\theta}_i = (\beta_{0i}, \beta_{1i}, \beta_{2i}, \tau_i, \gamma_i)$, $i = 1, 2$, which do not both lie on any hyper-ray, $R_{k\pm}$. Due to piecewise twice-differentiability of ℓ_{n,σ^2} with respect to $\boldsymbol{\theta}$, h''_{n,σ^2} is undefined if and only if the Hessian $\mathbb{V}_{n,\sigma^2}$ is undefined. Thus, h''_{n,σ^2} is undefined at only those t's such that $\boldsymbol{\theta}_1 + t(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1) \in R_{k\pm}$, $k = 1, \ldots, n$, i.e. at

$$t_{k\pm} = \frac{\pm(x_k - \tau_1) - \gamma_1}{(\gamma_2 - \gamma_1) \pm (\tau_2 - \tau_1)}\,, \qquad k = 1, \ldots, n\,. \tag{5.47}$$

Note that these $t_{k\pm}$'s correspond to the intersections of the hyper-rays, $R_{k\pm}$'s, and are the only t's at which the Hessian, $\mathbb{V}_{n,\sigma^2}\big(\boldsymbol{\theta}_1 + t(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)\big)$, is discontinuous. Of course, the set $\mathcal{N} \equiv \{t_{1\pm}, \ldots, t_{n\pm}\}$ contains isolated values in $[0,1]$. At any other t, the Hessian is well defined and coincides with its directional counterpart, $\mathbb{V}^+_{n,\sigma^2}\big(\boldsymbol{\theta}_1 + t(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)\big)$. Therefore, by the chain rule, (5.17) is true for all $t \in [0,1] \setminus \mathcal{N}$.

If $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in R_{k\pm}$ for some k, then $\boldsymbol{\theta}_1 + t(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1) \in R_{k\pm}$ also for all $t \in [0,1]$. By symmetry, it suffices to consider R_{k+} only. Although the chain rule is inapplicable here, we show that (5.17) is algebraically valid for $t \notin \mathcal{N}$.

Given $t \notin \mathcal{N}$, write $\boldsymbol{\theta}' = \boldsymbol{\theta}_1 + t(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)$. Since $\boldsymbol{\theta}' \in R_{k+}$, we have $\boldsymbol{\theta}' = (\beta_0', \beta_1', \beta_2', \tau', \tau' - x_k)$. Then,

$$
\begin{aligned}
h_{n,\sigma^2}(t) &= -\frac{1}{2\sigma^2} S_n(\boldsymbol{\theta}') \;=\; -\frac{1}{2\sigma^2}\sum_i \big\{\varepsilon_i + b_i(t)\big\} \\
\text{where} \qquad b_i(t) &= f_{\boldsymbol{\theta}_0}(x_i) - \beta_0' - \beta_1' x_i - \beta_2'\, q_i(t)\,, \\
q_i(t) &= \frac{(x_i - x_k)^2}{4(\tau' - x_k)}\,\mathbf{1}\big\{x_k \le x_i \le 2\tau' - x_k\big\} + (x_i - \tau')\,\mathbf{1}\big\{x_i > 2\tau' - x_k\big\}\,.
\end{aligned}
$$

Differentiating h_{n,σ^2} twice with respect to t yields

$$
\begin{aligned}
h''_{n,\sigma^2}(t) &= -\frac{1}{\sigma^2}\sum_i \big[A_i(t) - B_i(t)\big]^2 + \frac{1}{\sigma^2}\sum_i \big[\varepsilon_i + b_i(t)\big]\big[-2C_i(t) + D_i(t)\big] \qquad (5.48) \\
\text{where} \quad A_i(t) &= (\beta_{02} - \beta_{01}) + (\beta_{12} - \beta_{11})x_i + (\beta_{22} - \beta_{21})\, q_i(t)\,, \\
B_i(t) &= \beta_2'(\tau_2 - \tau_1)\, r_i(t)\,, \\
C_i(t) &= (\beta_{22} - \beta_{21})(\tau_2 - \tau_1)\, r_i(t)\,, \\
r_i(t) &= \frac{(x_i - x_k)^2}{4(\tau' - x_k)^2}\,\mathbf{1}\big\{x_k \le x_i \le 2\tau' - x_k\big\} + \mathbf{1}\big\{x_i > 2\tau' - x_k\big\}\,, \\
D_i(t) &= \beta_2'(\tau_2 - \tau_1)^2\, s_i(t)\,, \\
s_i(t) &= \frac{(x_i - x_k)^2}{2(\tau' - x_k)^3}\,\mathbf{1}\big\{x_k \le x_i \le 2\tau' - x_k\big\}\,.
\end{aligned}
$$

To show that $(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)^T\big[\mathbb{V}^+_{n,\sigma^2}(\boldsymbol{\theta}')\big](\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)$ expands out to the RHS of (5.48), we partition $(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)$ and $\mathbb{V}^+_{n,\sigma^2}$ as follows.

Write $\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1 = (\boldsymbol{\beta}^*, \boldsymbol{\tau}^*)$, where

$$
\boldsymbol{\beta}^* = \begin{bmatrix} \beta_{02} - \beta_{01} \\ \beta_{12} - \beta_{11} \\ \beta_{22} - \beta_{21} \end{bmatrix}, \qquad \boldsymbol{\tau}^* = \begin{bmatrix} \tau_2 - \tau_1 \\ \tau_2 - \tau_1 \end{bmatrix} = (\tau_2 - \tau_1)\begin{bmatrix} 1 \\ 1 \end{bmatrix} = (\tau_2 - \tau_1)\,\mathbf{1}\,.
$$

Next, denote the i-th summand of $\mathbb{V}^+_{n,\sigma^2}$ by $\mathbb{V}^{(i)+}$, and write

$$\sigma^2\,\mathbb{V}^{(i)+}(\boldsymbol{\theta}') = \begin{bmatrix} \mathbb{A}_i(\boldsymbol{\theta}') & \mathbb{B}_i(\boldsymbol{\theta}') \\ \mathbb{C}_i(\boldsymbol{\theta}') & \mathbb{D}_i(\boldsymbol{\theta}') \end{bmatrix}$$

where the dimensions of $\mathbb{A}_i$ to $\mathbb{D}_i$ are, respectively, 3×3, 3×2, 2×3, and 2×2. Note that by (5.14) on page 88, $\mathbb{B}_i^T = \mathbb{C}_i$. Then,

$$\begin{aligned} & \sigma^2\,(\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)^T\big[\mathbb{V}^{(i)+}(\boldsymbol{\theta}')\big](\boldsymbol{\theta}_2-\boldsymbol{\theta}_1) \\ = \;& \boldsymbol{\beta}^{*^T}\big[\mathbb{A}_i(\boldsymbol{\theta}')\big]\boldsymbol{\beta}^* + 2(\tau_2-\tau_1)\boldsymbol{\beta}^{*^T}\big[\mathbb{B}_i(\boldsymbol{\theta}')\big]\,\mathbf{1} + (\tau_2-\tau_1)^2\,\mathbf{1}^T\big[\mathbb{D}_i(\boldsymbol{\theta}')\big]\,\mathbf{1}\,. \end{aligned} \tag{5.49}$$

We expand each of the three terms in (5.49). First, by (5.14),

$$\boldsymbol{\beta}^{*^T}\big[\mathbb{A}_i(\boldsymbol{\theta}')\big]\boldsymbol{\beta}^* = -\boldsymbol{\beta}^{*^T}\begin{bmatrix} 1 & x_i & q_i(t) \\ x_i & x_i^2 & x_i\,q_i(t) \\ x_i^2 & x_i\,q_i(t) & \big[q_i(t)\big]^2 \end{bmatrix}\boldsymbol{\beta}^* = -\big[A_i(t)\big]^2\,.$$

Next, for $j = 1,2,3,4$, let α'_{ji} denote the value of α_{ji} evaluated at $\boldsymbol{\theta}'$ instead of $\boldsymbol{\theta}$. Then,

$$\begin{aligned} & (\tau_2-\tau_1)^2\,\mathbf{1}^T\big[\mathbb{D}_i(\boldsymbol{\theta}')\big]\,\mathbf{1} \\ = \;& \sigma^2(\tau_2-\tau_1)^2\big\{V^+_{\tau\tau,i}(\boldsymbol{\theta}') + V^+_{\tau\gamma,i}(\boldsymbol{\theta}') + V^+_{\gamma\tau,i}(\boldsymbol{\theta}') + V^+_{\gamma\gamma,i}(\boldsymbol{\theta}')\big\} \\ = \;& \beta'_2(\tau_2-\tau_1)^2\times \\ & \Big\{ \quad -\beta'_2(\alpha'_{1i}+\alpha'^2_{2i}) + \big[\varepsilon_i+b_i(t)\big]\frac{\mathbf{1}\{x_k < x_i \le 2\tau'-x_k\}}{2(\tau'-x_k)} \\ & \quad +2\beta'_2\alpha'_{2i}\alpha'_{3i} + \big[\varepsilon_i+b_i(t)\big]\frac{\alpha'_{4i}}{2(\tau'-x_k)^2}\times \\ & \qquad \Big[\mathbf{1}\{x_k < x_i \le 2\tau'-x_k\} + \mathbf{1}\{x_k \le x_i \le 2\tau'-x_k\}\Big] \\ & \quad -\beta'_2\alpha'^2_{3i} + \big[\varepsilon_i+b_i(t)\big]\frac{\alpha'^2_{4i}}{2(\tau'-x_k)^3}\,\mathbf{1}\{x_k \le x_i \le 2\tau'-x_k\} \qquad \Big\}\,. \end{aligned}$$

After some lengthy algebra, the coefficient of $\big[\varepsilon_i+b_i(t)\big]$ within the outer braces ({ }) reduces to $s_i(t)$, and that of β'_2, to $-\big[r_i(t)\big]^2$. Assembling, we have

$$\begin{aligned} (\tau_2-\tau_1)^2\,\mathbf{1}^T\big[\mathbb{D}_i(\boldsymbol{\theta}')\big]\,\mathbf{1} &= \beta'_2(\tau_2-\tau_1)^2\Big\{-\beta'_2\big[r_i(t)\big]^2 + \big[\varepsilon_i+b_i(t)\big]\,s_i(t)\Big\} \\ &= -\Big\{\big[B_i(t)\big]^2 - \big[\varepsilon_i+b_i(t)\big]\,D_i(t)\Big\}\,. \end{aligned}$$

Comparing to the i-th summand in (5.48), it remains to show $(\tau_2 - \tau_1)\boldsymbol{\beta}^{*^T}\big[\mathbb{B}_i(\boldsymbol{\theta}')\big]\,\mathbf{1} = A_i(t)\,B_i(t) - C_i(t)$. First, write

$$
\begin{aligned}
\big[\mathbb{B}_i(\boldsymbol{\theta}')\big]\,\mathbf{1} &= \begin{bmatrix} v_{1i} \\ v_{2i} \\ v_{3i} \end{bmatrix} \\
\text{where} \qquad v_{1i} &= \beta_2'(\alpha_{1i}' + \alpha_{2i}' - \alpha_{3i}')\,, \\
v_{2i} &= v_{1i}\,x_i\,, \\
v_{3i} &= \beta_2'(\alpha_{1i}'\alpha_{4i}' + \gamma'\alpha_{2i}'^3 - \gamma'\alpha_{2i}'^2\alpha_{3i}') - \big[\varepsilon_i + b_i(t)\big](\alpha_{1i}' + \alpha_{2i}' - \alpha_{3i}')\,, \\
\gamma' &= \tau' - x_k\,.
\end{aligned}
$$

More tedious algebra yields

$$
v_{1i} = \beta_2'\,r_i(t)\ , \quad v_{2i} = \beta_2'\,x_i\,r_i(t)\ , \quad v_{3i} = \beta_2'\,q_i(t)\,r_i(t) - \big[\varepsilon_i + b_i(t)\big]r_i(t)\,.
$$

Now, it is easy to verify that $(\tau_2 - \tau_1)\boldsymbol{\beta}^{*^T}\big[\mathbb{B}_i(\boldsymbol{\theta}')\big]\,\mathbf{1} = A_i(t)\,B_i(t) - C_i(t)$. ■

Proof of Lemma 5.5

Dropping the dependence on $\boldsymbol{\theta}_0$ and σ^2 in the notation, denote the i-th summands of $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)$ and $\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)$ by $\boldsymbol{U}^{(i)}$ and $\mathbb{I}^{(i)}$, respectively. Recall from Section 5.5.2 that $\mathbb{I}^{(i)} = \mathrm{Cov}\big[\boldsymbol{U}^{(i)}\big] = \sigma^{-2}\big[\boldsymbol{\nabla} f_i\big|_{\boldsymbol{\theta_0}}\big]\big[\boldsymbol{\nabla} f_i\big|_{\boldsymbol{\theta_0}}\big]^T$.

Next, define for all $\boldsymbol{w} \in \mathbb{R}^5$

$$
\begin{aligned}
\sigma_i^2(\boldsymbol{w}) &= \boldsymbol{w}^T\,\mathbb{I}^{(i)}\,\boldsymbol{w} = \left(\boldsymbol{w} \cdot \frac{1}{\sigma}\boldsymbol{\nabla} f_i\Big|_{\boldsymbol{\theta_0}}\right)^2 \\
B_n(\boldsymbol{w}) &= \sum_{i=1}^{n} \sigma_i^2(\boldsymbol{w}) = \boldsymbol{w}^T\big[\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)\big]\boldsymbol{w}
\end{aligned}
$$

For the zero-mean multivariate sequence $\big\{\boldsymbol{U}^{(i)}\big\}_{i=1}^n$ to satisfy the Lindeberg Condition, we require that the scalar zero-mean sequence $\big\{\boldsymbol{w}^T\boldsymbol{U}^{(i)}\big\}_{i=1}^n$ satisfy the usual Lindeberg Condition (see [37]) for all $\boldsymbol{w} \neq 0$. That is, we wish to show

$$
\max_{1\le i\le n} \frac{\sigma_i^2(\boldsymbol{w})}{B_n(\boldsymbol{w})} \longrightarrow 0 \qquad \forall\ \boldsymbol{w} \in \mathbb{R}^5,\ \boldsymbol{w} \neq \mathbf{0}\,. \tag{5.50}
$$

First, from the proof of Theorem 5.2, Assertion [1.], we take the quantities N, η^*, and $\mu_{n,j}$'s of (5.18). Then, by (5.19), we have

$$B_n(\boldsymbol{w}) = \left|\boldsymbol{w}\right|^2 \left[\frac{\boldsymbol{w}}{|\boldsymbol{w}|}\right]^T \left[\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)\right] \left[\frac{\boldsymbol{w}}{|\boldsymbol{w}|}\right] \geq \left|\boldsymbol{w}\right|^2 \left(\min_{1\leq j\leq 5} \mu_{n,j}\right) \geq n\eta^*\left|\boldsymbol{w}\right|^2 \ \forall\ n > N\,.$$

Next, recall the bound from (5.34) in Section 5.6.2 and apply the Cauchy-Schwarz inequality to see that

$$\begin{aligned}
\sigma_i^2(\boldsymbol{w}) &\leq \frac{1}{\sigma^2}\left(\left|\boldsymbol{w}\right| \left| \boldsymbol{\nabla} f_i\Big|_{\boldsymbol{\theta_0}} \right|\right)^2 \\
&\leq \frac{9}{\sigma^2}\,|\boldsymbol{w}|^2\Big(a^2\,\mathbf{1}\big\{|x_i| \leq a\big\} + x_i^2\,\mathbf{1}\big\{|x_i| > a\big\}\Big) \\
&\leq \frac{9}{\sigma^2}\,|\boldsymbol{w}|^2(a^2 + x_i^2) \\
\therefore \qquad \max_{1\leq i\leq n} \frac{\sigma_i^2(\boldsymbol{w})}{B_n(\boldsymbol{w})} &\leq \frac{9}{n\eta^*\sigma^2}\left\{a^2 + \left(\max_{1\leq i\leq n} |x_i|\right)^2\right\}.
\end{aligned}$$

This $\boldsymbol{w}$-free upper bound has order $o(1)$, since $\max_i |x_i| = o(\sqrt{n})$ by condition [F]. We have therefore established (5.50).

Finally, for any given $\boldsymbol{w} \in \mathbb{R}^5$, $\boldsymbol{w} \neq \mathbf{0}$, the Lindeberg-Feller Central Limit Theorem (see [37]) implies that the sum $\sum_{i=1}^n \boldsymbol{w}^T\boldsymbol{U}^{(i)} = \boldsymbol{w}^T\,\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)$ is asymptotically normal. More precisely, we saw in Theorem 5.2, Assertion [1.], that conditions [A], [B′], and [E] yield a positive definite $n^{-1}\mathbb{I}_{n,\sigma^2}$ for all sufficiently large n, so that $n^{-1}B_n(\boldsymbol{w})$ is asymptotically strictly positive, and

$$\frac{\frac{1}{\sqrt{n}}\boldsymbol{w}^T\,\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)}{\sqrt{\boldsymbol{w}^T\left[\frac{1}{n}\mathbb{I}_{n,\sigma^2}(\boldsymbol{\theta}_0)\right]\boldsymbol{w}}} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{w}^T\boldsymbol{U}^{(i)}}{\sqrt{\frac{1}{n}B_n(\boldsymbol{w})}}$$

has a limiting univariate N(0,1) distribution. Hence, by the definition of a multivariate normal distribution, the zero-mean vector $\boldsymbol{U}_{n,\sigma^2}(\boldsymbol{\theta}_0)$ is asymptotically normal. ■

Chapter 6

Summary of Main Results

In this thesis, we have given some structurally simple conditions on the design (see Sections 4.4, 4.8.4, 5.2.1). These provide a practical guideline for data collection when considering bent-cable regression. They essentially stipulate that there be five detached x-regions which contain substantial fractions of observations. One of these regions is strictly within the underlying transition region, and two of them, strictly within each underlying linear phase. Furthermore, none of the x-values should coincide with an underlying join point,[1] nor are they allowed to accumulate around one. The average absolute and squared values of the x's should be finite even for arbitrarily large samples. Finally, if the response errors are not assumed to be normally distributed, then the largest absolute x-value of the dataset should grow[2] more slowly than the square-root of the sample size.

The above design restrictions are easily met in practice. For example, the design points may be randomly generated from a continuous distribution with a finite variance, whose density is bounded strictly above 0 over each of the five aforementioned x-regions. Alternatively, the design points may be placed systematically at evenly spaced intervals along the x-axis, provided that their range cover substantial portions of each phase of the underlying bent cable. In either case, the investigator's expertise in the subject matter is generally sufficient for determining these regions. However,

[1] This is required purely for notational convenience. See Chapter 4, page 33.

[2] . . . over repeated experimentation.

sequential schemes designed to provide observations close to the underlying join points would, for instance, violate the design constraints.

In the case of a non-trivial underlying bend region, we have shown that these few regularity conditions suffice to compensate for the intrinsic irregularity of the problem due to non-differentiability of the model's first partial derivatives. In particular, they ensure that any sufficiently large dataset contains information about the model parameters, and that the information is more or less proportional to the sample size. They also guarantee that the discontinuities of the Hessian matrix for the error sum-of-squares function are asymptotically negligible, leading to a well-behaved sum-of-squares surface over a decreasing neighborhood of the true parameter value (p. 39, remarks on Lemma 4.3; p. 90, proof of relation (5.18)). Thus, standard maximum likelihood asymptotic results apply to bent-cable regression via least squares.

For an underlying basic bent cable without a middle "transition segment," however, the maximum likelihood asymptotics are unusual and very impractical. The convergence is slow (no faster than $n^{-1/3}$), and the limiting distributions are difficult to work with in practice (see Section 4.8.5). These results prompted us to withdraw from pursuing the theory for a full bent cable whose underlying bend is missing. A further interesting conclusion is that the estimated cable in this case has a non-trivial bend with a probability of at least 1/2 (see Section 4.8.7).

Discussion

This leaves us the thus-far-unaddressed question of testing for a non-trivial bend parametrically. In the next chapter, we suggest alternative approaches for handling this difficulty associated with the highly irregular null model. A related idea is the irregularity of the log-likelihood when the observed responses contain substantial chance errors, as is apparent in two worked examples in Chapter 2. The wide range of bend-width values — including the null value of 0 — which are consistent with these datasets demonstrates how large random errors can heavily obscure the underlying nature of the transition. Moreover, as an extension to the broken stick for describing change-point phenomena, the bent cable is the simplest most parsimonious model that is continuously differentiable. Therefore, our results seem to suggest unavoidable difficulty in obtaining precise inference on the abruptness of change in biological contexts. In such cases, follow-up studies with refined measurements and/or more observations may be required.

Chapter 7

Future Work

We first suggest two alternative approaches to likelihood-based procedures for testing the null hypothesis of a trivial quadratic bend segment. At the end of this chapter, we discuss the somewhat ambiguous notion of a *region of transition*, whose width depends on the order of the polynomial cable bend. We propose an alternative measure to γ for gauging the abruptness of change when modeling data with a general linear-polynomial-linear model that has at least one continuous derivative.

7.1 Testing $\gamma = 0$

One of the biggest concerns of this thesis is the difficulty in assessing the sharpness of change in a continuous response-covariate relationship. The conventional F-test is unjustifiable for testing a *sharp kink* ($\gamma = 0$) versus a *smooth quadratic transition* ($\gamma > 0$) in the context of bent-cable regression (see Chapter 4, Section 4.8). Furthermore, formal hypothesis tests based on likelihood theory are difficult (see Chapter 6). Alternative testing techniques may include computer-intensive procedures such as Bayesian and bootstrap methods.

Existing literature on Bayesian techniques associated with broken-stick regression appears in [38] and [39]. For the more general bent-cable model, it may be reasonable to assume, for instance, an exponential prior (or some prior distribution from the gamma family) for γ, and a normal distribution for the response errors. An "informal"

α-level test of $\gamma = 0$ can then be conducted based on a $100(1-\alpha)\%$ credible region for γ that is computed from the posterior distribution.[1] To conduct a formal Bayesian hypothesis test, a continuous prior for γ is inappropriate, as it gives a zero probability of the point null (and a zero posterior odds ratio). A mixed discrete-continuous distribution with a non-zero point mass at $\gamma = 0$ is a more suitable prior in that case. Of course, the performance of either prior above — and, in general, of Bayesian bent-cable regression — remains to be investigated.

Bootstrap regression is likely no less computer-intensive than the Bayesian approach, but is perhaps more straightforward. An extensive discussion of several bootstrap regression techniques is available in [40], Chapters 6 and 7. For example, one could choose between parametric and non-parametric resampling. Let us first briefly discuss the latter.

As usual, regress the observed responses, $y_1, \ldots, y_n$, on the covariates, $x_1, \ldots, x_n$, according to the full bent-cable model (5.1) of Chapter 5. Label the least-squares parameter estimates (LSE's) as $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\tau}$, and $\widehat{\gamma}$. To test $\gamma = 0$, we now need an approximate null distribution of $\widehat{\gamma}$. For that, regress the y_i's on the x_i's according to the null model of a zero bend width. Label the LSE's of the four free parameters as $\widehat{\beta}_0^{(0)}$, $\widehat{\beta}_1^{(0)}$, $\widehat{\beta}_2^{(0)}$, and $\widehat{\tau}^{(0)}$. The fit produces "null-fit residuals," $\widehat{\varepsilon}_1^{(0)}, \ldots, \widehat{\varepsilon}_n^{(0)}$, and "null-fit responses," $\widehat{y}_1^{(0)}, \ldots, \widehat{y}_n^{(0)}$. The resampling step of the bootstrap method can then be applied to either the $\widehat{\varepsilon}_i^{(0)}$'s or the $\widehat{y}_i^{(0)}$'s. For the former, select a random sample of size n from the empirical distribution of the $\widehat{\varepsilon}_i^{(0)}$'s.[2] Label these resampled values $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$. Now, the "bootstrap responses," $\tilde{y}_1, \ldots, \tilde{y}_n$, are computed for the null model by taking $\tilde{y}_i = \widehat{\beta}_0^{(0)} + \widehat{\beta}_1^{(0)} x_i + \widehat{\beta}_2^{(0)}\big(x_i - \widehat{\tau}^{(0)}\big)\mathbf{1}\big\{x_i > \widehat{\tau}^{(0)}\big\} + \tilde{\varepsilon}_i$. Hence, a set of five "bootstrap LSE's" for the alternative model ($\gamma > 0$, unknown) can be computed by regressing the $\tilde{y}_i$'s on the x_i's according to model (5.1). Repeat the resampling step R times (for a large R) to produce a bootstrap distribution for the

[1] Lange *et al.* have applied a similar technique in [38] for deciding between a linear longitudinal model against a piecewise-linear one for describing the decline of CD4 T-cells in HIV patients. However, for reasons stated in Section 2.4, the choice between a broken stick and a bent cable must be made with caution when not employing a formal hypothesis test.

[2] In [40], it is suggested that the $\widehat{\varepsilon}_i^{(0)}$'s be rescaled or normalized before being resampled. However, we omit the details to preserve the flow in the description of the bootstrap algorithm here.

γ-estimates, $\tilde{\gamma}$'s. This empirical distribution approximates the true null distribution of the original full model LSE, $\widehat{\gamma}$. To compute an approximate P-value for testing $\gamma = 0$, we compare $\widehat{\gamma}$ to the bootstrap distribution of $\tilde{\gamma}$'s. In particular, the fraction of those $\tilde{\gamma}$'s that are no smaller than $\widehat{\gamma}$ is approximately the P-value of interest.

The above method of resampling errors for a fixed design is known as "model-based resampling." It is a non-parametric technique because the resamples are drawn from the empirical distribution of the $\widehat{\varepsilon}_i^{(0)}$'s.[3] These bootstrap residuals could be weighted or normalized if homoscedasticity is not assumed. (See [40] for the advantages of normalization in general.) The parametric version is to resample from an assumed error distribution. For example, if the ε_i's were believed to be normally distributed, one could resample from a $N(0, s^2)$ distribution, where s^2 is the residual mean square computed from the full model fit.

At this point, it is unclear how the Bayesian or bootstrap approach will perform in practice for testing $\gamma = 0$. Future research in this context may include applying both approaches to each dataset of Chapter 2, as well as some general simulation testing and analytical assessments.

7.2 What *Is* the Transition Region?

The motivation for this thesis was the notion of an abrupt transition between two linear phases when a region of smooth transition may be equally plausible. With our once-differentiable linear-quadratic-linear bent-cable model, it is natural to regard the region of transition as the x-interval over which the bend occurs. However, if the same dataset were alternatively modeled by, say, a twice-differentiable linear-quartic-linear bent cable, then, as we shall explain below, the quartic bend would often stretch over a wider x-interval. The notion of a region of transition is seemingly sensitive to the order of smoothness of the bent cable.

The difference between the bend regions is due to the difference in *sharpness* or *curvature* between the quadratic and quartic curves. For a transition region, $[-\gamma, \gamma]$, where γ is fixed, the quartic model sweeps closer to the broken stick underneath than

[3] See Footnote 2 on page 123.

the quadratic model does (Figure 7.1; also see [41]). Hence, for both models' bend segments to exhibit comparable curvatures, as would be required in the modeling context given a fixed dataset, the quartic bend must stretch over a wider interval than the quadratic one.

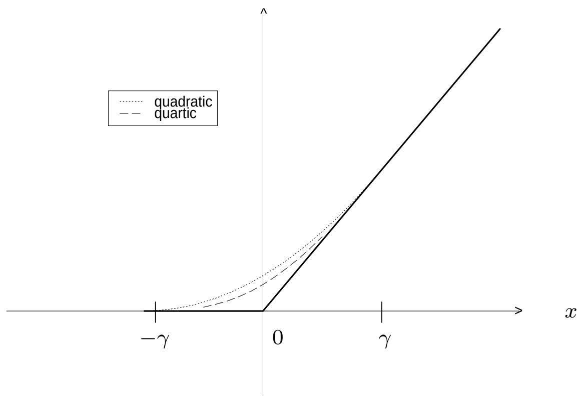

Figure 7.1: For the transition region of both quadratic and quartic models to stretch over the fixed interval $[-\gamma, \gamma]$, the sharper quartic bend must sweep underneath the more mildly curved quadratic bend.

It is important to ensure that the functional forms of the quadratic and quartic models' bend segments are unique so as to avoid ambiguity in the comparison of their respective domains. As we have seen in Chapter 5, a (full) bent-cable model is simply a rescaled basic bent cable. Here, we consider, without loss of generality, bent cables whose bends are centered at 0 only (from which bent cables with non-zero centers can be obtained by simple translations). In general, we refer to any such linear-polynomial-linear model as a *basic* bent cable if it has: (i) incoming and outgoing slopes of 0 and 1, respectively, (ii) a y-intercept of 0, (iii) a bend that stretches over $[-\gamma, \gamma]$, and (iv) an "m-th order smoothness," where m is half the order of the polynomial bend plus one (cf. Chapter 3, p. 27). Now, consider the one-parameter basic bent-cable model, $q(x; \gamma)$, of Section 4.8. It is the unique linear-quadratic-linear function which satisfies constraints (i) to (iv), with $m = 2$ (i.e. only one continuous x-derivative). To verify this uniqueness, consider a generic continuously differentiable function, g, of the form

$$g(x;\tau_1,\tau_2,\alpha) = \begin{cases} 0 & \text{if } x \le \tau_1 \\ \alpha(x-\tau_1)^2 & \text{if } x \in [\tau_1,\tau_2] \\ x & \text{if } x \ge \tau_2 \end{cases}$$

where $[\tau_1,\tau_2]$ is the bend region. As written, g is continuous at τ_1. To make it continuous also at τ_2, we need $\alpha(\tau_2-\tau_1)^2 = \tau_2$. So, rewrite g as

$$g(x;\tau_1,\tau_2) = \begin{cases} 0 & \text{if } x \le \tau_1 \\ \dfrac{\tau_2}{(\tau_2-\tau_1)^2}(x-\tau_1)^2 & \text{if } x \in [\tau_1,\tau_2] \\ x & \text{if } x \ge \tau_2 \end{cases} .$$

Now, the left and right x-derivatives evaluated at τ_1 are both 0, so that g' is continuous at τ_1. At τ_2, they are

$$g'_-(\tau_2) = \frac{2\tau_2}{\tau_2-\tau_1}\,, \qquad g'(\tau_2)_+ = 1\,.$$

In other words, g' is continuous at τ_2 if and only if

$$\frac{2\tau_2}{\tau_2-\tau_1} = 1 \quad \Longleftrightarrow \quad \tau_1 = -\tau_2\,.$$

Taking $\gamma = \tau_2$, we have $\alpha = (4\gamma)^{-1}$ and $\tau_1 = -\gamma$. Now, it is obvious that g is equivalent to $q(x;\gamma)$ of Section 4.8, which is therefore unique under the given constraints.

Similarly, it can be verified that the (centered) quartic basic bent-cable model (with third order smoothness, i.e. continuously twice-differentiable) has the unique form

$$u(x;\gamma) = \begin{cases} 0 & \text{if } x \le -\gamma \\ -\frac{x^4}{16\gamma^3} + \frac{3x^2}{8\gamma} + \frac{x}{2} + \frac{3\gamma}{16} & \text{if } x \in [-\gamma,\gamma] \\ x & \text{if } x \ge \gamma \end{cases} .$$

7.2.1 Dispersion of the Second Derivative Function

Both notions of smoothness and sharpness are directly related to the second derivative function. This function is continuous for u but not for q, and hence, u is *smoother*.

For a fixed bend region of $[-\gamma, \gamma]$, the second derivative has a lower peak for q than for u (Figure 7.2), the latter of which is more convex, and hence, *sharper*. However, we have argued that linear-polynomial-linear models of different orders fitted to the same dataset often produce different γ-values. The sensitivity of γ (a bend-width measure) to the model's order of smoothness makes the notion of a transition region somewhat ambiguous. If this ambiguity is ever an issue in practice, one may wish to gauge the abruptness or sharpness of change by a measure that is related to the idea of a bend width, but is less sensitive to the order of the polynomial bend. We propose using the *dispersion measure of the second derivative function*.

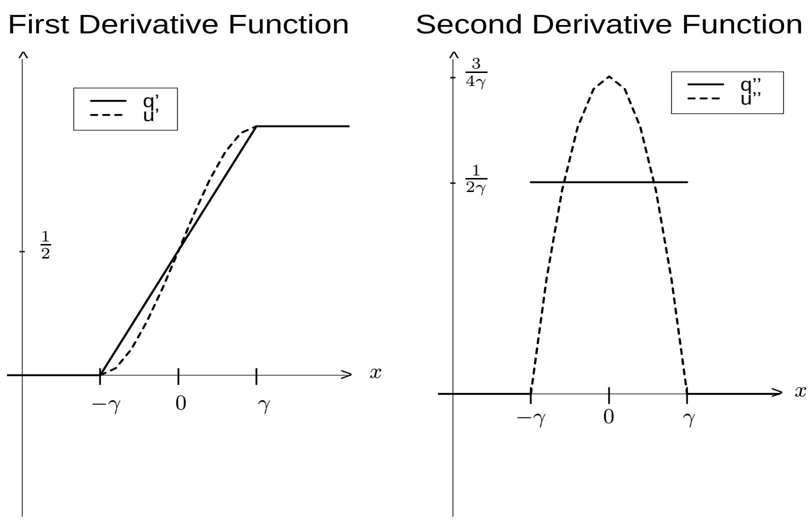

Figure 7.2: The first and second derivative functions of the (centered) quadratic basic bent cable, q, and the (centered) quartic basic bent cable, u. The term *basic* refers to the model constraints of (i) to (iv) on page 125.

First, note that we can view the quadratic or quartic basic bent cable's second derivative as a density function of some zero-mean random variable on $[-\gamma, \gamma]$. This is evident as both q'' and u'' are symmetric about 0, and always integrate to 1 over $[-\gamma, \gamma]$ (and over the entire real line). In fact, the same is true for any linear-polynomial-linear basic bent cable, g, by a simple application of the Fundamental Theorem of Calculus: the slope $g'(x)$ is always 0 at $x \leq -\gamma$ and 1 at $x \geq \gamma$, and g'' is continuous on $[-\gamma, \gamma]$; hence, $\int_{-\gamma}^{\gamma} g''(x)\, dx = g'(\gamma) - g'(-\gamma) = 1$.

Now, the dispersion of a density is large if the density is flat and/or widely spread out. Although q'' is flatter, if the γ-value associated with the quartic fit is larger, then u'' is more spread out. Thus, we may expect the *dispersion* of the "second-derivative density" to be relatively stable regardless of the model's order of smoothness. We can take the dispersion as the square root of the expected squared random variable. Then, the dispersion measure is $\sigma_q = \gamma_q/\sqrt{3}$ for the quadratic model, and $\sigma_u = \gamma_u/\sqrt{5}$ for the quartic model. (The subscript for γ denotes the corresponding bent-cable model.) Note that a larger divisor for the larger γ_u acts as the "stabilizer" for the dispersion measure.

For a general linear-polynomial-linear basic bent cable, g, that has a smoothness order m (and satisfies constraints (i) to (iv) of page 125), we can similarly compute its second derivative dispersion, $\sigma_m(\gamma)$, which is a function of the polynomial bend's half-width, γ. The linear-polynomial-linear *full* bent cable that corresponds to g has a bend centered at some τ with the same half-width, γ. In practice, $\sigma_m(\gamma)$ may be used instead of γ as a measure of the abruptness of the transition that f exhibits. However, one has yet to investigate the actual stability of $\sigma_m(\gamma)$ when different m-th order bent cables are fitted to a given real-life dataset.

Bibliography

[1] Neuman, M. J., Witting, D. A., Able, K. W. (2001). Relationships between otolith microstructure, otolith growth, somatic growth and ontogenetic transitions in two cohorts of windowpane. *J. Fish Biol.* **58**, 967–984.

[2] Barrowman, N. J., Myers, R. A. (2000). Still more spawner-recruitment curves: the hockey stick and its generalizations. *Can. J. Fish. Aquat. Sci.* **57**, 665–676.

[3] Naylor, R. E. L., Su, J. (1998). Plant development of triticale cv. Lasko at different sowing dates. *J. Agric. Sci.* **130**, 297–306.

[4] Routledge, R. D. (1991). Using time lags in estimating anaerobic threshold. *Can. J. Stat.* **19**, 233–236.

[5] Jones, M. C., Handcock, M. S. (1991). Determination of anaerobic threshold: what anaerobic threshold? *Can. J. Stat.* **19**, 236–239.

[6] Brown, C. C. (1987) in *Toxic Substances and Human Risk: Principles of Data Interpretation* (eds. Tardiff, R. G., Rodrick, J. V.) 237–268. Plenum, New York.

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

[8] Routledge, R. D. (1994). Least-squares estimation of a convex function with application to threshold estimation. Technical Report. Department of Mathematics and Statistics, Simon Fraser University, Burnaby, BC.

[9] Tishler, A., Zang, I. (1981). A new maximum likelihood algorithm for piecewise regression. *J. Amer. Stat. Assoc.* **76**, 980–987.

[10] Seber, G. A. F., Wild, C. J. (1989). *Nonlinear Regression*. Wiley, New York.

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

[12] Antonutto, G., Di Prampero, P. E. (1995). The concept of lactate threshold. A short review. *J. Sports Med. Phys. Fitness* **35**, 6–12.

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

[14] Moquin, A., Mazzeo, R. S. (2000). Effect of mild dehydration on the lactate threshold in women. *Med. Sci. Sports Exerc.* **32**, 396–402.

[15] Schneider, D. A., McLellan, T. M., Gass, G. C. (2000). Plasma catecholamine and blood lactate responses to incremental arm and leg exercise. *Med. Sci. Sports Exerc.* **32**, 608–613.

[16] Vachon, J. A., Bassett Jr., D. R., Clarke, S. (1999). Validity of the heart rate deflection point as a predictor of lactate threshold during running. *J. Appl. Physiol.* **87**, 452–459.

[17] 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. *J. Appl. Physiol.* **76**, 1144–1149.

[18] Kline, K. A. (1997). Metabolic Effects of Incremental Exercise on Arabian Horses Fed Diets containing Corn Oil and Soy Lecithin. M. S. thesis. Virginia Polytechnic Institute and State University, Blacksburg.

[19] Beaver, W. L., Wasserman, K., Whipp, B. J. (1985). Improved detection of lactate threshold during exercise using a log-log transformation. *J. Appl. Physiol.* **59**, 1936–1940.

[20] Bacon, D. W., Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika* **58**, 525–534.

[21] Feder, P. I. (1975). The log likelihood ratio in segmented regression. *Ann. Stat.* **3**, 84–97.

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

[23] Hinkley, D. V. (1971). Inference in two-phase regression. *J. Amer. Stat. Assoc.* **66**, 736–743.

[24] Gallant, A. R., Fuller, W. A. (1973). Fitting segmented polynomial regression models whose join points have to be estimated. *J. Amer. Stat. Assoc.* **68**, 144–147.

[25] Sprent, P. (1961). Some hypotheses concerning two phase regression lines. *Biometrics* **17**, 634–645.

[26] Hudson, D. J. (1966). Fitting segmented curves whose join points have to be estimated. *J. Amer. Stat. Assoc.* **61**, 1097–1129.

[27] Feder, P. I. (1975). On asymptotic distribution theory in segmented regression problems — identified case. *Ann. Stat.* **3**, 49–83.

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

[29] Gallant, A. R. (1975). Inference for nonlinear models. Mimeograph Series No. 875. Institute of Statistics , North Carolina State University, Rayleigh, NC.

[30] Ivanov, A. V. (1997). *Asymptotic Theory of Nonlinear Regression.* Kluwer Academic Publishers, Dordrecht.

[31] Golub, G. H., Van Loan, C. F. (1996). *Matrix Computations.* Johns Hopkins University Press, Baltimore.

[32] Stoer, J., Witzgall, C. (1970). *Convexity and Optimization in Finite Dimensions*, I. Springer-Verlag, Berlin.

[33] Kim, J., Pollard, D. (1990). Cube root asymptotics. *Ann. Stat.* **18**, 191–219.

[34] Breiman, L. (1968). *Probability.* Addison-Wesley, Reading.

[35] Rudin, W. (1964). *Principles of Mathematical Analysis*, ed. 2. McGraw-Hill, New York.

[36] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[37] Weisstein, E. W. (1999). *CRC Concise Encyclopedia of Mathematics.* CRC Press, Boca Raton.

[38] Lange, N., Carlin, B. P., Gelfand, A. E. (1992). Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers. *J. Amer. Stat. Assoc.* **87**, 615–626.

[39] Smith, A. F. M., Cook, D. G. (1980). Straight lines with a change-point: A Bayesian analysis of some renal transplant data. *Appl. Statist.* **29**, 180–189.

[40] Davison, A. C., Hinkley, D. V. (1997). *Bootstrap Methods and their Application.* Cambridge University Press, Cambridge.

[41] Zang, I. (1980). A smoothing-out technique for min-max optimization. *Math. Prog.* **19**, 61–77.