

# Adaptive and Multistage Procedures for Inference on Monotone Regression Functions in Observed Data Studies and Design Settings

by

Runlong Tang

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Statistics)  
in The University of Michigan  
2011

Doctoral Committee:

Associate Professor Moulinath Banerjee, Co-Chair  
Professor George Michailidis, Co-Chair  
Professor Michael B. Woodroffe  
Associate Professor Bin Nan  
Professor Michael R. Kosorok, University of North Carolina - Chapel Hill

© Runlong Tang 2011  

---

All Rights Reserved

*To Mom and Dad*

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisors, Moulinath Banerjee and George Michailidis. Without their guidance, encouragement and inspiration, this dissertation would not have been written. I would also like to thank Michael B. Woodrooffe and Michael R. Kosorok for their truly insightful discussions. Thanks to Bin Nan for being on my dissertation committee and to Ji Zhu and Susan Murphy for their encouragement and help in ways more than one. Finally, thanks to the department staff, Lu Ann, Mary Ann, Amy and Suleman, for their help on numerous occasions, especially in complicated official matters, and to my fellow graduate students in the department who were always available for discussion, advise and help.

# TABLE OF CONTENTS

<b>DEDICATION</b> . . . . .	ii
<b>ACKNOWLEDGEMENTS</b> . . . . .	iii
<b>LIST OF FIGURES</b> . . . . .	vii
<b>LIST OF TABLES</b> . . . . .	viii
<b>ABSTRACT</b> . . . . .	ix
<b>CHAPTER</b>	
<b>I. Introduction</b> . . . . .	1
<b>II. Asymptotics for Current Status Data under Varying Observation Time Sparsity</b> . . . . .	6
2.1 Introduction . . . . .	6
2.2 Formulation of the Problem and Overview of Findings . . . . .	10
2.3 Estimation . . . . .	13
2.4 Asymptotic Results . . . . .	16
2.4.1 Case One $\gamma \in (0, 1/3)$ . . . . .	16
2.4.2 Case Two $\gamma \in (1/3, 1)$ . . . . .	22
2.4.3 Case Three $\gamma = 1/3$ . . . . .	25
2.5 Adaptive Inference for $F$ at A Point . . . . .	36
2.6 A Practical Procedure and Simulations . . . . .	45
2.7 Conclusions . . . . .	50
2.8 Appendix . . . . .	52
2.8.1 Supplementary proofs for Case One $\gamma \in (0, 1/3)$ . . . . .	52
2.8.2 Proofs for Case Three $\gamma = 1/3$ . . . . .	57
<b>III. A Two-Stage Hybrid Procedure for Estimating an Inverse Regression Function</b> . . . . .	66

3.1	Introduction . . . . .	66
3.2	Two-Stage Procedures . . . . .	70
	3.2.1 Preliminaries: A Single-Stage Procedure . . . . .	70
	3.2.2 Procedures Based On Two-Stage Sampling Designs . . . . .	71
	3.2.3 Bootstrapping The Two-Stage Estimator . . . . .	72
3.3	Asymptotic Properties of Two-Stage Estimators . . . . .	73
	3.3.1 Two-Stage Estimator . . . . .	74
	3.3.2 Bootstrapped Two-Stage Estimator . . . . .	76
3.4	Performance Evaluation . . . . .	79
	3.4.1 Comparison of Two-Stage Procedures . . . . .	81
	3.4.2 Comparison of One- and Two-Stage Procedures . . . . .	83
3.5	Data Application . . . . .	85
3.6	Conclusions . . . . .	87
3.7	Appendix . . . . .	89
	3.7.1 Important Lemmas . . . . .	89
	3.7.2 Proofs for Results in Subsection 3.3.1 . . . . .	90
	3.7.3 Proofs for Results in Subsection 3.3.2 . . . . .	95
	3.7.4 Some Auxiliary Lemmas . . . . .	100

**IV. Two Stage Nonparametric Procedures for Estimating A Threshold Value of A Regressor . . . . . 102**

4.1	Introduction . . . . .	102
4.2	Preliminaries . . . . .	104
	4.2.1 One-Stage Isotonic Regression Procedure (OSIRP) . . . . .	104
	4.2.2 One-Stage Smoothed Isotonic Regression Procedure (OSSIRP) . . . . .	107
4.3	Two-Stage Procedures . . . . .	109
	4.3.1 Two-Stage Isotonic Regression Procedure $I+I$ (TSIRP) . . . . .	110
	4.3.2 Two-Stage Smoothed Isotonic Regression Procedure $I + I_s$ (TSSIRP) . . . . .	116
4.4	Practical Procedures . . . . .	118
	4.4.1 Practical OSIRP (POSIRP) . . . . .	118
	4.4.2 Practical OSSIRP (POSSIRP) . . . . .	119
	4.4.3 Practical TSIRP (PTSIRP) . . . . .	120
	4.4.4 Practical TSSIRP (PTSSIRP) . . . . .	122
4.5	Simulation Study . . . . .	124
4.6	An Application to Real Data . . . . .	127
	4.6.1 Preliminary Analysis . . . . .	129
	4.6.2 Practical Procedures . . . . .	131
4.7	Conclusion . . . . .	135
4.8	Appendix . . . . .	135
	4.8.1 Appendix for OSIRP . . . . .	135
	4.8.2 Appendix for OSSIRP . . . . .	136
	4.8.3 Appendix of TSIRP . . . . .	138

4.8.4 Appendix for TSSIRP . . . . .	144
<b>BIBLIOGRAPHY</b> . . . . .	150

## LIST OF FIGURES

<u>Figure</u>		
2.1	An illustration for showing $\{L_c\}$ is $O_P(1)$ . . . . .	44
2.2	A comparison of the coverage rates and average lengths . . . . .	48
3.1	The average delay as a function of system's loading. . . . .	67
3.2	Coverage rate plot grouped with different $\sigma$ 's . . . . .	82
3.3	Estimation of the variance function in POSP and PBTSP . . . . .	86
3.4	Comparing POSP and PBTSP . . . . .	88
4.1	The admissible region and optimal convergence rates . . . . .	117
4.2	Regression functions and their derivatives . . . . .	125
4.3	Scatter plots of data . . . . .	126
4.4	Coverage rates and average lengths with the sigmoid and quadratic functions . . . . .	127
4.5	Coverage rates and average lengths with the isotonic sine function .	128
4.6	Relationships between the horse power and fuel efficiency . . . . .	129
4.7	Relationships with different methods of air aspiration . . . . .	130
4.8	Relationships with the method of natural air aspiration . . . . .	131
4.9	Plots for the data analysis results . . . . .	133
4.10	The admissible region and optimal convergence rates . . . . .	146



## LIST OF TABLES

### Table

2.1	A comparison of the coverage rates and average lengths . . . . .	49
3.1	A comparison of coverage rates, average lengths and mean square errors	83
3.2	A comparison of POSP and PBTSP . . . . .	87
4.1	Data analysis results for the five practical procedures . . . . .	132

# ABSTRACT

Adaptive and Multistage Procedures for Inference on Monotone Regression  
Functions in Designed Studies and Observed Data Settings

by

Runlong Tang

Co-Chairs: Moulinath Banerjee and George Michailidis

This dissertation consisting of three works addresses two basic shape-restricted problems: the estimation of a distribution function at a given point in an observational study and the estimation of the inverse of an increasing function at a given point in a design setting.

In the first work, we study the estimation and hypothesis testing for a failure time distribution function at a point in the current status model with observation times supported on a grid of potentially unknown sparsity and with multiple subjects sharing the same observation time. This is of interest since observation time ties occur frequently with current status data. The grid resolution is specified as  $cn^{-\gamma}$  with  $c > 0$  being a scaling constant and  $\gamma > 0$  regulating the sparsity of the grid relative to the number of subjects  $n$ . The asymptotic behavior falls into three cases depending on  $\gamma$ : regular Gaussian-type asymptotics obtain for  $\gamma < 1/3$ , non-standard cube-root Chernoff-type asymptotics prevail when  $\gamma > 1/3$  and  $\gamma = 1/3$  serves as a boundary at which transition happens. The boundary limit distribution indexed by  $c$  is different from the Gaussian and Chernoff limit distributions in the previous cases but converges weakly to them as  $c$  goes to  $\infty$  and  $0$ , respectively. This relationship among the limit distributions allows us to develop an adaptive procedure to construct confidence intervals for the value of the failure time distribution at a point of interest without needing to specify or estimate  $\gamma$ , which is of enormous advantage from the perspective of inference. A simulation study of the adaptive procedure is presented.

In the second work, we consider a hybrid two-stage procedure (TSP) for estimating an inverse regression function at a given point, where isotonic regression is first implemented at stage one to obtain an initial estimate and then a local linear approximation is exploited over the vicinity of this estimate at stage two. The convergence rate of the second-stage estimate can attain the parametric rate  $n^{1/2}$ . Furthermore, a bootstrapped variant of TSP (BTSP) is introduced and its consistency properties established. This variant manages to overcome the slow speed of the convergence in distribution and the estimation of unknown parameters. Finally, the finite sample performance of BTSP is studied through simulations and the method is illustrated on a data set.

The third work shares the same basic problem with the second work. In practice, the previous hybrid two-stage procedure usually results in a biased estimator and thus confidence intervals with low coverage rates if the sample size is not large and the regression function is very locally nonlinear around the target point. For such cases, we propose to adopt isotonic regression and smoothed isotonic regression at stage two and denote the resulting two-stage procedures by TSIRP and TSSIRP, respectively. The convergence rate of the second-stage estimate from TSIRP is less than the parametric rate while that from TSSIRP can still achieve it. The performance of practical TSIRP and TSSIRP is studied by simulations.

# CHAPTER I

## Introduction

A model in statistics is a class of distribution functions indexed by an unknown parameter living in a mathematical space. According to the parameter spaces, statistical models are usually classified into three categories: parametric, nonparametric and semiparametric models, which are the combinations of the former two types of models.

In a nonparametric model, the parameter is usually an unknown function with an intuitive restriction. One popular constraint is posed on the smoothness of the unknown function, which leads to various kernel smoothing and spline methods. Another natural restriction is on the shape of the unknown function, such as monotonicity or convexity. The related statistical problems are usually called shape-restricted problems.

In this dissertation, we consider two basic shape-restricted regression problems: one is on the estimation of a distribution function, or more generally, an increasing function, at a given point in an observational study; the other is on the estimation of the inverse of an increasing function at a given point in a design setting. In both problems, a crucial step is to estimate the unknown increasing function by the so-called isotonic regression method. The resulting isotonic regression estimator of the unknown increasing function at a point is a local average, similar to a kernel smoothing estimator. However, different from the specified deterministic bandwidth in the kernel smoothing estimator, the bandwidth for the local average in isotonic regression is automatically chosen and random. In order to handle this randomness in the local average and rigorously derive theoretical results, the classical theory on central limit theorems for random vectors is not sufficient so that we exploit the approach provided by the theory of empirical processes, which is essentially a theory

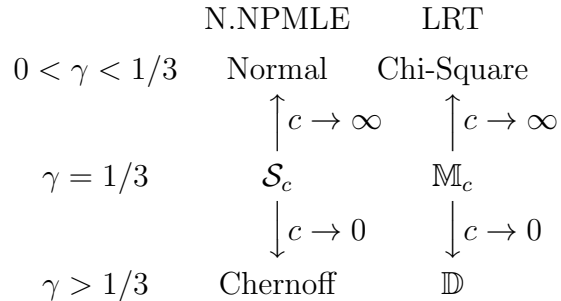
on functional central limit theorems.

Three works have been done to address the above two basic shape-restricted problems. They are organized into the following three chapters. The work in Chapter II is for the first problem and the works in Chapters III and IV for the second one. Next, we give a brief separate introduction for each work.

**Chapter II: Asymptotics for Current Status Data under Varying Observation Time Sparsity.** The current status model, being the simplest avatar of interval-censoring — one observes the individual at a single time-point and ascertains whether they are infected or not by that time — and therefore best suited to an understanding of the more general phenomenon, has unsurprisingly received much attention in the statistical literature.

In this work, we consider estimating the distribution function of the infection time at an observation time of interest and constructing confidence intervals for it. Since the infection distribution function is increasing, it turns out that its nonparametric maximum likelihood estimator (NPMLE) is equivalent to the isotonic regression of it. To construct confidence intervals, we exploit weak convergence results on both the normalized NPMLE (N.NPMLE) and the likelihood ratio test statistic (LRT). A standard assumption for the current status model is that the observation times are random observations from a Lebesgue density on the time-domain. However, this assumption, though convenient for mathematical tractability, is often an unnatural one. For example, health care providers inspecting patients may only observe them at specific limited times, depending on their schedules, with multiple patients at one time. In such cases, it seems more reasonable to assume that the observation time lies on a grid in an interval and that the number of the grid points (grid size) increases with the sample size. So, in this work, we assume that the observation time follows a distribution on a grid and the grid size increases to infinity as the sample size goes to infinity. It turns out that the rate of the increase of the grid size completely determines the asymptotic behavior of N.NPMLE and LRT. When the rate is low, there are so many individuals at each grid point that isotonization becomes asymptotically unnecessary and simple averages play the main role. It is no wonder that the limit distributions of N.NPMLE and LRT are normal or chi-square distributions. On the other hand, when the rate is high, there are so few individuals at each grid point that simple averages need to be isotonized and local averages with automatically determined random bandwidths play the main role. For this case,

the limit distributions are Chernoff and the so-called  $\mathbb{D}$  distributions, which are the same as those under the Lebesgue density assumption. This means if the grid is dense enough in the interval the discrete distribution on the grid is asymptotically equivalent to a Lebesgue density. Since the limit distributions are so different as the rate of grid size changes from low to high, there naturally arises an interesting question: does there exist a boundary rate? The answer is affirmative and the boundary rate is of order  $n^{1/3}$ . For this case, the grid is *just* dense enough to invoke isotonization and the limit distributions are different from those of the other two cases and depend on discrete time stochastic processes. In order to obtain more flexibility, we introduce a scaling constant  $c$  into the rate. That is, the rate is set to be of order  $(1/c)n^\gamma$  with constant  $c > 0$  and  $\gamma \in (0, 1)$ . We denote the limiting distributions of N.NPMLE and LRT for the boundary case with  $\gamma = 1/3$  by  $\mathcal{S}_c$  and  $\mathbb{M}_c$ , respectively. Clearly, if the scaler  $c$  depends on  $n$ , it can change the rate order and thus the limiting distributions. What is more interesting is that the scaler  $c$  also plays a similar role on the limit distribution side! More specifically, in one direction,  $\mathcal{S}_c$  and  $\mathbb{M}_c$  converge weakly to the normal and chi-square distributions of the sparse grid case as  $c$  goes to infinity; in the other direction,  $\mathcal{S}_c$  and  $\mathbb{M}_c$  converge weakly to the Chernoff and  $\mathbb{D}$  distributions of the dense grid case as  $c$  goes to zero. The following diagram illustrates these relationships among limiting distributions.



This diagram clearly shows that the normal and Chernoff distributions are the extremes of the boundary distribution  $\mathcal{S}_c$  and the chi-square and  $\mathbb{D}$  distributions are the extremes of  $\mathbb{M}_c$ . This exciting discovery provides a theoretically strong and natural support for a claim that the counterparts of the two perhaps most important classical distributions normal and chi-square in the isotonic regression world are Chernoff and  $\mathbb{D}$ . Further, from this discovery, an adaptive procedure is proposed to construct confidence intervals for the unknown infection distribution function at a point without estimating or specifying the grid size rate, which is certainly of practical interest.

**Chapter III: A Two-Stage Hybrid Procedure for Estimating an Inverse Regression Function.** It is often of interest to estimate  $d_0$ , the inverse of an increasing regression function  $f$  at a point of interest. For example, for a drug, there is usually an increasing relationship between its dose and toxicity level. It is important in developing the drug to identify the maximum safety dose given a maximum tolerance level of toxicity. For such problems, it is natural to obtain the isotonic regression of the increasing function and then compute its inverse. Although this nonparametric approach has a significant advantage of not involving tuning parameters, the convergence rate of the estimator is only  $n^{1/3}$ , slower than the usual parametric rate  $n^{1/2}$ . In many dose-response-type problems we can actually design the related studies. Based on a zoom-in idea of improving the quality of the second-stage data, we propose a hybrid two-stage procedure with the intention to produce a estimate having a faster convergence rate.

From Taylor's expansion,  $f$  is locally linear at  $d_0$  under mild assumptions. This means if we can first identify a small neighborhood around  $d_0$ , then we can effectively fit  $f$  locally with a line. In this way, a nonparametric problem is asymptotically transformed into a parametric one and the convergence rate may be accelerated. More specifically, at the first stage we use some sample points to obtain a small interval around  $d_0$  by isotonic regression. Then, at the second stage, all the remaining doses are evenly allocated at the two ends of the interval since this is the most efficient allocation for fitting a line over an interval by least squares. It turns out the convergence rate of the second-stage estimator can reach  $n^{1/2}$ , which means the parametric rate is achieved in a *nonparametric* problem! However, when using the limit distribution with the parametric rate to construct confidence intervals for  $d_0$ , we face two difficulties: one is on the estimation of  $f'(d_0)$ , the slope of the unknown function at the unknown target point appearing in the limit distribution; the other one is a little implicit, which is the slow speed of convergence in distribution indicated by simulation studies. To deal with these difficulties, we bootstrap the second-stage responses. The corresponding bootstrapped second-stage estimator is proved to be strongly consistent. Thus, we obtain a two-stage bootstrapped nonparametric procedure which achieves the parametric rate and at the same time avoids the above two difficulties in constructing confidence intervals for  $d_0$ . We have also proposed a practical version of the bootstrapped procedure, which performs well in simulations,

and applied this practical procedure to an engineering problem.

**Chapter IV: Two Stage Nonparametric Procedures for Estimating a Threshold Value of A Regressor.**

The successful application of the previous hybrid two-stage procedure in practice relies on the sample size and the local linearity of the regression function. If the sample size is not large so that the second-stage sampling interval is relatively long and the regression function is quite locally nonlinear over the second-stage sampling interval, the local linear fitting step in the hybrid two-stage procedure would usually bring significant bias into the estimate of  $d_0$ , which would further cause low coverage rates for confidence intervals. In order to address this issue, we propose alternative two-stage nonparametric procedures.

More specifically, after identifying a small neighborhood of  $d_0$ , we consider fitting  $f$  locally by isotonic regression again instead of a line via least squares. This time, the second-stage design points are distributed within the interval instead of only at the two ends. Intuitively, this approach would have more nonparametric flavor and thus should be able to handle regression functions with more nonlinearities better than the previous hybrid two-stage procedure, though the convergence rate might not achieve the parametric rate  $n^{1/2}$ . Suppose the order of the interval length is  $n^{-\gamma}$  for  $\gamma \in (0, 1/3)$ . Then the convergence rate of the second-stage estimator turns out to be  $n^{(1+\gamma)/3}$ , faster than  $n^{1/3}$  but still slower than the parametric rate. When this two-stage isotonic regression procedure is extended to a  $K$ -stage one, the convergence rate of the  $K$ th-stage estimator increasingly converges to the parametric rate as  $K$  goes to infinity. This discovery reveals an essential difference in terms of convergence rates between parametric methods and a special nonparametric method — isotonic regression. Further, this discovery drives us to ask the interesting question whether a  $K$ -stage nonparametric procedure can reach or even exceed the parametric rate if a nonparametric method with a one-stage convergence rate faster than  $n^{1/3}$  is employed. We have considered smoothed isotonic regression, a combination of isotonic regression and kernel smoothing, whose one-stage convergence rate can be  $n^{2/5}$ . It turns out that the  $K$ -stage procedure with this smoothed isotonic regression can *not* exceed the parametric rate. This, to some extent, indicates that the difference between parametric approaches and nonparametric ones is so fundamental that continuously improving data quality could not make nonparametric approaches outperform parametric ones. However, although the parametric rate can not be exceeded, a heuristic derivation shows that the two-stage smoothed isotonic regression procedure could achieve it.



## CHAPTER II

# Asymptotics for Current Status Data under Varying Observation Time Sparsity

### 2.1 Introduction

The current status model is one of the most well-studied survival models in statistics. An individual at risk for an event of interest is monitored at a random status or observation time, and an indicator of whether the event has occurred is recorded. An interesting feature of this kind of data is that the underlying survival function for the event time can only be estimated nonparametrically at the  $n^{1/3}$  rate when the status time is a continuous random variable. Under mild conditions on the survival distribution, the limiting distribution of the estimator in this setting is the non-Gaussian Chernoff's distribution. This is in contrast to right-censored data where the underlying survival function can be estimated nonparametrically at rate  $\sqrt{n}$  under right-censoring and is 'pathwise norm-differentiable' in the sense of *van der Vaart (1991)*, admitting regular estimators and normal limits. Interestingly, when the status time distribution has finite support, the survival function estimator for current status data simplifies to a binomial random variable and is also  $\sqrt{n}$  estimable and regular, with a normal limiting distribution. This intriguing change in limiting distribution under different degrees of sparsity of the status time distribution is the focus of this paper.

While the distinction between finite support and continuously distributed status times is clear philosophically, the practical distinction between these two settings is less clear. For example, suppose we observe a large number of ties in status times which are distributed across a large number of distinct times, as can happen when there is only one status time per day or only one status time per week but there are a moderately large number of days or weeks in the study; for example, a group of

health care providers inspecting people in a community (at risk for infection, say) over a large number of days may only observe them at a specific time everyday, depending on their schedule, with multiple patients being inspected at the same time. How large should the number of distinct status times relative to  $n$  (the total number of individuals inspected) be for the distribution of the estimates of the survival function to be closer to the Chernoff limit than the Gaussian? Under certain configurations, will the distribution be closer to a limiting distribution that is neither Gaussian nor Chernoff? These kinds of questions are the focus of this paper. Specifically, *we wish to assess the possible, relevant limiting distributions that can arise under different levels of sparsity of the status time distribution as well as develop a statistical procedure that is capable of adapting to whatever asymptotic regime is most suitable for the data.* This technically challenging problem is practically important since the most relevant asymptotic regime is seldom known in practice except for the very extreme settings where either there are only a few distinct status times (in which case a normal approximation should work well) or when there are no ties at all in the status times (the setting of a continuous status time).

The current status model, being the simplest avatar of interval-censoring — one observes the individual at a single time-point and ascertains whether they are infected or not by that time — and therefore best suited to an understanding of the more general phenomenon, has unsurprisingly received much attention in the statistical literature. The model itself goes back to [Ayer et al. \(1955\)](#) and was subsequently studied in [Turnbull \(1976\)](#) in a more general framework; asymptotic properties for the nonparametric maximum likelihood estimator (NPMLE) of the survival distribution were first obtained in [Groeneboom and Wellner \(1992\)](#) and involved techniques radically different from those in ‘classical’ survival analysis, and since then there has been a notable body of work on both the methodological and asymptotic fronts: see, for example, [Huang \(1996\)](#) where current status data under the Cox PH model is studied; [Lin et al. \(1998\)](#) and [Shiboski \(1998\)](#) for work on additive hazards regression and generalized additive models, respectively, for current status data; [Sun \(1999\)](#) for current status data under unequal censoring; [Banerjee and Wellner \(2001\)](#) and [Banerjee and Wellner \(2005\)](#) for a study of pointwise likelihood ratio tests for the survival distribution in the current status model that lead to asymptotically pivotal methods for inference in this model with broader implications for monotone function estimation; [Jewell et al. \(2003\)](#), [Groeneboom et al. \(2008a\)](#) and [Groeneboom et al.](#)

(2008b) for current status model under competing risks with the last two papers providing a comprehensive description of the highly involved asymptotics that comes into play; and *Ma and Kosorok* (2005) for some recent semiparametric analysis involving partly linear transformation models with current status data. The above list is only a sample of the enormous body of work involving the current status model and is meant to reflect some of the main directions along which methodological and theoretical research have evolved over the last two decades, and also directions in which the authors have taken a more active interest, but should amply illustrate the wide range of problems that present themselves within the context of this relatively simple model. Interestingly though, the problem of determining the correct asymptotic approximation in current status data with ties and the development of an adaptive inference scheme, as discussed in the first two paragraphs, has not been satisfactorily resolved thus far in the rather large literature.

Denoting the survival distribution by  $F$  and with  $x_0$  being a point of interest in the time domain, the goal is to determine how the behavior of the NPMLE of  $F(x_0)$  depends on the relative magnitude of the number of distinct observation times,  $K(n)$ , to the sample size  $n$ . This change in the model is important not only because it more closely mirrors applications than the Lebesgue density assumption, as discussed above, but also because it induces surprisingly non-trivial complications in the limiting distribution and consequently in conducting inference. Since  $F$ , as will be shown later, is only identifiable at the distinct observation times, one can think about  $K(n)$  as the effective number of parameters in the model. We will be investigating the behavior of  $\hat{F}$  (the NPMLE of  $F$ ) under a discrete observation time scheme where  $K(n)$ , the determinant of the sparsity of the observation time scheme, will be allowed to go to infinity with  $n$  at different rates. High rates correspond to dense observation time schemes and low rates to sparse ones. It should be noted here that some authors have indeed studied the current status model or closely related variants under discrete observation time settings. *Yu et al.* (1998) have studied the asymptotic properties of the NPMLE of  $F$  in the current status model with discrete observation times and more recently *Maathuis and Hudgens* (2010) have considered nonparametric inference for competing risks current status data under discrete or grouped observation times. However, these papers consider situations where the observation times are i.i.d. copies from a *fixed discrete distribution* (but not necessarily finitely supported) on the time-domain and are therefore not geared towards studying the effect of the

trade-off between  $n$  and  $K(n)$ , i.e. *the effect of the relative sparsity of the number of distinct observation times to the size of the cohort of individuals* on inference for  $F$ . In both these papers, the (pointwise) estimates of  $F$  are asymptotically normal and  $\sqrt{n}$  consistent; however, in situations, where the number of distinct observation times is large relative to the sample size, such normal approximations based on a fixed discrete observation time distribution might be suspect; see, for example, Section 5.1 of [Maathuis and Hudgens \(2010\)](#) where this is illustrated via simulations. We will return to the paper [Maathuis and Hudgens \(2010\)](#) later in our concluding discussion. [Zhang et al. \(2001\)](#) consider isotonic regression with grouped data in the monotone density estimation problem. The asymptotic properties of the isotonic estimator are shown to depend on the order of magnitude of the grouping intervals (which are allowed to depend on the sample size) and the corresponding limit distributions determined but no effective inference schemes are developed. More recently, [Wang and Shen \(2010\)](#) have considered monotone regression estimators based on grouped data using B-splines and have studied their asymptotic properties in terms of the number of knots as the sample size increases.

The rest of the paper is organized as follows. In Section 2, we introduce our mathematical formulation of the problem and present an overview of the basic results at a high level. Section 3 discusses the characterizations of the estimators of interest while Section 4 presents the main asymptotic results and major proofs. Section 5 addresses the important question of *adaptive inference* in the current status model: given a time-domain  $[a, b]$  and current status data observed at times on a regular grid on  $[a, b]$  of an *unknown level of sparsity*, how do we make inference on  $F$ ? What asymptotic approximations should we use? We provide an answer to this question that *circumvents the need to determine the sparsity of the grid*, and therefore provides a tremendous advantage from the point of view of inference. Section 6 provides simulation results for a practical version of the adaptive inference procedure developed in Section 5. Section 7 concludes with a discussion of the findings of this paper and their implications for monotone regression models in general, as well as more complex forms of interval censoring and interval censoring with competing risks.

## 2.2 Formulation of the Problem and Overview of Findings

Let  $\{T_{i,n}\}_{i=1}^n$  be i.i.d. random survival times following some unknown distribution  $F$  with Lebesgue density  $f$  concentrated on the time-domain  $[a', b']$  with  $0 \leq a' < b' < \infty$  (or supported on  $[a', \infty)$  if no such  $b'$  exists) and  $\{X_{i,n}\}$  are i.i.d. observation times drawn from a discrete probability measure  $H_n$  supported on a regular grid on  $[a, b]$  with  $a' \leq a < b < b'$ .<sup>1</sup> Also,  $T_{i,n}$  and  $X_{i,n}$  are assumed to be independent for each  $i$ . However,  $\{T_{i,n}\}$  are not observed; rather, we observe  $\{Y_{i,n} = 1\{T_{i,n} \leq X_{i,n}\}\}$ . This puts us in the setting of a binary regression model with  $Y_{i,n}|X_{i,n} \sim \text{Bernoulli}(F(X_{i,n}))$ . We denote the support of  $H_n$  by  $\{t_{i,n}\}_{i=1}^K$  where the  $i$ -th grid point  $t_{i,n} = a + i\delta$ , the unit spacing  $\delta = \delta(n) = cn^{-\gamma}$  (also referred to as the *grid resolution*) with  $\gamma \in (0, 1)$  and  $c > 0$ , and the number of grid points  $K = K(n) = \lfloor (b-a)/\delta \rfloor$ . On this grid, the distribution  $H_n$  is viewed as a discretization of an absolutely continuous distribution  $G$ , whose support contains  $[a, b]$  and whose Lebesgue density is denoted as  $g$ . More specifically,  $H_n\{t_{i,n}\} = G(t_{i,n}) - G(t_{i-1,n})$ , for  $i = 2, 3, \dots, K-1$ ,  $H_n\{t_{1,n}\} = G(t_{1,n})$  and  $H_n\{t_{K,n}\} = 1 - G(t_{K-1,n})$ . For simplicity, these discrete probabilities are denoted as  $p_{i,n} = H_n\{t_{i,n}\}$  for  $i = 1, 2, \dots, K$ . Let  $x_0 \in (a, b)$  be a point around which we are interested in determining the properties of the NPMLE of  $F$  as a function of the grid resolution  $\delta$  or equivalently in terms of  $(c, \gamma)$ . In what follows, we refer to the pair  $(X_{i,n}, Y_{i,n})$  as  $(X_i, Y_i)$ , suppressing the dependence on  $n$ , but the *triangular array nature* of our observed data should be kept in mind. Similarly, the subscript  $n$  is suppressed elsewhere when no confusion will be caused.

Given the point of interest  $x_0 \in (a, b)$ , define  $t_l = t_{l,n}$  to be the largest grid-point less than or equal to  $x_0$  and  $t_r = t_{r,n} = t_{l+1,n}$  the smallest grid point larger than  $x_0$ . As will be seen, the NPMLE  $\hat{F}$  is only identifiable up to its values at the grid-points. To define it on the interval  $[a, b]$  one usually resorts to some convention of extension. Our estimate  $\hat{F}$  will be taken to be the unique right-continuous step function with potential jumps only at the grid-points that maximizes the likelihood function of the observed data over all piece-wise constant right-continuous distribution functions supported on  $[a, b]$ .

We start with the estimation of  $F(t_l)$  using  $\hat{F}(t_l)$ . The key features of the effect of sparsity on the asymptotics are best illustrated by focusing on  $t_l$  (which can be

---

<sup>1</sup>The regularity of the grid corresponds to evenly spaced inspection times which is satisfied in many clinical and engineering applications (patients inspected daily at a clinic or machines inspected routinely every week, for example). It also makes the subsequent derivations simpler to present without compromising the complexity of the intrinsic mathematical issues involved.

viewed as a natural surrogate for  $x_0$ ). We will demonstrate, explicitly, that resolving the inference problem for  $F(t_l)$  allows us to resolve the inference problem for  $F(x_0)$  with only minor modifications. The reason behind using  $t_l$  instead of  $x_0$  primarily stems from the fact that the derivations can be presented in a much cleaner form; to deal with  $F(x_0)$  one is forced to resort to a combination of  $\hat{F}(t_l)$  and  $\hat{F}(t_r)$  at least for sparser grids ( $\gamma \leq 1/3$ ) and while this in itself does not require any significant technical innovation, the derivations tend to become somewhat more cumbersome and tedious. We focus our attention on three key questions: (a) What is the limit distribution of  $\hat{F}(t_l) - F(t_l)$  under appropriate normalization? (b) What can we say about the asymptotic properties of the likelihood ratio test statistic (LRS) for testing the hypotheses  $H_{0,n} : F(t_l) = \theta_l \leftrightarrow H_{1,n} : F(t_l) \neq \theta_l$  when  $H_{0,n}$  holds? (c) How can the methods for making inference on  $F(t_l)$  be modified to make inference for  $F(x_0)$ ?

The NPMLE of  $F$ , say  $\hat{F}$ , is characterized by the vector

$$\{\hat{F}(t_i)\}_{i=1}^K = \operatorname{argmin}_{u_1 \leq u_2 \leq \dots \leq u_K} \sum_{j=1}^K (\bar{Z}_i - u_i)^2 N_i,$$

where  $N_i$  is the number of  $X_j$ 's that equal  $t_i$  and  $\bar{Z}_i$  is the sample mean of all the  $Y_j$ 's such that  $X_j = t_i$ , or in the other words the sample proportion of patients inspected at time  $t_i$  that are infected by that time. Some mathematics shows that for  $\gamma \in (0, 1/3)$ , with probability increasing to 1, the  $\bar{Z}_i$ 's are ordered in  $i$  and therefore furnish the minimizer of the optimization problem in the above display. Basically, for  $\gamma \in (0, 1/3)$ , the grid is sparse enough so that the naive averages at each inspection time, which provide empirical estimates of  $F$  at those corresponding inspection times, are automatically ordered and there is no 'strength borrowed' from nearby inspection times. It will be shown that  $n^{(1-\gamma)/2} (\hat{F}(t_l) - F(t_l))$  converges to a normal distribution. Note that the rate of convergence is always faster than  $n^{1/3}$  and we are in the setting of 'standard' asymptotics.

On the contrary, when  $\gamma \in (1/3, 1)$ , the customary 'non-standard' asymptotics that are prevalent in the usual treatment of the current status model (as in [Groeneboom and Wellner \(1992\)](#) and [Banerjee and Wellner \(2001\)](#)) take over. Now, successive grid points are close by each other, the naive averages are no longer ordered and isotonic algorithms kick in, producing boundary solutions that are highly non-linear functionals of the empirical distribution and the asymptotic distribution turns out to be drastically different:  $n^{1/3} (\hat{F}(t_l) - F(t_l))$  converges to a multiple, which

depends among other parameters upon  $c$ , of Chernoff's distribution, or equivalently a multiple of the slope at 0 of the greatest convex minorant of the stochastic process  $\{X(t) = W(t) + t^2 : t \in \mathbb{R}\}$  with  $W(t)$  being two sided Brownian motion starting from 0.

The case  $\gamma = 1/3$  acts as the boundary scenario at which the transition between standard and non-standard asymptotics happens and interestingly enough, the behavior of  $\hat{F}(t_l)$  in this case is different from either of the previous cases. Basically, when  $\gamma = 1/3$ , the grid points are 'close enough', so that the naive averages are no longer the best estimates of  $F$ . On the other hand, the resolution of the grid exactly matches  $n^{-1/3}$ , the order of the local neighborhoods of the point  $t_l$  that are asymptotically relevant to the estimation of  $F$  whenever  $\gamma \in [1/3, 1)$ . As a consequence, the limit distribution of  $n^{1/3}(\hat{F}(t_l) - F(t_l))$  is driven, not by a process in continuous time as in the case when  $\gamma \in (1/3, 1)$ , but a process essentially defined on  $c\mathbb{Z}$ , where  $\mathbb{Z}$  is the set of integers. In fact, the relevant limit process can be written down as the linear interpolant of the set of points  $\mathcal{P}_c = \{ck, \alpha W(ck) + \beta c^2 k(1+k) : k \in \mathbb{Z}\}$ , where the constants  $\alpha$  and  $\beta$  depend on the underlying parameters of the problem, and the limit distribution is that of the left derivative at 0 of the greatest convex minorant of this process, denoted subsequently by  $\mathcal{S}_c$ . The process  $\mathcal{P}_c$  can be viewed as a discretized version of the process  $X$ : note that for  $\gamma = 1/3$ , instead of Brownian motion the random component of the relevant process is the restriction of Brownian motion to the set  $c\mathbb{Z}$  and the quadratic drift term  $t^2$ , essentially the integral of the function  $t$ , is replaced by the sum of the first  $k$  integers. As far as the asymptotics of LRS for testing  $H_{0,n} : F(t_l) = \theta_l$  are concerned, for the case  $\gamma \in (0, 1/3)$  we get the usual  $\chi_1^2$  distribution, for  $\gamma \in (1/3, 1)$  we obtain the pivotal limit  $\mathbb{D}$  of Banerjee and Wellner (2001), the same as when the covariates come from a Lebesgue density, and for  $\gamma = 1/3$  we obtain a discrete analogue of  $\mathbb{D}$  which can be written as a functional of  $\mathcal{P}_c$ , though this is no longer pivotal.

In Section 5, we address the rather interesting question of how to make inferences on  $F$  from current status data observed on a regular grid in the time-domain by using the results developed in this paper. We argue that irrespective of the inherent resolution parameter  $\gamma$  of the given grid, which is usually unknown in practice, one can always assume  $\gamma$  to be  $1/3$ , the boundary value, at the cost of adjusting  $c$ . Basically, depending on the true inherent grid resolution, an adjusted data-driven  $c$ , say  $\hat{c}$ , is shown to provide the correct asymptotic quantiles — namely those of

$\mathcal{S}_{\hat{c}}$  — for the distribution of  $n^{1/3}(\hat{F}(t_l) - F(t_l))$ . Thus, the ‘boundary asymptotics’ can be used to approximate both the ‘standard’ ( $\gamma \in (0, 1/3)$ ) and ‘non-standard’ ( $\gamma \in (1/3, 1)$ ) situations and provide an effective means of setting confidence intervals for  $F(t_l)$  without having to contend with the difficult problem of estimating the inherent resolution of the grid. Similar phenomena are observed with the likelihood ratio statistics. From the point of view of conducting effective inference, we view this as the *key contribution* of our work.

## 2.3 Estimation

In this section, we consider the shape-restricted nonparametric maximum likelihood estimator (NPMLE) of  $F$  and the likelihood ratio test statistic (LRS) for testing the value of  $F$  at a point of interest. The characterizations of these estimators are well-known from the current status literature but we include a description tailored for the setting of this paper.

The likelihood function of the data  $\{(X_i, Y_i)\}$  is given by

$$L_n(F) = \prod_{j=1}^n F(X_j)^{Y_j} (1 - F(X_j))^{1-Y_j} p_{\{i: X_j=t_i\}} = \prod_{i=1}^K F_i^{Z_i} (1 - F_i)^{N_i - Z_i} p_i^{N_i},$$

where  $p_{\{i: X_j=t_i\}}$  denotes the probability that  $X_j$  equals a genetic grid point  $t_i$ ,  $F_i$  is an abbreviation for  $F(t_i)$ ,  $N_i = \sum_{j=1}^n \{X_j = t_i\}$  is the number of the observation at  $t_i$ ,  $Z_i = \sum_{j=1}^n Y_j \{X_j = t_i\}$  is the sum of the responses at  $t_i$ ,  $\{\cdot\}$  stands for both a set and its indicator function with the meaning depending on the context of use, and  $F$  is generically understood as either a distribution or the vector  $(F_1, F_2, \dots, F_K)$ , which sometimes is also written as  $\{F_i\}_{i=1}^K$ . Then, the log-likelihood function  $\log(L_n(F))$  is given by

$$l_n(F) = \sum_{i=1}^K N_i \log p_i + \sum_{i=1}^K \{[\bar{Z}_i \log F_i + (1 - \bar{Z}_i) \log(1 - F_i)] N_i\},$$

where  $\bar{Z}_i = Z_i/N_i$  is the average of the responses at  $t_i$ .

Denote the basic shape-restricted maximizer as

$$\{F_i^*\}_{i=1}^K = \operatorname{argmax}_{F_1 \leq \dots \leq F_K} l_n(F).$$



From the theory of isotonic regression (see, for example, [Robertson et al. \(1988\)](#)), we have

$$\operatorname{argmax}_{F_1 \leq \dots \leq F_K} l_n(F) = \operatorname{argmin}_{F_1 \leq \dots \leq F_K} \sum_{i=1}^K [(\bar{Z}_i - F_i)^2 N_i].$$

Thus,  $\{F_i^*\}_{i=1}^K$  is the weighted isotonic regression of  $\{\bar{Z}_i\}_{i=1}^K$  with weights  $\{N_i\}_{i=1}^K$ , and exists uniquely. We conventionally define the shape-restricted NPMLE of  $F$  as the following right-continuous step function on  $[a, b]$ :

$$\hat{F}(t) = \begin{cases} 0 & \text{if } t \in [a, t_1); \\ F_i^*, & \text{if } t \in [t_i, t_{i+1}), i = 1, \dots, K-1; \\ F_K^*, & \text{if } t \in [t_K, b]. \end{cases} \quad (2.1)$$

A popular and straightforward method of setting confidence intervals for  $F$  at a point of interest  $t \in (a, b)$  is first to derive the asymptotic distribution of  $\hat{F}(t)$  and then to construct for  $F(t)$  the so-called Wald-type confidence intervals of the form  $\hat{F}(t)$  plus and minus terms that depend on both the level of confidence and the asymptotic distribution. Another popular but slightly more involved approach to construct confidence intervals is through the inversion of a likelihood ratio test for the value of  $F(t)$ . More specifically, we first derive the asymptotic null distribution of the likelihood ratio test statistic (LRS) for the following hypothesis testing problem:

$$H_0 : F(t) = \theta \leftrightarrow H_1 : F(t) \neq \theta, \quad (2.2)$$

where  $t \in (a, b)$  and  $\theta \in (0, 1)$  and then obtain the so-called LR-type confidence intervals via inversion. Note that both  $t$  and  $\theta$  can depend on  $n$ . One specific instance of interest in this paper is that  $t = t_l$  and  $\theta = \theta_l = F(t_l)$ . In this case, the hypotheses (2.2) become

$$H_{0,n} : F(t_l) = \theta_l \leftrightarrow H_{1,n} : F(t_l) \neq \theta_l. \quad (2.3)$$

To construct the LRS, we next consider the constrained shape-restricted NPMLE of  $F$  under the null hypothesis of (2.3). Define

$$\{F_i^{*o}\} = \operatorname{argmax}_{F_1 \leq \dots \leq F_l = \theta_l \leq F_{l+1} \leq \dots \leq F_K} l_n(F).$$

It is well known that  $\{F_i^{*o}\}$  is well defined and

$$\begin{aligned} \{F_i^{*o}\}_{i=1}^{l-1} &= \theta_l \wedge \operatorname{argmin}_{F_1 \leq \dots \leq F_{l-1}} \sum_{i=1}^{l-1} [(\bar{Z}_i - F_i)^2 N_i], \\ \{F_i^{*o}\}_{i=r}^K &= \theta_l \vee \operatorname{argmin}_{F_r \leq \dots \leq F_K} \sum_{i=r}^K [(\bar{Z}_i - F_i)^2 N_i], \end{aligned}$$

where the minimum and maximum operators ( $\wedge$  and  $\vee$ ) are interpreted as being taken component-wise. Specifically,  $F_l^{*o} = \theta_l$  as required by the null hypothesis. Thus, the constrained NPMLE of  $F$ , similar to the unconstrained one, can be defined as the following right-continuous step function on  $[a, b]$ :

$$\hat{F}^o(t) = \begin{cases} 0 & t \in [a, t_1); \\ F_i^{*o}, & t \in [t_i, t_{i+1}), i = 1, \dots, K-1; \\ F_K^{*o}, & t \in [t_K, b]. \end{cases} \quad (2.4)$$

For the details underlying the above characterization of the constrained estimator of  $F$ , we refer the readers to [Banerjee \(2000\)](#) and [Banerjee and Wellner \(2001\)](#). Thus, the LRS is given by

$$2 \log \lambda_n = 2[l_n(\hat{F}) - l_n(\hat{F}^o)]. \quad (2.5)$$

An asymptotic  $1 - \eta$  confidence interval for  $\theta_l$  is given by the set of all  $0 < \theta < 1$  such that  $2 \log \lambda_n(\theta)$ , the LRS for testing  $H_{0,n} : F(t_l) = \theta$  versus its complement, lies below the  $(1 - \eta)$ 'th quantile of the limit distribution of the LRS under  $H_{0,n} : F(t_l) = \theta_l$ .

Next, we provide characterizations of  $\hat{F}$  and  $\hat{F}^o$  as slopes of *the greatest convex minorants* (GCMs) of random processes, which prove useful for deriving the asymptotics for  $\gamma \in [1/3, 1)$ . First, consider the characterization of  $\hat{F}$ . Define, for  $t \in [a, b]$ ,

$$G_n(t) = \mathbb{P}_n\{x \leq t\}, \quad V_n(t) = \mathbb{P}_n y\{x \leq t\}, \quad (2.6)$$

where  $\mathbb{P}_n$  is the empirical probability measure based on the data  $\{(X_i, Y_i)\}$ . Then, we have, for each  $x \in [a, b]$ ,

$$\hat{F}(x) = LS \circ GCM \{(G_n(t), V_n(t)), t \in [a, b]\} (G_n(x)). \quad (2.7)$$

Here,  $GCM(\cdot)$  denotes the greatest convex minorant of a set of points in  $\mathbb{R}^2$ . For any finite collection of points in  $\mathbb{R}^2$ , its GCM is a continuous piecewise linear convex

function and  $LS(\cdot)$  denotes the *left slope or derivative function* of a convex function. The term GCM will also be generically used in connection with functions from  $\mathbb{R} \mapsto \mathbb{R}$ . For such a function  $H$ ,  $GCM(H)$  will denote the greatest convex minorant of  $H$ .

Next, consider the characterization of  $\hat{F}^o$ . Define, for  $s \in [a - t_l, b - t_l]$ ,

$$\tilde{G}_n(s) = G_n(t_l + s) - G_n(t_l), \quad \tilde{V}_n(s) = V_n(t_l + s) - V_n(t_l). \quad (2.8)$$

Then, we have, for  $s \in [a - t_l, 0)$ ,

$$\hat{F}_l^o(t_l + s) = \theta_l \wedge LS \circ GCM \left\{ (\tilde{G}_n(u), \tilde{V}_n(u)), u \in [a - t_l, 0) \right\} (\tilde{G}_n(s)) \quad (2.9)$$

and for  $s \in [t_r - t_l, b - t_l)$ ,

$$\hat{F}_r^o(t_l + s) = \theta_l \vee LS \circ GCM \left\{ (\tilde{G}_n(u), \tilde{V}_n(u)), u \in [0, b - t_l] \right\} (\tilde{G}_n(s)). \quad (2.10)$$

Therefore, we have, for  $s \in [a - t_l, b - t_l]$ ,

$$\begin{aligned} \hat{F}^o(t_l + s) &= \hat{F}_l^o(t_l + s) \{s \in [a - t_l, 0)\} \\ &\quad + \hat{F}_r^o(t_l + s) \{s \in [t_r - t_l, b - t_l]\} + \theta_l \{s \in [0, t_r - t_l]\}. \end{aligned} \quad (2.11)$$

The above characterizations will be further exploited in the next section.

## 2.4 Asymptotic Results

In this section, we primarily study the asymptotics of  $\hat{F}$  and  $2 \log(\lambda_n)$  for three cases with different values of  $\gamma \in (0, 1)$ . We first consider Case One  $\gamma \in (0, 1/3)$ , then Case Two  $\gamma \in (1/3, 1)$  and finally Case Three  $\gamma = 1/3$ .

### 2.4.1 Case One $\gamma \in (0, 1/3)$

In this subsection, we consider the asymptotics for  $\gamma \in (0, 1/3)$ . First, we state further technical assumptions :

**(A1.1)** There exists a constant  $f_l > 0$  such that  $f(x) > f_l$  for every  $x \in [a, b]$ .

**(A1.2)** There exists a constant  $g_l > 0$  such that  $g(x) \geq g_l$  for every  $x \in [a, b]$ .

**(A1.3)** Assume  $a' < a$  and  $F(a) > 0$ .

We denote the above assumptions together as **(A1)**. The assumptions **(A1.1)** and **(A1.2)** require that both  $f$  and  $g$  have lower bounds ‘globally’. The assumption **(A1.3)** is technically tailored to establish Proposition 2.4.9, the key technical tool behind the arguments in this subsection. In fact, the strict assumption **(A1.3)** can be replaced by a weaker one which allows  $a' = a$  but requires that there exists  $d \in (a, x_0)$  such that  $F(d) > 0$ . With this weaker assumption, Proposition 2.4.9 needs some technical adjustments but all the theorems in this subsection still hold without any modification. To emphasize the main idea of the argument, without having to deal with the aforementioned adjustments, we retain the strong assumption **(A1.3)**. By **(A1.2)**, each discrete probability  $p_i \geq g_l \delta$ . Denote  $m_l = ng_l \delta = g_l cn^{1-\gamma}$ , which can be interpreted as the minimum average number of observations at a grid point. The following lemma shows the consistency of  $F_l^*$ ,  $F_r^*$  and  $\hat{F}(x_0)$ .

*Lemma 2.4.1.* If  $\gamma \in (0, 1/3)$  and **(A1)** holds, we have

$$F_l^* - F(t_l) \xrightarrow{P} 0, \quad F_r^* - F(t_r) \xrightarrow{P} 0, \quad \text{and} \quad \hat{F}(x_0) \xrightarrow{P} F(x_0).$$

The joint limiting distribution of  $\hat{F}(t_l)$  and  $\hat{F}(t_r)$  is described below.

*Theorem 2.4.2.* If  $\gamma \in (0, 1/3)$  and **(A1)** holds, we have

$$\left( \sqrt{N_l}(\hat{F}(t_l) - F(t_l)), \sqrt{N_r}(\hat{F}(t_r) - F(t_r)) \right) \xrightarrow{d} \sqrt{F(x_0)(1 - F(x_0))} N(0, I_2),$$

where  $I_2$  is the  $2 \times 2$  identity matrix.

*Remark 2.4.3.* From Theorem 2.4.2, the quantities  $\hat{F}(t_l)$  and  $\hat{F}(t_r)$  with proper centering and scaling are asymptotically uncorrelated and independent. In fact, they are essentially the averages of the responses at the two grid points  $t_l$  and  $t_r$  and are therefore based on responses corresponding to different sets of individuals. Consequently, there is no dependence between them in the long run. Intuitively speaking,  $\gamma \in (0, 1/3)$  corresponds to very sparse grids with successive grid points *far enough* so that the responses at different grid points fail to influence each other.

Note that  $N_l/(np_l)$  converges to 1 in probability and that  $np_l/cg(x_0)n^{1-\gamma}$  converges to 1 for  $\gamma \in (0, 1/3)$ . Then the result of Theorem 2.4.2 can be rewritten as follows:

$$\left( n^{(1-\gamma)/2}(\hat{F}(t_l) - F(t_l)), n^{(1-\gamma)/2}(\hat{F}(t_r) - F(t_r)) \right) \xrightarrow{d} \alpha c^{-\frac{1}{2}} N(0, I_2), \quad (2.12)$$

where  $\alpha = \sqrt{F(x_0)(1 - F(x_0))/g(x_0)}$ . This formulation will be used later and the

parameter  $\alpha$  will be seen to play a critical role in the asymptotic behavior of  $\hat{F}(t_i)$  when  $\gamma \in [1/3, 1)$  as well.

Next, we consider the asymptotics of the log-likelihood ratio test statistic (LRS)  $2 \log \lambda_n$  for testing the hypotheses (2.3).

*Theorem 2.4.4.* If  $\gamma \in (0, 1/3)$  and **(A1)** holds, under the null hypothesis of (2.3), we have  $2 \log \lambda_n \xrightarrow{d} \chi_1^2$ .

From Theorem 2.4.2 and Theorem 2.4.4, we see that the asymptotic distributions are Normal and Chi-Squared, standard limit distributions for parametric problems.

In the following, we discuss testing and estimating  $F(x_0)$ . Denote  $2 \log \lambda'_n$  as the LRS based on  $\hat{F}$  and  $\hat{F}^o$  for the hypotheses (2.2) with  $t = x_0$  and  $\theta = \theta_0$ . Then, this test statistic may not have a proper limit distribution, as shown by the following theorem.

*Theorem 2.4.5.* Suppose  $\gamma \in (0, 1/3)$  and **(A1)** holds. If there exists  $\epsilon > 0$  and a subsequence  $\{n_k\}$  of  $\{n\}$  such that  $\min(x_0 - t_l, t_r - x_0)/\delta > \epsilon$ , then under the null hypothesis of (2.2) with  $t = x_0$  and  $\theta = \theta_0$ , we have  $2 \log \lambda'_{n_k} \xrightarrow{P} 0$  as  $k \rightarrow \infty$ .

An example satisfying the conditions of Theorem 2.4.5 is as follows. Suppose  $a = 0$ ,  $x_0 = 1/2$ ,  $c = 1$ ,  $\gamma = 1/4$  and  $n_k = (2k + 1)^4$  for  $k \in \mathbb{N}$ . Then  $(x_0 - t_l)/\delta = (t_r - x_0)/\delta = 1/2 > 0$ . In fact, this example also concretely explains why we do not make inference about  $F(x_0)$  using  $\hat{F}(x_0)$ . The basic reason is that there may not exist any scaling such that the scaled difference  $\hat{F}(x_0) - F(x_0)$  converges to a tight non-degenerate random limit. More specifically, assume the setting of the above example and further suppose  $F(t) = t$  for  $t \in [0, 1]$ . Then, we have  $\hat{F}(x_0) - F(x_0) = (\hat{F}(t_l) - t_l) + (t_l - x_0)$ . By (2.12), the proper scaler for the first term  $\hat{F}(t_l) - t_l$  is  $n^{3/8}$  and the limit distribution is normal. However, this scaler times the second term  $(x_0 - t_l)$  gives  $\sqrt{2k - 1}/2$  (along  $\{n_k\}$ ), which diverges to infinity as  $k$  goes to infinity.

**Inference for  $F(x_0)$ :** It turns out, however, that inference for  $F(x_0)$  can still be made provided the estimator  $\hat{F}$  is altered slightly. Recall that  $\hat{F}$  is defined as the right-continuous step function that assumes the value  $F_i^*$  at  $t_i$ . Consider now the piecewise linear interpolant of the  $F_i^*$ 's. More precisely, the modified estimator is

given by

$$\tilde{F}(t) = \begin{cases} 0, & \text{if } t \in [a, t_1) \\ \frac{t_{i+1}-t}{t_{i+1}-t_i} F_i^* + \frac{t-t_i}{t_{i+1}-t_i} F_{i+1}^*, & \text{if } t \in [t_i, t_{i+1}) \text{ and } i = 1, 2, \dots, K-1; \\ F_K^*, & \text{if } t \in [t_K, b]. \end{cases}$$

Then, we can estimate  $F(x_0)$  by  $\tilde{F}(x_0)$ , whose consistency is ensured by the following lemma.

*Lemma 2.4.6.* If  $\gamma \in (0, 1/3)$  and **(A1)** holds, we have  $\tilde{F}(x_0) \xrightarrow{P} F(x_0)$ .

Define two proportions  $p_n = (x_0 - t_l)/(t_r - t_l)$  and  $q_n = 1 - p_n$  and a properly normalized random quantity

$$R_{\tilde{F}} = \frac{1}{\sqrt{p_n^2 + q_n^2}} \sqrt{\frac{N_l + N_r}{2}} (\tilde{F}(x_0) - F(x_0)).$$

The asymptotic distribution of  $\tilde{F}(x_0)$  is the content of the next theorem.

*Theorem 2.4.7.* If  $\gamma \in (0, 1/3)$ , **(A1)** holds, and  $f''$  is bounded in a neighborhood of  $x_0$ , we have, for  $\gamma \in (1/5, 1/3)$ ,

$$R_{\tilde{F}} \xrightarrow{d} \sqrt{F(x_0)(1 - F(x_0))} Z.$$

Further, if  $\{n_k\}$  is a subsequence of  $\{n\}$  such that  $p_{n_k}$  converges to some  $p \in (0, 1)$ , we have, along this subsequence and for  $\gamma = 1/5$ ,

$$R_{\tilde{F}} \xrightarrow{d} \sqrt{F(x_0)(1 - F(x_0))} Z + \frac{1}{2} \frac{pq}{\sqrt{p^2 + q^2}} g(x_0)^{1/2} c^{5/2} f'(x_0).$$

Furthermore, if  $f'(x_0) \neq 0$ , we have, for  $\gamma \in (0, 1/5)$ ,

$$R_{\tilde{F}} \xrightarrow{P} \text{Sign}(f'(x_0))\infty,$$

where  $q = 1 - p$  and  $Z$  follows  $N(0, 1)$ .

*Remark 2.4.8.* Since  $N_l/N_r$  converges in probability to 1,  $R_{\tilde{F}}$  is asymptotically equivalent to  $(p_n^2 + q_n^2)^{-1/2} \sqrt{N_i} (\tilde{F}(x_0) - F(x_0))$  with  $i = l$  or  $r$ . Note that the coefficient  $(p_n^2 + q_n^2)^{-1/2}$  lies between 1 and  $\sqrt{2}$ . Comparing the asymptotic results of Theorem 2.4.7 with  $\gamma \in (1/5, 1/3)$  and that of Theorem 2.4.2 reveals that  $\tilde{F}(x_0)$  effectively combines the data around  $x_0$  to successfully achieve smaller asymptotic variance than

$\hat{F}(t_l)$  and  $\hat{F}(t_r)$ .

### 2.4.1.1 Proofs

The proofs of the above lemmas and theorems rely heavily on some fundamental propositions that we state below and prove in the appendix. In what follows we will need to consider the vector of averages of responses over the grid-points:  $\{\bar{Z}_i\}_{i=1}^k$ . Since  $\bar{Z}_i$  is not defined when  $N_i = 0$ , to avoid ambiguity we set  $\bar{Z}_i = 0$  whenever this happens. This can be done without affecting the asymptotic results, since the probability of the event  $\{N_i > 0, i = 1, 2, \dots, K\}$  goes to 1. See, for example, Lemma 2.8.1 in the appendix where a stronger fact is established.

*Proposition 2.4.9.* If  $\gamma \in (0, 1/3)$  and **(A1)** holds, we have

$$P(\bar{Z}_1 \leq \bar{Z}_2 \leq \dots \leq \bar{Z}_K) \rightarrow 1.$$

The above proposition says that on a sequence of sets with probability increasing to 1, the vector  $\{\bar{Z}_i\}_{i=1}^k$  is ordered and therefore the isotonization algorithm involved in finding the NPMLE of  $F$  yields  $\{F_i^*\}_{i=1}^K = \{\bar{Z}_i\}_{i=1}^K$  on these sets. In other words, asymptotically, isotonization has no effect and the naive estimates obtained by averaging the responses at each grid point produce the NPMLE. This lemma is really at the heart of the asymptotic derivations for  $\gamma < 1/3$  because it effectively reduces the problem of studying the  $F_i^*$ 's which are obtained through a complex non-linear algorithm to the study of the asymptotics of the  $\bar{Z}_i$ , which are linear statistics and can be handled readily using standard central limit theory. For the following lemmas, we use weaker assumptions (than **(A1)** on  $f$  and  $g$ ):

**(A1.1')**  $f(\cdot)$  has a positive lower bound in a neighborhood of  $x_0$ .

**(A1.2')**  $g(\cdot)$  has a positive lower bound in a neighborhood of  $x_0$ .

We refer to these two assumptions together as **(A1')**.

*Proposition 2.4.10.* If  $\gamma \in (0, 1/3)$  and **(A1')** hold, we have

$$P(\bar{Z}_{l-1} \leq F_l \leq \bar{Z}_r) \rightarrow 1.$$

Now we state a further weaker assumption on  $F$ .

**(A1.1'')**  $F(\cdot)$  is positive and continuous in a neighborhood of  $x_0$ .

*Proposition 2.4.11.* If  $\gamma \in (0, 1)$  and **(A1.1'')** and **(A1.2')** hold, we have

$$(\bar{Z}_l - F_l, \bar{Z}_r - F_r) \xrightarrow{P} 0,$$

$$\left(\sqrt{N_l}(\bar{Z}_l - F_l), \sqrt{N_r}(\bar{Z}_r - F_r)\right) \xrightarrow{d} \sqrt{F(x_0)(1 - F(x_0))}N(0, I_2).$$

*Proof of Lemma 2.4.1.* We only show  $\hat{F}(t_l) - F(t_l) \rightarrow 0$  in probability. The other two can be shown similarly. We have

$$\hat{F}(t_l) - F(t_l) = (F_l^* - F_l)B_n + (F_l^* - F_l)(1 - B_n) =: T_1 + T_2,$$

where  $B_n = \{\bar{Z}_1 \leq \dots \leq \bar{Z}_K\}$  and ‘=:’ means the right-hand side notations are defined by the left-hand terms. By Proposition 2.4.9,  $B_n$  converges to 1 in probability. Thus,  $T_2$  converges to 0 in probability. On the other hand, by Proposition 2.4.11,  $\bar{Z}_l - F_l$  converges to 0 in probability. On  $B_n$ , the  $\bar{Z}_i$ ’s are already monotone, and the vector  $\{F_i^*\}$  obtained via isotonization coincides with  $\{\bar{Z}_i\}$ . We conclude that  $T_1 = (\bar{Z}_l - F_l) B_n \xrightarrow{P} 0 \cdot 1 = 0$ . So,  $\hat{F}(x_0)$  is a consistent estimator of  $F(x_0)$ .  $\square$

*Proof of Theorem 2.4.2.* Denote  $T_{nl} = \sqrt{N_l}(\hat{F}(t_l) - F(t_l))$  and  $T_{nr} = \sqrt{N_r}(\hat{F}(t_r) - F(t_r))$ . Then, we have  $(T_{nl}, T_{nr}) = S_1 + S_2$ , where  $S_2$  and  $S_1$  are  $(T_{nl}, T_{nr})(1 - B_n)$  and  $(\sqrt{N_l}(\bar{Z}_l - F_l), \sqrt{N_r}(\bar{Z}_r - F_r)) B_n$ , respectively, with  $B_n$  as in the preceding proof. By Proposition 2.4.9,  $B_n$  converges to 1 in probability. Then,  $S_2$  converges to 0 in probability. On the other hand, by Proposition 2.4.11 and Slutsky’s Lemma, we have  $S_1 \xrightarrow{d} \sqrt{F(x_0)(1 - F(x_0))}N(0, I_2) \cdot 1$ .  $\square$

*Proof of Theorem 2.4.4.* We have

$$2 \log \lambda_n = (2 \log \lambda_n)C_n + (2 \log \lambda_n)(1 - C_n),$$

where  $C_n = B_n\{\bar{Z}_{l-1} \leq \theta_l \leq \bar{Z}_r\}$  with  $B_n$  as in the preceding proof. By Proposition 2.4.9 and Proposition 2.4.10,  $C_n$  converges to 1 in probability under the null hypothesis. Thus, the second term above converges to 0 in probability. Denote the first term above as  $T_1$ . Then, we have

$$T_1 = 2 \left\{ \left[ \bar{Z}_l \log \left( \frac{\bar{Z}_l}{\theta_l} \right) + (1 - \bar{Z}_l) \log \left( \frac{1 - \bar{Z}_l}{1 - \theta_l} \right) \right] N_l \right\} C_n =: T_{11}C_n.$$

By routine Taylor expansions of  $\log(\bar{Z}_l)$  around  $\theta_l$  and  $\log(1 - \bar{Z}_l)$  around  $1 - \theta_l$  up



to a third order term and some straightforward algebra, we have

$$T_{11} = \frac{N_l(\bar{Z}_l - \theta_l)^2}{\theta_l(1 - \theta_l)} + o_p(1).$$

Note that  $\theta_l \rightarrow F(x_0)$ . Finally, by Proposition 2.4.11, we have  $T_{11} \xrightarrow{d} 1 \cdot \chi_1^2 + 0 = \chi_1^2$ , which completes the proof.  $\square$

*Proof of Theorem 2.4.5.* We have  $2 \log \lambda'_{n_k} = 2 \log \lambda'_{n_k} A_{n_k}$  where  $A_n = 1 - \{\bar{Z}_1 \leq \dots \leq \bar{Z}_K\} \{\bar{Z}_l \leq \theta_0 \leq \bar{Z}_r\}$ , since the likelihood ratio is identically 1 whenever  $A_n = 0$ . Now,  $P(\bar{Z}_l \leq \theta_0 \leq \bar{Z}_r) \rightarrow 1$  under the null hypothesis and the conditions of the theorem. To see this, consider  $P(\bar{Z}_l \leq \theta_0) = P(n^{(1-\gamma)/2}(\bar{Z}_l - F(t_l)) \leq n^{(1-\gamma)/2}(F(x_0) - F(t_l)))$ . Now,  $n^{(1-\gamma)/2}(\bar{Z}_l - F(t_l))$  is  $O_p(1)$  and  $n^{(1-\gamma)/2}(F(x_0) - F(t_l))$  is easily seen to be larger than a constant times  $n^{(1-\gamma)/2} \delta$  along the subsequence  $\{n_k\}$ . Since  $\delta$  is of order  $n^{-\gamma}$  and  $1 - 3\gamma > 0$ , it follows that  $n^{(1-\gamma)/2}(F(x_0) - F(t_l))$  diverges to  $\infty$  along  $\{n_k\}$ . So  $P(\bar{Z}_l \leq \theta_0) \rightarrow 1$  along  $\{n_k\}$  and a similar argument shows that  $P(\theta_0 \leq \bar{Z}_r) \rightarrow 1$ .

This, together with Proposition 2.4.9, implies that  $A_{n_k}$  converges to 0 in probability under the null hypothesis. Therefore, we have  $2 \log \lambda'_{n_k} \rightarrow 0$  in probability.  $\square$

For the proofs of Lemma 2.4.6 and Theorem 2.4.7, see the appendix.

## 2.4.2 Case Two $\gamma \in (1/3, 1)$

We next deal briefly with the case  $\gamma \in (1/3, 1)$  before considering the most interesting boundary case  $\gamma = 1/3$ . Our treatment will be condensed since the asymptotics for this case follow the same patterns as when the observation times possess a Lebesgue density. The notations, however, are developed in detail and will also be used for the boundary case.

For this case, we assume that both  $f$  and  $g$  are positive and continuously differentiable in a neighborhood of  $x_0$ . In order to study the asymptotics of the isotonic regression estimator  $\hat{F}(t_l)$  of  $F(t_l)$  and the likelihood ratio statistic for testing the hypotheses (2.3), the following localized processes will be of interest: for

$u \in I_n = [(a - t_l)n^{1/3}, (b - t_l)n^{1/3}]$ , define

$$\mathbb{X}_n(u) = n^{1/3}(\hat{F}(t_l + un^{-1/3}) - F(t_l)), \quad (2.13)$$

$$\mathbb{Y}_n(u) = n^{1/3}(\hat{F}^o(t_l + un^{-1/3}) - F(t_l)). \quad (2.14)$$

Define subsets of  $I_n$  as follows:  $I_n^l = [(a - t_l)n^{1/3}, 0)$ ,  $I_n^m = [0, (t_r - t_l)n^{1/3})$ ,  $I_n^r = [(t_r - t_l)n^{1/3}, (b - t_l)n^{1/3}]$  and  $\tilde{I}_n^r = I_n^r \cup I_n^m = [0, (b - t_l)n^{1/3}]$ . Note that  $I_n = I_n^l \cup I_n^m \cup I_n^r$  goes to  $\mathbb{R}$ ,  $I_n^m$  goes to  $\{0\}$ , and  $\tilde{I}_n^r$  is a left-side enlargement of  $I_n^r$ . Next, define the following normalized processes on  $I_n$ :

$$G_n^*(h) = g(x_0)^{-1}n^{1/3}\tilde{G}_n(hn^{-1/3}), \quad (2.15)$$

$$V_n^*(h) = g(x_0)^{-1}n^{2/3}\left(\tilde{V}_n(hn^{-1/3}) - F(t_l)\tilde{G}_n(hn^{-1/3})\right), \quad (2.16)$$

where  $\tilde{G}_n$  and  $\tilde{V}_n$  are defined in (2.8). After some straight forward algebra, from (2.7), (2.11), (2.13), and (2.14), we have the following technically useful characterizations of  $\mathbb{X}_n$  and  $\mathbb{Y}_n$ : for  $u \in I_n$ ,

$$\mathbb{X}_n(u) = LS \circ GCM \{(G_n^*(h), V_n^*(h)), h \in I_n\} (G_n^*(u)); \quad (2.17)$$

$$\mathbb{Y}_n(u) = \begin{cases} 0 \wedge LS \circ GCM \{(G_n^*(h), V_n^*(h)), h \in I_n^l\} (G_n^*(u)), & \text{if } u \in I_n^l, \\ 0, & \text{if } u \in I_n^m, \\ 0 \vee LS \circ GCM \{(G_n^*(h), V_n^*(h)), h \in \tilde{I}_n^r\} (G_n^*(u)), & \text{if } u \in I_n^r. \end{cases} \quad (2.18)$$

For constants  $\kappa_1 > 0$  and  $\kappa_2 > 0$ , denote

$$X_{\kappa_1, \kappa_2}(h) = \kappa_1 W(h) + \kappa_2 h^2, \text{ for } h \in \mathbb{R},$$

$$X_{\kappa_1, \kappa_2}^l(h) = \kappa_1 W(h) + \kappa_2 h^2, \text{ for } h \in (-\infty, 0],$$

$$X_{\kappa_1, \kappa_2}^r(h) = \kappa_1 W(h) + \kappa_2 h^2, \text{ for } h \in [0, \infty),$$

where  $W$  is a two-sided Brownian motion with  $W(0) = 0$ . Let  $G_{\kappa_1, \kappa_2}$ ,  $G_{\kappa_1, \kappa_2}^l$  and  $G_{\kappa_1, \kappa_2}^r$  be the greatest convex minorants of  $X_{\kappa_1, \kappa_2}$ ,  $X_{\kappa_1, \kappa_2}^l$  and  $X_{\kappa_1, \kappa_2}^r$ , respectively.

Then, define, for  $h \in \mathbb{R}$ ,

$$\begin{aligned} g_{\kappa_1, \kappa_2}(h) &= LS(G_{\kappa_1, \kappa_2})(h), \\ g_{\kappa_1, \kappa_2}^o(h) &= (0 \wedge LS(G_{\kappa_1, \kappa_2}^l)(h))\{h \in (-\infty, 0)\} \\ &\quad + (0 \vee LS(G_{\kappa_1, \kappa_2}^r)(h))\{h \in (0, \infty)\}. \end{aligned}$$

Denote  $\alpha = \sqrt{F(x_0)(1 - F(x_0))/g(x_0)}$  and  $\beta = f(x_0)/2$ . Let  $\mathcal{L}^p[-C, C]$  denote the class of Borel measurable real-valued functions defined on  $[-C, C]$  that possess a finite  $p$ 'th moment with respect to Lebesgue measure on  $[-C, C]$ . Let ' $\rightsquigarrow$ ' denote 'weak convergence' in addition to  $\xrightarrow{d}$ . With these notations, we have the following theorem on the joint distributional convergence of  $(\mathbb{X}_n, \mathbb{Y}_n)$ .

*Theorem 2.4.12* (Weak Convergence of  $(\mathbb{X}_n, \mathbb{Y}_n)$ ). Suppose  $f$  and  $g$  are positive and continuously differentiable in a neighborhood of  $x_0$ . Then, the finite dimensional marginals of the process  $(\mathbb{X}_n, \mathbb{Y}_n)$  converge weakly to those of the process  $(g_{\alpha, \beta}, g_{\alpha, \beta}^o)$ . Furthermore,

$$\{(\mathbb{X}_n(u), \mathbb{Y}_n(u)), u \in [-C, C]\} \rightsquigarrow \{(g_{\alpha, \beta}(u), g_{\alpha, \beta}^o(u)), u \in [-C, C]\}$$

in  $(\mathcal{L}^p[-C, C])^2$  for each  $p \geq 1$  and  $C > 0$ .

*Remark 2.4.13.* Note that  $\mathbb{X}_n(0) = n^{1/3}(\hat{F}(t_l) - F(t_l))$ . Then, by Theorem 2.4.12, it converges in distribution to  $g_{\alpha, \beta}(0)$ . By the Brownian scaling results on Page 1724 of [Banerjee and Wellner \(2001\)](#), we have, for  $h \in \mathbb{R}$ ,

$$(g_{\alpha, \beta}(h), g_{\alpha, \beta}^o(h)) \stackrel{d}{=} (\alpha^2 \beta)^{1/3} (g_{1,1}((\beta/\alpha)^{2/3}h), g_{1,1}^o((\beta/\alpha)^{2/3}h))$$

Hence, by noting  $g_{1,1}(0) \stackrel{d}{=} 2\mathcal{Z}$ , we have a concise result:

$$n^{1/3}(\hat{F}(t_l) - F(t_l)) \xrightarrow{d} (\alpha^2 \beta)^{1/3} g_{1,1}(0) \stackrel{d}{=} \left( \frac{4f(x_0)F(x_0)(1 - F(x_0))}{g(x_0)} \right)^{1/3} \mathcal{Z}. \quad (2.19)$$

Thus, the limit distribution of  $\hat{F}(t_l)$  is exactly the same as one would encounter in the current status model with survival distribution  $F$  and the observation times drawn from a Lebesgue density function  $g$ .

**Inference for  $F(x_0)$  :** Note that  $n^{1/3}(\hat{F}(x_0) - F(x_0)) = \mathbb{X}_n(0) + n^{1/3}(F(t_l) - F(x_0))$ . Since, under the conditions of Theorem 2.4.12,  $n^{1/3}(F(t_l) - F(x_0))$  converges to 0 for

$\gamma \in (1/3, 1)$ , we conclude that  $n^{1/3}(\hat{F}(x_0) - F(x_0))$  converges in distribution to the right-hand side of (2.19), whose distribution is well-characterized.

The following result states the asymptotic distribution of the LRS  $2 \log \lambda_n$  for testing the hypotheses (2.3).

*Theorem 2.4.14* (Weak Convergence of  $2 \log \lambda_n$ ). Under the null hypothesis in (2.3), i.e.  $F(t_l) = \theta_l$ , we have

$$2 \log \lambda_n \xrightarrow{d} \mathbb{D} = \int_{\mathbb{R}} (g_{1,1}(u)^2 - g_{1,1}^o(u)^2) du.$$

*Remark 2.4.15.* This is the same limit distribution as obtained in [Banerjee and Wellner \(2001\)](#) under a Lebesgue density assumption on the observation times. This distribution also appears in [Banerjee \(2007\)](#) in connection with likelihood ratio tests in general monotone regression models. The proofs of the above theorems are omitted as they can be established via arguments similar to those in [Banerjee \(2007\)](#) using continuous mapping for slopes of greatest convex minorants. Note that the limit distribution of the likelihood ratio statistic is free of nuisance parameters.

*Remark 2.4.16.* Results similar to [Theorem 2.4.12](#) and [Theorem 2.4.14](#) hold for the case  $\gamma \geq 1$ . More specifically, the slight difference is that it is more convenient to consider  $F(x_0)$  directly instead of  $F(t_l)$  for the case  $\gamma \geq 1$ . This, together with [Theorem 2.4.12](#) and [Theorem 2.4.14](#), means that a discrete random observation time with a dense enough grid, i.e.  $\gamma > 1/3$ , leads to the same asymptotics as a continuous random observation time.

### 2.4.3 Case Three $\gamma = 1/3$

Now we consider the most interesting boundary case  $\gamma = 1/3$ . For this case, we assume as in the previous section that both  $f$  and  $g$  are positive and continuously differentiable in a neighborhood of  $x_0$ . Let the localized processes  $\mathbb{X}_n(u)$  and  $\mathbb{Y}_n(u)$  be exactly as in [Subsection 2.4.2](#) on Case Two  $\gamma \in (1/3, 1)$ . Since the order of the unit grid spacing  $\delta$  is exactly  $n^{-1/3}$ , equal to the order of the increment  $un^{-1/3}$  in  $\tilde{G}_n$  and  $\tilde{V}_n$ ,  $\mathbb{X}_n$  and  $\mathbb{Y}_n$  have potential jumps only at  $ci$  for  $i \in \mathcal{I}_n = (I_n/c) \cap \mathbb{Z}$ . Thus, it is equivalent to consider  $\mathbb{X}_n$  and  $\mathbb{Y}_n$  on those  $ci$ 's. Define  $\mathcal{I}_n^l = (I_n^l/c) \cap \mathbb{Z}$  and  $\mathcal{I}_n^m, \mathcal{I}_n^r$  and  $\tilde{\mathcal{I}}_n^r$  in analogous fashion. Note that  $\mathcal{I}_n, \mathcal{I}_n^l, \mathcal{I}_n^r$  and  $\tilde{\mathcal{I}}_n^r$  go to  $\mathbb{Z}, -\mathbb{N}, \mathbb{N}$  and  $\{0\} \cup \mathbb{N}$ , respectively, as  $n$  goes to infinity and that  $\mathcal{I}_n^m$  always equals  $\{0\}$ . We then have, for

$i \in \mathcal{I}_n$ ,

$$\mathbb{X}_n(ci) = LS \circ GCM \{(G_n^*(ck), V_n^*(ck)), k \in \mathcal{I}_n\} (G_n^*(ci)), \quad (2.20)$$

and

$$\mathbb{Y}_n(ci) = \begin{cases} 0 \wedge LS \circ GCM \{(G_n^*(ck), V_n^*(ck)), k \in \mathcal{I}_n^l\} (G_n^*(ci)), & \text{if } i \in \mathcal{I}_n^l; \\ 0, & \text{if } i \in \mathcal{I}_n^m; \\ 0 \vee LS \circ GCM \{(G_n^*(ck), V_n^*(ck)), k \in \tilde{\mathcal{I}}_n^r\} (G_n^*(ci)), & \text{if } i \in \mathcal{I}_n^r. \end{cases} \quad (2.21)$$

Denote  $\mathcal{P}_c(k) = (\mathcal{P}_{1,c}(k), \mathcal{P}_{2,c}(k)) = (ck, \alpha W(ck) + \beta c^2 k(1+k))$ , for  $k \in \mathbb{Z}$ , as a discrete process in  $\mathbb{R}^2$ , where  $\alpha, \beta$  and  $W$  are defined as in Subsection 2.4.2. Based on  $\mathcal{P}_c$ , we define a discrete process, which will be related to  $\mathbb{X}_n$ :

$$\mathbb{X}(ci) = LS \circ GCM \{\mathcal{P}_c(k), k \in \mathbb{Z}\} (ci), \quad (2.22)$$

and a discrete process, which will be related to  $\mathbb{Y}_n$ :

$$\mathbb{Y}(ci) = \begin{cases} 0 \wedge LS \circ GCM \{\mathcal{P}_c(k), k \in -\mathbb{N}\} (ci), & \text{if } i \in -\mathbb{N}; \\ 0, & \text{if } i = 0; \\ 0 \vee LS \circ GCM \{\mathcal{P}_c(k), k \in \{0\} \cup \mathbb{N}\} (ci), & \text{if } i \in \mathbb{N}. \end{cases} \quad (2.23)$$

For simplicity of notation, for the remainder of this section, we will often write an integer interval as a usual interval with two integer endpoints. This will, however, not cause confusion since the interpretation of the interval will be immediate from the context.

For  $(\mathbb{X}_n, \mathbb{Y}_n)$ , the following joint distributional convergence holds.

*Theorem 2.4.17* (Weak Convergence of  $(\mathbb{X}_n, \mathbb{Y}_n)$ ). For each non-negative integer  $N$ , we have

$$\{(\mathbb{X}_n(ci), \mathbb{Y}_n(ci)), i \in [-N, N]\} \rightsquigarrow \{(\mathbb{X}(ci), \mathbb{Y}(ci)), i \in [-N, N]\}.$$

*Proof.* For each non-negative integer  $N$ , take an integer  $M > N$ . Then, we have the following two claims:

**Claim 1:** There exist (integer-valued) random variables  $L_n, L'_n < -M$  and  $U_n, U'_n >$

$M$  such that  $L_n, L'_n, U_n$  and  $U'_n$  are all  $O_P(1)$  and that

$$\begin{aligned}
& GCM \{(G_n^*(ck), V_n^*(ck)), k \in [L_n, U_n]\} \\
&= GCM \{(G_n^*(ck), V_n^*(ck)), k \in \mathbb{Z}\} |[G_n^*(cL_n), G_n^*(cU_n)], \\
& \quad GCM \{(G_n^*(ck), V_n^*(ck)), k \in [L'_n, -1]\} \\
&= GCM \{(G_n^*(ck), V_n^*(ck)), k \in -\mathbb{N}\} |[G_n^*(cL'_n), G_n^*(-c)], \\
& \quad GCM \{(G_n^*(ck), V_n^*(ck)), k \in [0, U'_n]\} \\
&= GCM \{(G_n^*(ck), V_n^*(ck)), k \in \{0\} \cup \mathbb{N}\} |[G_n^*(0), G_n^*(cU'_n)].
\end{aligned}$$

**Claim 2:** There also exist (integer-valued) random variables  $L, L' < -M$  and  $U, U' > M$  such that  $L, L', U, U'$  are  $O_P(1)$  and that

$$\begin{aligned}
GCM \{\mathcal{P}_c(k), k \in [L, U]\} &= GCM \{\mathcal{P}_c(k), k \in \mathbb{Z}\} |[cL, cU], \\
GCM \{\mathcal{P}_c(k), k \in [L', -1]\} &= GCM \{\mathcal{P}_c(k), k \in -\mathbb{N}\} |[cL', -c], \\
GCM \{\mathcal{P}_c(k), k \in [0, U']\} &= GCM \{\mathcal{P}_c(k), k \in \{0\} \cup \mathbb{N}\} |[0, cU'].
\end{aligned}$$

The proofs of Claim 1 and Claim 2 consist of technically important localization arguments. See Lemma 2.4.22 for the proof of Claim 1 and Lemma 2.8.7 in the appendix for that of Claim 2. Intuitively speaking, Claim 1 ensures that the restriction of the greatest convex minorant of the process  $(G_n^*, V_n^*)$  (which is involved in the construction of  $\mathbb{X}_n$  and  $\mathbb{Y}_n$ ) to a bounded domain can be made equal, eventually, with arbitrarily high probability, to the greatest convex minorant of the restriction of  $(G_n^*, V_n^*)$  to that domain, provided the domain is chosen appropriately large. A similar fact holds for the greatest convex minorant of the process  $\mathcal{P}_c$ , which is involved in the construction of  $\mathbb{X}$  and  $\mathbb{Y}$ . These equalities then translate to the left-derivatives of the GCM's involved and the proof is completed by invoking a continuous mapping theorem for the GCM's of the restriction of  $(G_n^*, V_n^*)$  on bounded domains, along with Claims 1 and 2, which enable the use of a key approximation lemma – a simple extension of Lemma 4.2 in *Prakasa Rao (1969)* – stated below.

*Lemma 2.4.18.* Suppose that for each  $\epsilon > 0$ ,  $\{W_{n\epsilon}\}$ ,  $\{W_n\}$  and  $\{W_\epsilon\}$  are sequences of random vectors,  $W$  is a random vector and such that

1.  $\lim_{\epsilon \rightarrow 0} \overline{\lim}_{n \rightarrow \infty} \mathbb{P}(W_{n\epsilon} \neq W_n) = 0$ ,
2.  $\lim_{\epsilon \rightarrow 0} \mathbb{P}(W_\epsilon \neq W) = 0$ ,

3.  $W_{n\epsilon} \rightsquigarrow W_\epsilon$ , as  $n \rightarrow \infty$  for each  $\epsilon > 0$ .

Then  $W_n \rightsquigarrow W$ , as  $n \rightarrow \infty$ .

From Claims 1 and 2, for every (small)  $\epsilon > 0$ , there exists an integer  $M_\epsilon$  large enough such that

$$P(M_\epsilon > \max\{|L_n|, |L'_n|, U_n, U'_n, |L|, |L'|, U, U'\}) > 1 - \epsilon.$$

Denote, for  $i \in [-N, N]$ ,

$$\begin{aligned} \mathbb{X}_n^{M_\epsilon}(ci) &= LS \circ GCM \{(G_n^*(ck), V_n^*(ck)), k \in [\pm M_\epsilon]\} (G_n^*(ci)), \\ \mathbb{X}^{M_\epsilon}(ci) &= LS \circ GCM \{\mathcal{P}_c(k), k \in [\pm M_\epsilon]\} (ci), \end{aligned}$$

and for  $i \in [-N, -1]$ ,  $\{0\}$  and  $[1, N]$

$$\mathbb{Y}_n^{M_\epsilon}(ci) = \begin{cases} 0 \wedge LS \circ GCM \{(G_n^*(ck), V_n^*(ck)), k \in [-M_\epsilon, -1]\} (G_n^*(ci)) \\ 0, \\ 0 \vee LS \circ GCM \{(G_n^*(ck), V_n^*(ck)), k \in [0, M_\epsilon]\} (G_n^*(ci)), \end{cases}$$

$$\mathbb{Y}(ci) = \begin{cases} 0 \wedge LS \circ GCM \{\mathcal{P}_c(k), k \in [-M_\epsilon, -1]\} (ci), & \text{if } i \in [-N, -1]; \\ 0, & \text{if } i = 0; \\ 0 \vee LS \circ GCM \{\mathcal{P}_c(k), k \in [0, M_\epsilon]\} (ci), & \text{if } i \in [1, N]. \end{cases}$$

Denote  $[\pm N] = [-N, N]$  and

$$\begin{aligned} A_n &= \{ \{ (\mathbb{X}_n^{M_\epsilon}(ci), \mathbb{Y}_n^{M_\epsilon}(ci)), i \in [\pm N] \} \neq \{ (\mathbb{X}_n(ci), \mathbb{Y}_n(ci)), i \in [\pm N] \} \}, \\ A &= \{ \{ (\mathbb{X}^{M_\epsilon}(ci), \mathbb{Y}^{M_\epsilon}(ci)), i \in [\pm N] \} \neq \{ (\mathbb{X}(ci), \mathbb{Y}(ci)), i \in [\pm N] \} \}. \end{aligned}$$

Then, the following three facts hold:

**Fact 1:**  $\lim_{\epsilon \rightarrow 0} \overline{\lim}_{n \rightarrow \infty} \mathbb{P}(A_n) = 0$ .

**Fact 2:**  $\lim_{\epsilon \rightarrow 0} \mathbb{P}(A) = 0$ .

**Fact 3:**  $\{ (\mathbb{X}_n^{M_\epsilon}(ci), \mathbb{Y}_n^{M_\epsilon}(ci)), i \in [\pm N] \} \rightsquigarrow \{ (\mathbb{X}^{M_\epsilon}(ci), \mathbb{Y}^{M_\epsilon}(ci)), i \in [\pm N] \}$ , as  $n \rightarrow \infty$  for each  $\epsilon > 0$ .

Since  $A_n$  and  $A$  are subsets of  $\{M_\epsilon \leq \max\{|L_n|, U_n, |L|, U, |L'_n|, U'_n, |L'|, U'\}\}$ , whose probability is less than  $\epsilon$ , Facts 1 and 2 hold. Fact 3 will be proved in Lemma 2.4.23 in Section 2.4.3.1. Therefore, by Lemma 2.4.18, we have the finite dimensional weak convergence.  $\square$

*Remark 2.4.19.* From Theorem 2.4.17, we have  $n^{1/3}(\hat{F}(t_l) - F(t_l)) = \mathbb{X}_n(0)$  converges weekly to  $\mathbb{X}(0) = LS \circ GCM\{\mathcal{P}_c(k), k \in \mathbb{Z}\}(0)$ . Note that  $\mathbb{X}(0)$  is precisely the random variable  $\mathcal{S}_c$  defined in Section 2.

*Theorem 2.4.20 (Weak Convergence of  $2 \log \lambda_n$ ).* Under the null hypothesis of (2.3), i.e.  $F(t_l) = \theta_l$ , we have

$$2 \log \lambda_n \rightsquigarrow \mathbb{M}_c = \frac{cg(x_0)}{F(x_0)(1 - F(x_0))} \sum_{i=-\infty}^{\infty} (\mathbb{X}(ci)^2 - \mathbb{Y}(ci)^2).$$

*Proof.* Denote  $J_n = \{1 \leq i \leq n : \hat{F}(X_i) \neq \hat{F}^o(X_i)\}$ . Then, we have

$$2 \log \lambda_n = 2(I - II) \text{ with } I = \sum_{i \in J_n} \psi(X_i, Y_i; \hat{F}), \quad II = \sum_{i \in J_n} \psi(X_i, Y_i; \hat{F}^o)$$

and  $\psi(x, y; F) = y \log F(x) + (1 - y) \log(1 - F(x))$ . By Taylor expansions of  $\log \hat{F}(X_i)$  and  $\log(1 - \hat{F}(X_i))$  around the point  $t_l \equiv t_{l,n}$  (with  $F_l = F(t_l)$ ) we have  $I = I(1) + I(2) + I(3) + I(4)$ , where

$$\begin{aligned} I(1) &= \log F_l \sum_{i \in J_n} Y_i + \log(1 - F_l) \sum_{i \in J_n} (1 - Y_i), \\ I(2) &= \sum_{i \in J_n} \left( \frac{Y_i}{F_l} - \frac{1 - Y_i}{1 - F_l} \right) (\hat{F}(X_i) - F_l), \\ I(3) &= -\frac{1}{2} \sum_{i \in J_n} \left( \frac{Y_i}{F_l^2} + \frac{1 - Y_i}{(1 - F_l)^2} \right) (\hat{F}(X_i) - F_l)^2, \\ I(4) &= \frac{1}{3} \sum_{i \in J_n} \left( \frac{Y_i}{F_{li}^{*3}} - \frac{1 - Y_i}{(1 - F_{li}^{**})^3} \right) (\hat{F}(X_i) - F_l)^3, \end{aligned}$$

and  $F_{li}^*$  and  $F_{li}^{**}$  both lie between  $F_l$  and  $\hat{F}(X_i)$ . By noting that  $\{X_i : i \in J_n\}$  is the union of several whole “blocks” of  $X_{(i)}$ ’s in the unconstrained isotonic regression and that over each block  $\hat{F}$  equals the average of the corresponding block responses, we



have, for each nonnegative integer  $k$ ,

$$\sum_{i \in J_n} (Y_i - \hat{F}(X_i))(\hat{F}(X_i) - F_l)^k = 0. \quad (2.24)$$

Using this fact, we have

$$\begin{aligned} I(2) &= \frac{1}{F_l(1-F_l)} \sum_{i \in J_n} \left( \hat{F}(X_i) - F_l \right)^2, \\ I(3) &= -\frac{1}{2}I(2) - \frac{1}{2} \left( \frac{1}{F_l^2} - \frac{1}{(1-F_l)^2} \right) \sum_{i \in J_n} \left( \hat{F}(X_i) - F_l \right)^3. \end{aligned}$$

Thus,  $I$  simplifies to

$$I = I(1) + \frac{1}{2F_l(1-F_l)} \sum_{i \in J_n} \left( \hat{F}(X_i) - F_l \right)^2 + R_1,$$

where  $R_1$  is the remaining term.

Similarly, we have

$$II = II(1) + \frac{1}{2F_l(1-F_l)} \sum_{i \in J_n} \left( \hat{F}^o(X_i) - F_l \right)^2 + R_2,$$

where  $II(1) = I(1)$  and  $R_2$  is the remaining term, by noting that the corresponding equation of (2.24) still holds since  $\{X_i : i \in J_n\}$  is the union of whole “blocks” of those two constrained isotonic regressions and a special interval containing  $t_l$  over which  $\hat{F}^o$  always equals to  $F_l$ . In fact, the special interval consists of  $t_l$  and the “truncated” blocks of those two constrained isotonic regressions.

We will show both  $R_1$  and  $R_2$  are  $o_P(1)$  in Lemma 2.8.14. Thus, we have:

$$2 \log \lambda_n = \frac{1}{F_l(1-F_l)} \sum_{i \in J_n} \left[ \left( \hat{F}(X_i) - F_l \right)^2 - \left( \hat{F}^o(X_i) - F_l \right)^2 \right] + o_P(1).$$

Denote  $D_n = \{1 \leq j \leq K : \hat{F}(t_j) \neq \hat{F}^o(t_j)\}$ . Then  $D_n$  is either an empty set or an integer interval including  $l$ . Denote  $\tilde{D}_n = D_n - l$  also by  $[-L_n, U_n]$ . For empty  $D_n$ ,  $[-L_n, U_n]$  is understood as  $\emptyset$ . A summation with an empty index set is defined to be 0.

Next, for  $1 \leq j \leq K$ , let  $W'_j = \#\{1 \leq i \leq n : X_i = t_j\}$ , the number of observation

times equal to  $t_j$ . Then,

$$\begin{aligned}
& \sum_{i \in J_n} \left[ \left( \hat{F}(X_i) - F_l \right)^2 - \left( \hat{F}^o(X_i) - F_l \right)^2 \right] \\
&= \sum_{j \in D_n} \left[ \left( \hat{F}(t_j) - F_l \right)^2 - \left( \hat{F}^o(t_j) - F_l \right)^2 \right] W'_j \\
&= \sum_{j \in \tilde{D}_n} \left[ \left( \hat{F}(t_{l+j}) - F_l \right)^2 - \left( \hat{F}^o(t_{l+j}) - F_l \right)^2 \right] W'_{l+j} \\
&= \sum_{j \in [-L_n, U_n]} \left( \mathbb{X}_n^2(cj) - \mathbb{Y}_n^2(cj) \right) n^{-2/3} W'_{l+j},
\end{aligned}$$

using the definitions of the process  $\mathbb{X}_n$  and  $\mathbb{Y}_n$ . On the other hand,

$$W'_{l+j} = \sum_{i=1}^n \{t_{l+j-1} < X_i \leq t_{l+j}\} = g(x_0) n^{2/3} (G_n^*(cj) - G_n^*(c(j-1))),$$

which means  $n^{-2/3} W'_{l+j} = g(x_0) W_j$  with  $W_j = G_n^*(cj) - G_n^*(c(j-1))$ . Then, we have:

$$\begin{aligned}
2 \log \lambda_n &= \frac{g(x_0)}{F_l(1-F_l)} \sum_{j \in [-L_n, U_n]} \left( \mathbb{X}_n^2(cj) - \mathbb{Y}_n^2(cj) \right) W_j + o_P(1) \\
&= \frac{cg(x_0)}{F_l(1-F_l)} S_n + R_3 + o_P(1),
\end{aligned}$$

where  $S_n = \sum_{j \in [-L_n, U_n]} \left( \mathbb{X}_n^2(cj) - \mathbb{Y}_n^2(cj) \right)$  and  $R_3$  is the remaining term.

Similarly, denote  $E = \{j \in \mathbb{Z} : \mathbb{X}(cj) \neq \mathbb{Y}(cj)\}$  for the limiting processes  $\mathbb{X}$  and  $\mathbb{Y}$ . Then,  $E$  is either an empty set or an integer interval usually including 0. Denote  $E$  also by  $[-L, U]$ . For empty  $E$ ,  $[-L, U]$  is understood as  $\emptyset$  as before. Let  $S = \sum_{j \in [-L, U]} \left( \mathbb{X}^2(cj) - \mathbb{Y}^2(cj) \right)$ .

We will show  $S_n \rightsquigarrow S$  in Lemma 2.8.11. On the other hand, it will be shown that  $R_3 = o_P(1)$  in Lemma 2.8.14. Therefore, we complete the proof by noticing  $F_l \rightarrow F(x_0)$  and applying Slutsky's Lemma.  $\square$

*Remark 2.4.21.* The limit distribution  $\mathbb{M}_c$  in Theorem 2.4.20 is, up to a constant, the difference of two discrete stochastic processes summed over the (discrete) time axis and depends on parameters. Owing to the discreteness of the time axis, there is no scaling to filter out the nuisance parameters in  $\mathbb{M}_c$  so that we do not have a universal limit distribution for the likelihood ratio statistic in this boundary case. As

a comparison, in the case with  $\gamma > 1/3$  or with a continuous design density, Brownian scaling ensures that all the nuisance parameters in the expression for the likelihood ratio statistic disappear, leaving us with the universal limit  $\mathbb{D}$ . For more details on the nature of the Brownian scaling arguments, see, for example, Pages 1724-1725 of Banerjee and Wellner (2001).

**Inference for  $F(x_0)$  :** The linearly interpolated estimator that we used to make inference for  $F(x_0)$  when  $\gamma < 1/3$  cannot be made to work in this situation. Inference on  $F(x_0)$  with the linear interpolant can, however, be made if we consider a slightly altered setting for the grid of observation times. This and related issues are discussed towards the end of the next section on adaptive inference for  $F$  at a point.

### 2.4.3.1 Technical Details

*Lemma 2.4.22 (Localization Argument).* Claim 1 in the proof of Theorem 2.4.17 holds.

*Proof.* We only prove the first equality. The others can be established through analogous arguments. Denote  $\mathbb{K}_n(x) = GCM \{(G_n(t), V_n(t)), t \in [a, b]\} (x)$  for  $x \in [0, 1]$ . Recall  $t_{l+j} = t_l + jcn^{-1/3}$  for  $j \in \mathbb{Z}$ . Then, it is sufficient to show there exist  $x_{L_n} < t_{l-M}$  and  $x_{U_n} > t_{l+M}$  such that

- (1) both  $t_l - x_{L_n}$  and  $x_{U_n} - t_l$  are  $O_P(n^{-1/3})$ , and
- (2)  $GCM \{(G_n(t), V_n(t)), t \in [x_{L_n}, x_{U_n}]\} = \mathbb{K}_n|[G_n(x_{L_n}), G_n(x_{U_n})]$ .

Let  $x_{L_n}$  and  $x_{U_n}$  be the largest grid point less than  $t_{l-M}$  and the smallest grid point larger than or equal to  $t_{l+M}$ , at which  $\hat{F}$  jumps. Thus, the slopes of  $\mathbb{K}_n$  change at (and  $\mathbb{K}_n$  and  $CSD$  agree on)  $G_n(x_{L_n})$  and  $G_n(x_{U_n})$ .

With the above choices of  $x_{L_n}$  and  $x_{U_n}$ , (2) is satisfied. We next show that (1) holds. It is sufficient to show both  $t_{l-M} - x_{L_n}$  and  $x_{U_n} - t_{l+M}$  are  $O_P(n^{-1/3})$ . We next show the former and the latter can be established in the same way. Denote  $T_n = t_{l-M} - x_{L_n}$  and let  $S_n$  be the smallest non-negative value such that  $\hat{F}$  changes its value at  $t_{l-M} + S_n$ . First, by the definition of  $\hat{F}$ , for all  $\beta$ , we have

$$V_n(t_{l-M} + \beta) \geq \mathbb{K}_n(t_{l-M}) + \hat{F}(t_{l-M}) [G_n(t_{l-M} + \beta) - G_n(t_{l-M})].$$

Denote, for all  $\beta$ ,

$$\Gamma_n(\beta) = [V_n(t_{l-M} + \beta) - V_n(t_{l-M})] - \hat{F}(t_{l-M}) [G_n(t_{l-M} + \beta) - G_n(t_{l-M})].$$

Then, it is easy to see that  $\Gamma_n(0) = 0$  and  $\Gamma_n$  achieves its minimum at both  $-T_n$  and  $S_n$ . Thus,  $\Gamma_n(-T_n) \leq 0$ .

Letting  $\mathbb{P}$  denote the distribution of  $(Y_{1n}, X_{1n})$  (note the suppression of dependence on  $n$ ) and  $\mathbb{P}_n$  the empirical measure of  $n$  i.i.d. observations from this distribution, we have via simple algebra,

$$\begin{aligned} \Gamma_n(\beta) = & (\mathbb{P}_n - \mathbb{P})g_{n1}(x, y; \beta) - \gamma_n(\mathbb{P}_n - \mathbb{P})g_{n2}(x; \beta) \\ & + \mathbb{P}g_{n1}(x, y; \beta) - \gamma_n\mathbb{P}g_{n2}(x; \beta), \end{aligned}$$

where  $\gamma_n = \hat{F}(t_{l-M}) - F(t_{l-M})$ ,  $g_{n1}(x, y; \beta) = (y - F(t_{l-M}))g_{n2}(x; \beta)$  and  $g_{n2}(x; \beta) = \{x \leq t_{l-M} + \beta\} - \{x \leq t_{l-M}\}$ .

Since  $T_n$ ,  $S_n$  and  $\gamma_n$  are  $o_P(1)$  by a standard consistency argument, we can analyze locally. Now, for all sufficiently small  $\epsilon > 0$ , we can find, a neighborhood  $\mathcal{N}$  of 0 (that may depend on  $\epsilon$ ), such that for all sufficiently large  $n$  (depending on  $\epsilon$ ), the following facts hold for every  $\beta \in \mathcal{N}$ :

**Fact 1:**  $|\mathbb{P}g_{n1}(x, y; \beta) - (1/2)g(x_0)f(x_0)\beta^2| \leq \epsilon\beta^2 + O(n^{-2/3})$ .

**Fact 2:**  $|\mathbb{P}g_{n2}(x; \beta) - g(x_0)\beta| \leq K\beta^2 + O(n^{-1/3})$ .

**Fact 3:**  $|(\mathbb{P}_n - \mathbb{P})g_{n1}(x, y; \beta)| \leq \epsilon\beta^2 + O_P(n^{-2/3})$ .

**Fact 4:**  $|(\mathbb{P}_n - \mathbb{P})g_{n2}(x; \beta)| \leq \epsilon\beta^2 + O_P(n^{-2/3})$ .

The  $O$  and  $O_P$  terms in the above facts are non-negative and can depend on  $\epsilon$ , but not on  $\beta \in \mathcal{N}$ . The constant  $K$  can again depend on  $\epsilon$ . For the purpose of continuity, the proofs of the above facts are provided in Lemmas [2.8.3](#), [2.8.4](#), [2.8.5](#), and [2.8.6](#). Next, choose and fix  $\epsilon_0$  small enough such that the above facts hold and furthermore  $\epsilon_0 < f(x_0)g(x_0)/8$ . The above four facts imply that:

$$\begin{aligned} & |\Gamma_n(\beta) - (1/2)f(x_0)g(x_0)\beta^2 + \gamma_n g(x_0)\beta| \\ & \leq 3\epsilon_0\beta^2 + \gamma_n K\beta^2 + O_P(n^{-2/3}) + |\gamma_n|O(n^{-1/3}). \end{aligned}$$

Given  $\eta > 0$ , there exists  $N_\eta$  such that for all  $n \geq N_\eta$ , we have  $\gamma_n K < \epsilon_0$  and  $-T_n$  and  $S_n$  lie in  $\mathcal{N}$  with probability at least  $1 - \eta$ . Denote this event as  $E_n$ . Then, on  $E_n$ ,  $A_n(\beta) \leq \Gamma_n(\beta) \leq B_n(\beta)$ , where

$$\begin{aligned} A_n(\beta) &= C_1\beta^2 - \gamma'_n\beta - O_P(n^{-2/3}) - |\gamma'_n|O(n^{-1/3}), \\ B_n(\beta) &= C_2\beta^2 - \gamma'_n\beta + O_P(n^{-2/3}) + |\gamma'_n|O(n^{-1/3}), \end{aligned}$$

and  $C_1 = (f(x_0)g(x_0)/2 - 4\epsilon_0) > 0$ ,  $C_2 = (f(x_0)g(x_0)/2 + 4\epsilon_0) > 0$  are constants and  $\gamma'_n = \gamma_n g(x_0)$ . Note that  $\gamma'_n$  and  $\gamma_n$  have the same stochastic order. From the definition of  $T_n$  and  $S_n$ ,

$$\min_{\beta} \Gamma_n(\beta) = \max\{\Gamma_n(-T_n), \Gamma_n(S_n)\} \geq \max\{A_n(-T_n), A_n(S_n)\}.$$

Since one of  $C_1(-T_n)^2 - \gamma'_n(-T_n)$  and  $C_1S_n^2 - \gamma'_nS_n$  must be non-negative, we readily conclude that  $\min_{\beta} \Gamma_n(\beta) \geq -O_P(n^{-2/3}) - |\gamma'_n|O(n^{-1/3})$ . Next, note that  $B_n(\cdot)$  attains its minimum at  $\gamma'_n/2C_2$ . Hence, we also have

$$\begin{aligned} \min_{\beta} \Gamma_n(\beta) &\leq \Gamma_n(\gamma'_n/(2C_2)) \leq B_n(\gamma'_n/(2C_2)) \\ &\leq -(\gamma'_n)^2/(4C_2) + O_P(n^{-2/3}) + |\gamma'_n|O(n^{-1/3}). \end{aligned}$$

We conclude that

$$-O_P(n^{-2/3}) - |\gamma'_n|O(n^{-1/3}) \leq -(\gamma'_n)^2/(4C_2) + O_P(n^{-2/3}) + |\gamma'_n|O(n^{-1/3}),$$

which can be written equivalently as

$$(\gamma'_n)^2 - |\gamma'_n|O(n^{-1/3}) - O_P(n^{-2/3}) \leq 0.$$

This implies  $|\gamma'_n| \leq O_P(n^{-1/3}) =: \xi_n$  and thus  $|\gamma'_n|$  is  $O_P(n^{-1/3})$ . Further,

$$0 = \Gamma_n(0) \geq \Gamma_n(-T_n) \geq A_n(-T_n) = C_1T_n^2 + \gamma'_nT_n - O_P(n^{-2/3}) - |\gamma'_n|O(n^{-1/3}),$$

showing that  $T_n$  must be less in absolute value than the maximum of the roots of the

quadratic equation  $C_1x^2 + \gamma'_n x - O_P(n^{-2/3}) - |\gamma'_n|O(n^{-1/3}) = 0$ . This leads to

$$|T_n| \leq \frac{\xi_n + \sqrt{\xi_n^2 + 4C_1O_P(n^{-2/3}) + 4C_1\xi_nO(n^{-1/3})}}{2C_1} =: \tilde{\xi}_n,$$

and  $\tilde{\xi}_n$  is again  $O_P(n^{-1/3})$ . Thus, for  $\tau > 0$ ,

$$P(n^{1/3}T_n > \tau) \leq P(E_n^c) + P(\{n^{1/3}T_n > \tau\} \cap E_n) \leq P(E_n^c) + P(n^{1/3}\tilde{\xi}_n > \tau),$$

which is less than  $2\eta$  for all  $n$  sufficiently large given  $\tau$  large enough. Therefore,  $T_n$  is  $O_P(n^{-1/3})$  and the proof is complete.  $\square$

*Lemma 2.4.23.* Fact 3 in the proof of Theorem 2.4.17 holds.

*Proof.* We view  $\{(\mathbb{X}_n^{M_\epsilon}(ci), \mathbb{Y}_n^{M_\epsilon}(ci)), i \in [\pm N]\}$  as the image of a  $2(2N + 1)$  dimensional vector function  $\Psi$  defined on the vector of two processes  $\mathcal{W}_n = (\mathcal{G}_{nf}, \mathcal{V}_{nf})$ , where  $\mathcal{G}_{nf} = \{G_n^*(ck), k \in [\pm M_\epsilon]\}$  and  $\mathcal{V}_{nf} = \{V_n^*(ck), k \in [\pm M_\epsilon]\}$ . More specifically, denote  $\Psi = (\{\Psi_i\}_{i=-N}^N, \{\Psi_i^l\}_{i=-N}^{-1}, 0, \{\Psi_i^r\}_{i=1}^N)$ , where

$$\Psi_i(\kappa, \lambda) = \max_{-M_\epsilon \leq u < i; i \leq v \leq M_\epsilon} \min \{(\lambda(v) - \lambda(u))/(\kappa(v) - \kappa(u))\} \text{ for } i \in [\pm N],$$

$$\Psi_i^l(\kappa_l, \lambda_l) = \max_{-M_\epsilon \leq u < i; i \leq v < 0} \min \{(\lambda(v) - \lambda(u))/(\kappa(v) - \kappa(u))\} \wedge 0 \text{ for } i \in [-N, 1],$$

$$\Psi_i^r(\kappa_r, \lambda_r) = \max_{0 \leq u < i; i \leq v \leq M_\epsilon} \min \{(\lambda(v) - \lambda(u))/(\kappa(v) - \kappa(u))\} \vee 0 \text{ for } i \in [1, N];$$

$\kappa$  and  $\lambda$  are functions defined on  $[\pm M_\epsilon]$ . Then,  $\mathbb{X}_n^{M_\epsilon}(ci) = \Psi_i((G_n^*(kc), V_n^*(kc)) : k \in [\pm M_\epsilon])$  for  $i \in [\pm N]$ ,  $\mathbb{Y}_n^{M_\epsilon}(ci) = \Psi_i^l((G_n^*(kc), V_n^*(kc)) : k \in [\pm M_\epsilon])$  for  $i \in [-N, -1]$  and  $\mathbb{Y}_n^{M_\epsilon}(ci) = \Psi_i^r((G_n^*(kc), V_n^*(kc)) : k \in [\pm M_\epsilon])$  for  $i \in [1, N]$ . Similarly,  $\{(\mathbb{X}_n^{M_\epsilon}(ci), \mathbb{Y}_n^{M_\epsilon}(ci)), i \in [\pm N]\}$  is the image of the same  $2(2N + 1)$  dimensional vector function  $\Psi$  defined on the vector of two processes  $\mathcal{W} = (\mathcal{G}_f, \mathcal{V}_f)$ , where  $\mathcal{G}_f = \{\mathcal{P}_{1,c}(k), k \in [\pm M_\epsilon]\}$  and  $\mathcal{V}_f = \{\mathcal{P}_{2,c}(k), k \in [\pm M_\epsilon]\}$ . Letting  $\mathcal{H}$  be the set of real-valued functions defined on the integer interval  $[\pm M_\epsilon]$ , view  $(\kappa, \lambda)$  as an element of  $\mathcal{H}^2$  equipped with the topology of pointwise convergence on each co-ordinate. Note that  $\mathcal{W}_n$  and  $\mathcal{W}$  are random elements assuming values in  $\mathcal{H}^2$ . Thus  $(\kappa_m, \lambda_m)$  is defined to converge to  $(\kappa_0, \lambda_0)$  if  $\kappa_m(j) \rightarrow \kappa_0(j)$  for all  $j \in [\pm M_\epsilon]$  and the same holds for  $\lambda_m$  and  $\lambda_0$ . Then,  $\Psi$  can be viewed as a map from  $\mathcal{H}^2$  to  $\mathbb{R}^{2(2N+1)}$ . The continuity of  $\Psi$  follows immediately from the fact that each  $\Psi_i, \Psi_i^l, \Psi_i^r$  is continuous: this follows from the fact that the max and min in the definitions of each of these functions are

taken with respect to a finite number of elements. We show in Lemmas 2.8.8 and 2.8.9, that

$$\mathcal{G}_{nf} \xrightarrow{P^*} \mathcal{G}_f \tag{2.25}$$

$$\mathcal{V}_{nf} \rightsquigarrow \mathcal{V}_f \tag{2.26}$$

Then, by Slutsky’s lemma, we have  $\mathcal{W}_n \rightsquigarrow \mathcal{W}$ . The conclusion of the lemma is now a direct outcome of the continuous mapping theorem.  $\square$

## 2.5 Adaptive Inference for $F$ at A Point

The goal in this section is to develop a scheme for constructing confidence intervals for  $F(t_l)$  (and later also for  $F(x_0)$ ) without knowledge of the underlying grid resolution controlled by the parameters  $\gamma$  and  $c$ , given current status data over an equally spaced grid of observation times on  $[a, b]$ . To this end, we first investigate the relationships between the three different asymptotic limits for the distribution of  $\hat{F}(t_l)$  that were derived in the previous section, under different values of  $\gamma$ . Our first result relates the distribution of  $\mathcal{S}_c$  to the standard normal.

*Theorem 2.5.1.* As  $c \rightarrow \infty$ , we have  $\sqrt{c}\mathcal{S}_c \xrightarrow{d} \alpha Z$ , where  $Z$  follows the standard normal distribution.

*Remark 2.5.2.* Recall that  $\mathcal{S}_c$  does depend on both  $\alpha$  and  $\beta$ . So, rigorously, it should be written as  $\mathcal{S}_{c,\alpha,\beta}$ . The notation  $\mathcal{S}_c$ , though convenient, is perhaps a bit unfortunate since dependence on other parameters in the problem are ignored. We need to keep this dependence in mind.

Our next result investigates the case where  $c$  goes to 0.

*Theorem 2.5.3.* As  $c \rightarrow 0$ , we have  $\mathcal{S}_c \xrightarrow{d} g_{\alpha,\beta}(0) \stackrel{d}{=} 2(\alpha^2\beta)^{1/3}\mathcal{Z}$ .

*Remark 2.5.4.* This result is somewhat easier to visualize at a heuristic level. Recall that  $\mathcal{S}_c$  is the left-slope of the GCM of the process  $\mathcal{P}_c$  at the point 0, the process itself being defined on the grid  $c\mathbb{Z}$ . As  $c$  goes to 0, the grid becomes finer and the process  $\mathcal{P}_c$  will eventually be substituted by its limiting version, namely  $X_{\alpha,\beta}$ . Thus, in the limit  $\mathcal{S}_c$  becomes  $g_{\alpha,\beta}(0)$ , the left-slope of the GCM of  $X_{\alpha,\beta}$  at 0. The representation of this limit in terms of  $\mathcal{Z}$  was established in Remark 2.4.13 following Theorem 2.4.12.

We are now in a position to propose our inference scheme. We focus on the so-called ‘Wald-type’ intervals for  $F(t_l)$ , i.e. intervals of the form  $\hat{F}(t_l)$  plus and

minus terms depending on the sample size and the large sample distribution of the estimator. Recall that the grid resolution is determined by unknown parameters  $\gamma > 0$  and  $c > 0$ . We now *pretend* that  $\gamma$  is exactly equal to  $n^{-1/3}$ . This allows us to calculate the value of the corresponding parameter  $c$ , say  $\hat{c}$ , via the relation:  $\lfloor (b-a)/(\hat{c}n^{-1/3}) \rfloor = K (= \lfloor (b-a)/(cn^{-\gamma}) \rfloor)$ , where  $K$  is the number of grid points. Some algebra shows that

$$\hat{c} = \hat{c}_n = cn^{1/3-\gamma} + O(n^{1/3-2\gamma}) = cn^{1/3-\gamma}(1 + O(n^{-\gamma})).$$

Thus, the calculated parameter  $\hat{c}$  actually depends on  $n$ , and goes to  $\infty$  and 0 for  $\gamma \in (0, 1/3)$  and  $\gamma \in (1/3, 1)$ , respectively. We propose to use the distribution of  $\mathcal{S}_{\hat{c}}$  as an approximation to the distribution of  $n^{1/3}(\hat{F}(t_l) - F(t_l))$  for  $\gamma \in (0, 1)$ . Thus, an adaptive approximate  $1 - \eta$  confidence interval for  $F(t_l)$  is given by

$$\left[ \hat{F}(t_l) - n^{-1/3} q(\mathcal{S}_{\hat{c}}, 1 - \eta/2), \hat{F}(t_l) + n^{-1/3} q(\mathcal{S}_{\hat{c}}, (\eta/2)) \right], \quad (2.27)$$

where  $\eta > 0$  and  $q(X, p)$  stands for the lower  $p$ th quantile of a random variable  $X$  with  $p \in (0, 1)$ .

The above adaptive confidence interval provides the correct asymptotic calibration, irrespective of the true value of  $\gamma$ . If  $\gamma$  happens to be  $1/3$ , then the adaptive confidence interval is constructed with the correct asymptotic result. Usually this situation is rare so that it is important to know the performance of the adaptive confidence interval for  $\gamma \neq 1/3$ . If *we knew* that  $\gamma \in (1/3, 1)$ , then, by the result (2.19) and the symmetry of  $g_{\alpha, \beta}(0)$ , the true confidence interval would be

$$\left[ \hat{F}(t_l) \pm n^{-1/3} q(g_{\alpha, \beta}(0), (1 - \eta/2)) \right]. \quad (2.28)$$

Recall that  $\hat{c}$  goes to 0 for  $\gamma \in (1/3, 1)$ . Thus, by Theorem 2.5.3, the quantile sequence  $q(\mathcal{S}_{\hat{c}}, p)$  converges to  $q(g_{\alpha, \beta}(0), p)$  since  $g_{\alpha, \beta}(0)$  is a continuous random variable. So, the adaptive confidence interval (2.27) converges to the true one (2.28) obtained when  $\gamma$  is in  $(1/3, 1)$ .

That the adaptive procedure also works with  $\gamma \in (0, 1/3)$  will be shown by using Theorem 2.5.1. When the true  $\gamma$  is known to be less than  $1/3$ , from the result (2.12) and the symmetry of the standard normal random variable  $Z$ , the confidence interval



is given by

$$\left[ \hat{F}(t_l) \pm n^{-(1-\gamma)/2} \alpha c^{-1/2} q(Z, (1 - \eta/2)) \right]. \quad (2.29)$$

We then only need to show that, for every  $p \in (0, 1)$ , as  $n \rightarrow \infty$ ,

$$\frac{n^{-1/3} q(\mathcal{S}_{\hat{c}}, p)}{n^{-(1-\gamma)/2} \alpha c^{-1/2} q(Z, p)} = \frac{n^{-1/3} c^{1/2}}{n^{-(1-\gamma)/2} \hat{c}^{1/2}} \cdot \frac{\hat{c}^{1/2} q(\mathcal{S}_{\hat{c}}, p)}{\alpha q(Z, p)} = I \cdot II \rightarrow 1.$$

Recall that  $\hat{c}$  goes to  $\infty$  for  $\gamma \in (0, 1/3)$ . By Theorem 2.5.1, we have  $II \rightarrow 1$  as  $n \rightarrow \infty$ . On the other hand, we can see  $I$  simplifies to  $(1 + O(n^{-\gamma}))^{-1/2}$  and therefore goes to 1. Thus, the adaptive confidence interval (2.27) also converges to the true one (2.29) obtained when  $\gamma$  is known to be in  $(0, 1/3)$ .

While the major advantage of our method lies in that it does not require making a call on the degree of sparsity  $\gamma$  — in that respect, it *adjusts automatically* to the inherent rate of growth of the number of distinct observation times to the number of individuals, and that is an extremely desirable property — there are some implementational issues with the method that should be pointed out.

Firstly, note that nuisance parameters, namely  $g(x_0)$  and  $f(x_0)$ , do need to be estimated from the data. Of the two  $f(x_0)$  is more difficult to estimate accurately. Estimation of nuisance parameters is however unavoidable using the Wald-type intervals we have been dealing with, *even with  $\gamma$  known*. This nuisance parameter problem is somewhat easier for  $\gamma \in (0, 1/3)$ , since one at most needs an estimate of  $g(x_0)$  in that situation and this is readily available via standard smoothing techniques. For  $\gamma \in (1/3, 1)$ , both nuisance parameters enter into the limit distribution of  $n^{1/3}(\hat{F}(t_l) - F(t_l))$ . Of course, with  $\gamma$  *known but not equal to 1/3*, one could have dispensed with the Wald-type intervals altogether, considering instead intervals using LRS inversion. This would have obviated the need to estimate nuisance parameters, since in either case,  $\gamma \in (0, 1/3)$  or  $\gamma \in (1/3, 1)$ , the LRS is asymptotically pivotal with known limit distributions  $\chi_1^2$  and  $\mathbb{D}$  respectively. Recall that in the boundary scenario, i.e. for  $\gamma = 1/3$ , the asymptotic distribution of the LRS is no longer pivotal, so the usual advantageous feature of likelihood ratios is absent in this situation.

Secondly, note that unlike the cases  $\gamma \in (0, 1/3)$  and  $\gamma \in (1/3, 1)$ , the limit distribution of  $n^{1/3}(\hat{F}(t_l) - F(t_l))$  in the boundary case does not admit a natural scaling in terms of a fixed known distribution; with  $\gamma \in (0, 1/3)$ , the standard normal plays this role and for  $\gamma \in (1/3, 1)$ , Chernoff's distribution. This means that the quantiles

of the limit distribution in the boundary case cannot be computed by multiplying the quantiles of a generic distribution by an estimated factor and must be generated time and again starting from the parent process  $\mathcal{P}_c$  which depends on  $\alpha$  and  $\beta$ , or to be more precise, from an estimated parent process generated by using consistent estimates of the unknown parameters  $\alpha$  and  $\beta$ . This is, however, not a terribly major issue in these days of fast computing and in our opinion, the mileage obtained in terms of adaptivity more than compensates for the lack of scaling.

Finally, one might wonder if resampling techniques could be used for adaptive estimation. The problem, however, lies in the fact that while the usual  $n$  out of  $n$  bootstrap works for  $\gamma \in (0, 1/3)$ , it fails under the non-standard asymptotic regimes that operate for  $\gamma \in [1/3, 1)$ , as is clear from the work of [Abrevaya and Huang \(2005\)](#), [Kosorok \(2008\)](#) and [Sen et al. \(2010\)](#). Since  $\gamma$  is unknown, it is impossible to decide whether to use the standard bootstrap or not. One could argue that the  $m$  out of  $n$  bootstrap or subsampling will work irrespective of the value of  $\gamma$  but, again, the problem with using these methods is that they require knowledge of the convergence rate and this is unknown as it depends on the true value of  $\gamma$ .

Similar to the relationship among  $\mathcal{S}_c$ ,  $Z$  and  $\mathcal{Z}$ , revealed by [Theorem 2.5.1](#) and [Theorem 2.5.3](#), we can establish a corresponding relationship among  $\mathbb{M}_c$ ,  $\chi_1^2$  and  $\mathbb{D}$ .

*Theorem 2.5.5.* As  $c \rightarrow \infty$ , we have  $\mathbb{M}_c \xrightarrow{d} \chi_1^2$ .

*Theorem 2.5.6.* As  $c \rightarrow 0$ , we have  $\mathbb{M}_c \xrightarrow{d} \mathbb{D}$ .

From these two theorems, we can similarly develop another adaptive procedure based on the LRS of the boundary case for testing the hypotheses [\(2.3\)](#) to construct the so-called LR-type confidence intervals for  $F(t_l)$  without the knowledge of the true value of  $\gamma \in (0, 1)$ . These adaptive LR-type confidence intervals are also asymptotically correct, regardless the value of  $\gamma$ . As pointed out before, the asymptotic distribution of the LRS for the boundary case  $\gamma = 1/3$  is no longer pivotal and the same nuisance parameters  $g(x_0)$  and  $f(x_0)$  need to be estimated in order to construct the adaptive LR-type confidence intervals. This is a reasonable price for not knowing the true value of  $\gamma$ .

**Inference for  $F(x_0)$  :** To make adaptive inference on  $F(x_0)$  given  $x_0 \in (a, b)$ , consider a slightly altered setting for the grid. Suppose  $p \in [0, 1)$  and  $q = 1 - p$ . Let  $t_l = x_0 - p\delta$ ,  $t_r = x_0 + q\delta$ ,  $t_{l-i} = t_l - i\delta$ ,  $t_{r+j} = t_r + j\delta$  for  $i = 1, \dots, l - 1$  and

$j = 1, \dots, K - r$  with both  $t_1 - a$  and  $b - t_K$  in  $(0, \delta]$ . Since the unconstrained and constrained isotonic regressions  $\hat{F}$  and  $\hat{F}^o$  are well-defined only at the grid-points, the related random quantities  $\mathbb{X}_n$  and  $\mathbb{Y}_n$  are also well-defined without any modification.

Under the new setting of the grid, first consider the boundary case  $\gamma = 1/3$ , for which Theorem 2.4.17 still holds. Define  $\tilde{F}(x_0) = q\hat{F}(t_l) + p\hat{F}(t_r)$ , which is the linearly interpolated estimator for  $F(x_0)$ . Then, we have  $n^{1/3}(\tilde{F}(x_0) - F(x_0)) = q\mathbb{X}_n(0) + p\mathbb{X}_n(1) + n^{1/3}(F(t_l) - F(x_0))$ , which converges weakly to  $q\mathbb{X}(0) + p\mathbb{X}(1) - cpf(x_0)$  given  $f$  is continuous around  $x_0$ . Then, we can construct confidence intervals for  $F(x_0)$ . However, with our original setting for the grid,  $n^{1/3}(\tilde{F}(x_0) - F(x_0))$  might not converge weakly, because the  $p$  and  $q$  would depend on  $n$  without necessarily converging to a limit. To see this, let  $a = 0$ ,  $b = 1$ ,  $x_0 = 1/2$ ,  $c = 1$ ,  $n_{1,k} = (2k)^3$  and  $n_{2,k} = (2k + 1)^3$  for  $k \in \mathbb{N}$ . Then, along the first subsequence  $\{n_{1,k}\}$ ,  $n_{1,k}^{1/3}(\tilde{F}(x_0) - F(x_0))$  converges weakly to  $\mathbb{X}_n(0)$ . But along the second subsequence  $\{n_{2,k}\}$ ,  $n_{2,k}^{1/3}(\tilde{F}(x_0) - F(x_0))$  converges weakly to  $(\mathbb{X}_n(0) + \mathbb{X}_n(1))/2 + f(x_0)/2$ . Since the limiting distributions along different subsequences are different,  $n^{1/3}(\tilde{F}(x_0) - F(x_0))$  does not converge weakly.

Next, consider the case  $\gamma \in (1/3, 1)$ , for which the above decomposition of  $n^{1/3}(\tilde{F}(x_0) - F(x_0))$  still holds. Since  $n^{1/3}(F(t_l) - F(x_0))$  converges to 0 and  $(\mathbb{X}_n(0), \mathbb{X}_n(1))$  converges weakly to  $g_{\alpha,\beta}(0)(1, 1)$ , we have  $n^{1/3}(\tilde{F}(x_0) - F(x_0))$  converges weakly to  $g_{\alpha,\beta}(0)$  by noticing  $p + q = 1$ . On the other hand, similar to the argument for Theorem 2.5.3,  $(\mathbb{X}(0), \mathbb{X}(1))$  converges weakly to  $g_{\alpha,\beta}(0)(1, 1)$  as  $c$  goes to 0. Thus, the limit distribution for  $\gamma = 1/3$ ,  $q\mathbb{X}(0) + p\mathbb{X}(1) - cpf(x_0)$ , converges weakly to  $g_{\alpha,\beta}(0)$ , the limit distribution for  $\gamma \in (1/3, 1)$ , by noticing that  $cpf(x_0)$  converges to 0 as  $c$  goes to 0. This means that adaptive inference on  $F(x_0)$  can be made for large values of  $\gamma$ .

Finally, consider the case  $\gamma \in (0, 1/3)$ . Similar to the  $R_{\tilde{F}}$  defined before Theorem 2.4.7, here we generically denote  $R_{\tilde{F}} = (p^2 + q^2)^{-1/2}[(N_l + N_r)/2]^{1/2}(\tilde{F}(x_0) - F(x_0))$ . Similar to the argument for Theorem 2.4.7, we have that  $R_{\tilde{F}}$  converges weakly to  $[F(x_0)(1 - F(x_0))]^{1/2}N(0, 1)$  for  $\gamma \in (1/5, 1/3)$ . Since both  $N_l$  and  $N_r$  are asymptotically equivalent to  $cg(x_0)n^{1-\gamma}$ , we conclude that  $c^{1/2}n^{(1-\gamma)/2}(\tilde{F}(x_0) - F(x_0))$  converges weakly to  $(p^2 + q^2)^{1/2}\alpha N(0, 1)$ . On the limiting distribution side, *another decomposition* of  $n^{1/3}(\tilde{F}(x_0) - F(x_0))$  is exploited for  $\gamma = 1/3$ . We have  $n^{1/3}(\tilde{F}(x_0) - F(x_0)) = qn^{1/3}(\hat{F}(t_l) - F(t_l)) + pn^{1/3}(\hat{F}(t_r) - F(t_r)) + qn^{1/3}(F(t_l) - F(x_0)) + pn^{1/3}(F(t_r) - F(x_0))$ . Similar to the argument for Theorem 2.4.12, the above first two terms as a vector converges weakly to  $(\mathbb{X}(0), \mathbb{X}'(0))$  with  $\mathbb{X}'(0) \stackrel{d}{=} \mathbb{X}(0)$ ; the above last two terms converge

to  $-pqcf(x_0)$  and  $pqcf(x_0)$ , respectively. Thus, for  $\gamma = 1/3$ ,  $n^{1/3}(\tilde{F}(x_0) - F(x_0))$  converges weakly to  $q\mathbb{X}(0) + p\mathbb{X}'(0)$ . Now, similar to the argument for Theorem 2.5.1, we have that, as  $c$  goes to  $\infty$ ,  $\sqrt{c}(\mathbb{X}(0), \mathbb{X}'(0))$  converges weakly to  $\alpha(Z_1, Z_2)$ , where  $Z_1$  and  $Z_2$  are independent standard normal distributions. Thus,  $\sqrt{c}(q\mathbb{X}(0) + p\mathbb{X}'(0))$  converges weakly to  $(p^2 + q^2)^{1/2}\alpha N(0, 1)$  as  $c$  goes to  $\infty$ . This means that adaptive inference on  $F(x_0)$  can also be made for values of  $\gamma$  in  $(1/5, 1/3)$ .

Similar to Theorem 2.4.7, for  $\gamma = 1/5$  and  $\gamma \in (0, 1/5)$  with  $f'(x_0) \neq 0$ , we have  $c^{1/2}n^{(1-\gamma)/2}(\tilde{F}(x_0) - F(x_0))$  converges weakly to  $(p^2 + q^2)^{1/2}\alpha N(0, 1) + (1/2)pqc^{5/2}f'(x_0)$  and  $Sign(f'(x_0))\infty$ , respectively. Thus, the adaptive procedure could still be applied with an adjustment for the bias term  $(1/2)pqc^{5/2}f'(x_0)$  for  $\gamma = 1/5$  but is not available for  $\gamma \in (0, 1/5)$  any more. This basically means that when the grid resolution under the new grid setting is very sparse, the adaptive procedure would lose its effectiveness for inference at  $x_0$ . In comparison, in the original grid setting, the adaptive procedure would work for all values of  $\gamma \in (0, 1/3)$  for inference on  $F$  at the point  $t_l$ .

### 2.5.0.2 Proofs

Here, we provide proofs of the main results in this section, apart from the proof of Theorem 2.5.6 which can be established by an extension of the ideas used in the proof of Theorem 2.5.3 and is skipped.

*Proof of Theorem 2.5.1.* For  $k \in \mathbb{Z}$ , let

$$\tilde{h}(k) = \alpha\sqrt{c}W(ck) + \beta c^{5/2}k(1+k), \quad h(k) = \alpha cW(k) + \beta c^{5/2}k(1+k).$$

Then, we have  $\{\tilde{h}(k), k \in \mathbb{Z}\} \stackrel{d}{=} \{h(k), k \in \mathbb{Z}\}$ . Thus,

$$\sqrt{c}\mathcal{S}_c \stackrel{d}{=} LS \circ GCM \{(ck, h(k)), k \in \mathbb{Z}\} (0).$$

Define  $\tilde{\mathcal{S}}_c = \sqrt{c}\mathcal{S}_c$ . Denote

$$A_c = \left\{ \frac{h(k)}{ck} < \frac{h(k+1)}{c(k+1)}, k = 1, 2, \dots \right\},$$

$$B_c = \left\{ \frac{h(-(k-1))}{c(k-1)} < \frac{h(-k)}{ck}, k = 2, 3, \dots \right\}, \quad C_c = \left\{ \frac{h(1)}{c} > \frac{-h(-1)}{c} \right\}.$$

Then, for  $\omega \in A_c B_c C_c$ , it is easy to see  $\tilde{\mathcal{S}}_c = -\alpha W(-1)$ . We will show in Lemma 2.5.7,  $A_c B_c C_c \xrightarrow{P} 1$ . Thus,  $\tilde{\mathcal{S}}_c = \tilde{\mathcal{S}}_c A_c B_c C_c + \tilde{\mathcal{S}}_c (1 - A_c B_c C_c) \xrightarrow{d} -\alpha W(-1) \stackrel{d}{=} \alpha Z$ ,

with  $Z \sim N(0, 1)$ . Therefore,  $\sqrt{c}\mathcal{S}_c \xrightarrow{d} \alpha Z$ . □

*Proof of Theorem 2.5.5.* We have

$$\mathbb{M}_c = \alpha^{-2} (\sqrt{c}\mathcal{S}_c)^2 + c\alpha^{-2} \sum_{i \neq 0} (\mathbb{X}(ci)^2 - \mathbb{Y}(ci)^2) =: T + R.$$

By Theorem 2.5.1, we have  $T \xrightarrow{d} Z^2 \sim \chi_1^2$ . It suffices to show  $R \xrightarrow{P} 0$ . Letting  $A_c$ ,  $B_c$  and  $C_c$  denote the same quantities as in the proof of Theorem 2.5.1 and letting

$$D_c = \left\{ \frac{h(-1) - h(-2)}{-c + 2c} < 0 \right\}, \quad E_c = \left\{ \frac{h(1)}{c} > 0 \right\},$$

for every  $\omega \in A_c B_c C_c D_c E_c$ , we have  $R = 0$ . We will show  $A_c B_c C_c D_c E_c \xrightarrow{P} 1$  in Lemma 2.5.7. Thus,  $R \xrightarrow{P} 0$ , which completes the proof. □

*Lemma 2.5.7.* Each of  $A_c$ ,  $B_c$ ,  $C_c$ ,  $D_c$  and  $E_c$  in the proof of Theorem 2.5.1 and Theorem 2.5.5 converges to 1 in probability.

*Proof.* It is easy to show  $C_c$ ,  $D_c$  or  $E_c$  converges to 1 in probability. The argument that  $A_c$  converges to one in probability is similar to that for  $B_c$  and only the former is established here. In order to show  $P(A_c) \rightarrow 1$ , it suffices to show  $P(A_c^c) \rightarrow 0$ . We have, for each  $k \in \mathbb{Z}$ ,

$$\begin{aligned} & P\left(\frac{h(k)}{ck} \geq \frac{h(k+1)}{c(k+1)}\right) \\ &= P\left(\frac{\alpha W(k)}{k} + \beta c^{3/2}(k+1) \geq \frac{\alpha W(k+1)}{k+1} + \beta c^{3/2}(k+2)\right) \\ &= P\left(\alpha \left[\frac{W(k)}{k} - \frac{W(k+1)}{k+1}\right] \geq \beta c^{3/2}\right) \\ &= P\left(N(0, 1) \geq \alpha^{-1} \beta c^{3/2} \sqrt{k(k+1)}\right) \\ &\leq 2^{-1} \exp\{-2^{-1} \alpha^{-2} \beta^2 c^3 k(k+1)\}, \end{aligned}$$

using the fact that  $W(k)/k - W(k+1)/(k+1) \sim N(0, (k(k+1))^{-1})$  and the inequality  $P(N(0, 1) > x) \leq 2^{-1} \exp\{-2^{-1} x^2\}$  for  $x \geq 0$  (See, for example, < 2 > on Page 317

of [Pollard \(2002\)](#)). Then, we have

$$\begin{aligned} P(A_c^e) &\leq \sum_{k=1}^{\infty} P\left(\frac{h(k)}{ck} \geq \frac{h(k+1)}{c(k+1)}\right) \leq \sum_{k=1}^{\infty} 2^{-1} \exp\{-2^{-1}\alpha^{-2}\beta^2 c^3 k^2\} \\ &\leq 2^{-1} \int_0^{\infty} \exp\{-2^{-1}\alpha^{-2}\beta^2 c^3 x^2\} dx = (\sqrt{2\pi}/4)\alpha\beta^{-1}c^{-3/2} \rightarrow 0, \end{aligned}$$

as  $c \rightarrow \infty$ . Thus,  $P(A_c) \rightarrow 1$ , which completes the proof.  $\square$

*Proof for Theorem 2.5.3.* We want to show that  $\mathcal{S}_c \xrightarrow{d} g_{\alpha,\beta}(0)$ , as  $c \rightarrow 0$ , where  $g_{\alpha,\beta}(0) = LS \circ GCM\{X_{\alpha,\beta}\}(0) = LS \circ GCM\{X_{\alpha,\beta}(t) : t \in \mathbb{R}\}(0)$  and  $\mathcal{S}_c = LS \circ GCM\{\mathcal{P}_c\}(0) = LS \circ GCM\{\mathcal{P}_c(k) : k \in \mathbb{Z}\}(0)$ . Since  $\mathcal{S}_c = \mathcal{S}'_c + \beta c$ , where  $\mathcal{S}'_c = LS \circ GCM\{\mathcal{P}'_c : k \in \mathbb{Z}\}(0)$  and  $\mathcal{P}'_c = \{(ck, \alpha W(ck) + \beta(ck)^2) : k \in \mathbb{Z}\}$ , it is sufficient to show  $\mathcal{S}'_c \xrightarrow{d} g_{\alpha,\beta}(0)$  as  $c \rightarrow 0$ . To make the notation simple and without causing confusion, in the following we still use  $\mathcal{P}_c$  and  $\mathcal{S}_c$  to denote  $\mathcal{P}'_c$  and  $\mathcal{S}'_c$ . Also, it will be useful to think of  $\mathcal{P}_c$  as a continuous process on  $\mathbb{R}$  formed by linearly interpolating the points  $\{ck, \mathcal{P}_{2,c}(ck) : k \in \mathbb{Z}\}$ , where  $\mathcal{P}_{2,c}(ck) = \alpha W(ck) + \beta(ck)^2 = X_{\alpha,\beta}(ck)$ . Note that viewing  $\mathcal{P}_c$  in this way keeps the GCM unaltered, i.e. the GCM of this continuous linear interpolated version is the same as that of the set of points  $\{ck, \mathcal{P}_{2,c}(ck) : k \in \mathbb{Z}\}$  and the slope-changing points of this piece-wise linear GCM are still grid-points of the form  $ck$ .

Let  $L$  and  $U$  be the largest negative and smallest nonnegative x-axis coordinates of the slope changing points of the GCM of  $X_{\alpha,\beta}$ . Similarly, let  $L_c$  and  $U_c$  be the largest negative and smallest nonnegative x-axis coordinates of the slope changing points of the GCM of  $\mathcal{P}_c$ . For  $K > 0$ , define  $g_{\alpha,\beta}^K(0) = LS \circ GCM\{X_{\alpha,\beta}(t) : t \in [-K, K]\}(0)$  and  $\mathcal{S}_c^K = LS \circ GCM\{\mathcal{P}_c(t) : t \in [-K, K]\}(0)$ .

We will show that, given  $\epsilon > 0$ , there exist  $M_\epsilon > 0$  and  $c(\epsilon)$  such that **(a)** for all  $0 < c < c(\epsilon)$ ,  $P(\mathcal{S}_c^{M_\epsilon} \neq \mathcal{S}_c) < \epsilon$  and **(b)**  $P(g_{\alpha,\beta}^{M_\epsilon}(0) \neq g_{\alpha,\beta}(0)) < \epsilon$ . These immediately imply that both **Fact 1:**  $\lim_{\epsilon \rightarrow 0} \limsup_{c \rightarrow 0} P(\mathcal{S}_c^{M_\epsilon} \neq \mathcal{S}_c) = 0$  and **Fact 2:**  $\lim_{\epsilon \rightarrow 0} P(g_{\alpha,\beta}^{M_\epsilon}(0) \neq g_{\alpha,\beta}(0)) = 0$  hold. We then show that **Fact 3:** For each  $\epsilon > 0$ ,  $\mathcal{S}_c^{M_\epsilon} \xrightarrow{d} g_{\alpha,\beta}^{M_\epsilon}(0)$  holds as well. Then, by Lemma 2.4.18, we have the conclusion  $\mathcal{S}_c \xrightarrow{d} g_{\alpha,\beta}(0)$ . Figure 2.1 illustrates the following argument.

Let  $\tau_{-2} < \tau_{-1} < \tau_1 < \tau_2$  be four consecutive slope changing points of  $G_{\alpha,\beta} = GCM\{X_{\alpha,\beta}\}$  with  $\tau_{-1}$  denoting the first slope changing point to the left of 0 and  $\tau_1$  the first slope changing point to the right. Since  $\tau_{-2}$  and  $\tau_2$  are  $O_P(1)$ , given

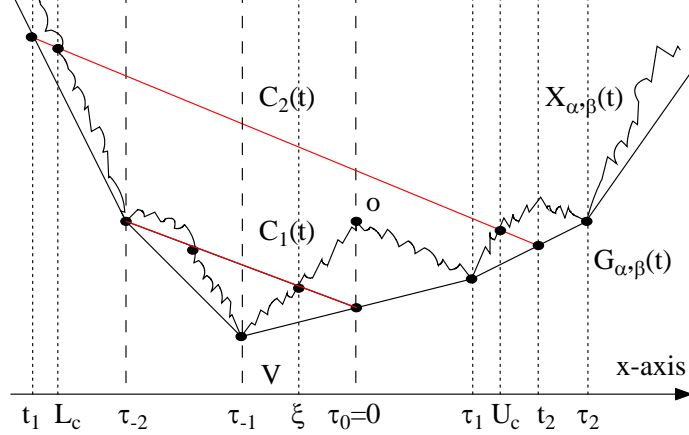


Figure 2.1: An illustration for showing  $\{L_c\}$  is  $O_P(1)$  in the proof of Theorem 2.5.3.

$\epsilon > 0$ , there exists  $M_\epsilon > 0$  such that  $P(-M_\epsilon < \tau_{-2} < \tau_2 < M_\epsilon) > 1 - \epsilon/4$ . Note that the event  $\{g_{\alpha,\beta}^{M_\epsilon}(0) = g_{\alpha,\beta}(0)\} \subset \{-M_\epsilon < \tau_{-2} < \tau_2 < M_\epsilon\}$  and it follows that  $P(g_{\alpha,\beta}^{M_\epsilon}(0) \neq g_{\alpha,\beta}(0)) < \epsilon/4 < \epsilon$ . Thus, (b) holds.

Next, consider the chord  $C_1(t)$  joining  $(0, G_{\alpha,\beta}(0))$  and  $(\tau_{-2}, G_{\alpha,\beta}(\tau_{-2}))$ . By the convexity of  $G_{\alpha,\beta}$  over  $[\tau_{-2}, 0]$  and  $\tau_{-1} \in (\tau_{-2}, 0)$  being a slope changing point,  $X_{\alpha,\beta}(\tau_{-1}) = G_{\alpha,\beta}(\tau_{-1}) < C_1(\tau_{-1})$ . But  $C_1(0) = G_{\alpha,\beta}(0) < X_{\alpha,\beta}(0)$  and it follows by the intermediate value theorem that  $\xi = \inf_{\tau_{-1} < t < 0} \{t : X_{\alpha,\beta}(t) = C_1(t)\}$  is well-defined (since the set in question is non-empty),  $\tau_{-1} < \xi < 0$ ,  $C_1(\xi) = X_{\alpha,\beta}(\xi)$  and on  $[\tau_{-1}, \xi)$ ,  $X_{\alpha,\beta}(t) < C_1(t)$ . Let  $V = \xi - \tau_{-1}$ . Since  $V$  is a continuous and positive random variable, there exists  $\delta(\epsilon) > 0$  such that  $P(V > \delta(\epsilon)) \geq 1 - \epsilon/4$ . Then, the event  $E_\epsilon = \{V > \delta(\epsilon)\} \cap \{-M_\epsilon < \tau_{-2}\}$  has probability larger than  $1 - \epsilon/2$ . For any  $c < c(\epsilon) =: \delta(\epsilon)$ , we claim that  $L_c \geq \tau_{-2}$  on the event  $E_\epsilon$  and the argument for this follows below.

If  $L_c < \tau_{-2}$ , consider the chord  $C_2(t)$  connecting two points  $(L_c, \mathcal{P}_{2,c}(L_c))$  and  $(U_c, \mathcal{P}_{2,c}(U_c))$ . This chord must lie strictly above the chord  $\{C_1(t) : \tau_{-1} \leq t \leq 0\}$  since it can be viewed as a restriction of a chord connecting two points  $(t_1, G_{\alpha,\beta}(t_1))$  and  $(t_2, G_{\alpha,\beta}(t_2))$  with  $t_1 \leq L_c < \tau_{-1} < 0 \leq U_c \leq t_2$ . It then follows that all points of the form  $\{ck, \mathcal{P}_{2,c}(ck) = X_{\alpha,\beta}(ck) : ck \in [L_c, U_c]\}$  must lie above  $C_2(t)$ . But there is at least one  $ck^*$  with  $\tau_{-1} < ck^* < \xi$  and such that  $X_{\alpha,\beta}(ck^*) < C_1(ck^*) < C_2(ck^*)$ , which furnishes a contradiction.

We conclude that for any  $c < c(\epsilon)$ ,  $P(-M_\epsilon < L_c) > 1 - \epsilon/2$ . A similar argument to the right-hand side of 0 shows that for the same  $c$ 's (by the symmetry of two-sided Brownian motion about the origin),  $P(U_c < M_\epsilon) > 1 - \epsilon/2$ . Hence  $P(-M_\epsilon < L_c <$

$U_c < M_\epsilon) > 1 - \epsilon$ . On this event, clearly  $\mathcal{S}_c^{M_\epsilon} = \mathcal{S}_c$  and it follows that for all  $c < c(\epsilon)$ ,  $P(\mathcal{S}_c^{M_\epsilon} \neq \mathcal{S}_c) < \epsilon$ . Thus, (a) also holds and Facts 1 and 2 are established.

It remains to establish Fact 3. This follows easily. For almost every  $\omega$ ,  $X_{\alpha,\beta}(t)$  is uniformly continuous on  $[\pm 2M_\epsilon]$ . It follows by elementary analysis that (for almost every  $\omega$ ) on  $[\pm M_\epsilon]$ , the process  $\mathcal{P}_c$ , being the linear interpolant of the points  $\{ck, X_{\alpha,\beta}(ck) : -M_\epsilon \leq ck \leq M_\epsilon\} \cup \{(-M_\epsilon, \mathcal{P}_{2c}(-M_\epsilon)), (M_\epsilon, \mathcal{P}_{2c}(M_\epsilon))\}$ , converges uniformly to  $X_{\alpha,\beta}$  as  $c \rightarrow 0$ . Thus, the left slope of the GCM of  $\{\mathcal{P}_c(t) : t \in [\pm M_\epsilon]\}$ , which is precisely  $\mathcal{S}_c^{M_\epsilon}$ , converges to  $g_{\alpha,\beta}^{M_\epsilon}(0)$  since the GCM of the restriction of  $X_{\alpha,\beta}$  to  $[\pm M_\epsilon]$  is almost surely differentiable at 0 (see, for example, the Lemma on Page 330 of [Robertson et al. \(1988\)](#) for a justification of this convergence).  $\square$

## 2.6 A Practical Procedure and Simulations

In this section, we provide a practical version of the adaptive procedure introduced in Section 5 to construct Wald-type confidence intervals for  $F(t_l)$ .

Recall that, in the adaptive procedure, we always specify  $\gamma = 1/3$  and thus estimate the value of  $c$  by a solution  $\hat{c}$  of the equation  $K = \lfloor (b-a)/\hat{c}n^{-1/3} \rfloor$ , where  $K$  is the number of grid points. To construct a  $1 - 2\eta$  confidence interval for  $F(t_l)$ , quantiles of  $\mathbb{X}_c(0)$  are needed. Since  $\mathbb{X}_c(0) = LS \circ GCM \{T(k), k \in \mathbb{Z}\} (0)$  ( $c$  is genetically used), we approximate  $\mathbb{X}_c(0)$  with

$$\mathbb{X}_{c,K_a}(0) = LS \circ GCM \{\mathcal{P}_c(k), k \in [-K_a - 1, K_a]\} (0)$$

for some large  $K_a \in \mathbb{N}$ . Further, since

$$\mathbb{X}_{c,K_a}(0) = LS \circ GCM \{(\mathcal{P}_{1,c}(k)/c, \mathcal{P}_{2,c}(k)/c), k \in [-K_a - 1, K_a]\} (0)$$

where  $\mathcal{P}_{1,c}(k)/c = k$  and  $\mathcal{P}_{2,c}(k)/c = \alpha W(ck)/c + \beta ck(1+k)$ , we have that  $\mathbb{X}_{c,K_a}(0)$  is the isotonic regression at  $k = 0$  of the data

$$\begin{aligned} & \{(k, \mathcal{P}_{2,c}(k)/c - \mathcal{P}_{2,c}(k-1)/c), k \in [-K_a, K_a]\} \\ &= \{(k, \alpha Z_k/\sqrt{c} + 2\beta ck), k \in [-K_a, K_a]\}, \end{aligned}$$

where  $\{Z_k\}_{k=-K_a}^{K_a}$  are i.i.d. from  $N(0, 1)$ ,  $\alpha = \sqrt{F(x_0)(1-F(x_0))/g(x_0)}$  and  $\beta = f(x_0)/2$ . To make this adaptive procedure practical, we next consider the estimation



of  $\alpha$  and  $\beta$ , or equivalently, the estimation of  $F(x_0)$ ,  $g(x_0)$  and  $f(x_0)$ .

First, we consider the estimation of  $F(x_0)$  and  $g(x_0)$ . Although  $F(x_0)$  can be consistently estimated by  $\hat{F}(t_l)$ , it is usually better to estimate  $F(x_0)$  by  $\rho\hat{F}(t_l) + (1 - \rho)\hat{F}(t_r)$  with  $\rho = (x_0 - t_l)/(t_r - t_l) \in [0, 1)$ . To estimate  $g(x_0)$ , suppose the design density  $g$  is constant within a small interval around  $x_0$ , which is chosen to be  $[t_{l-j^*}, t_{r+j^*}]$ , where  $j^*$  is defined below in the estimation of  $f(x_0)$ . Then, from the estimating equation  $(N_{l-j^*+1} + \dots + N_{r+j^*})/n = g(x_0)(t_{r+j^*} - t_{l-j^*})$ , one simple but consistent estimator of  $g(x_0)$  is given by  $\hat{g}(x_0) = (N_{l-j^*+1} + \dots + N_{r+j^*})/[n(t_{r+j^*} - t_{l-j^*})]$ .

Next, we consider the estimation of  $f(x_0)$ . In fact, we estimate  $f(t_l)$  using a local linear approximation. First, we identify a small interval around  $t_l$  and then approximate  $F$  over this interval by a line, whose slope gives the estimator of  $f(t_l)$ . We determine the interval according to the following several requirements. First, the sample proportion  $p_n$  in the interval should be larger than the sample proportion at each grid point, which is of order  $n^{-\gamma}$  for  $\gamma \in (0, 1)$ . For example, setting  $p_n$  be of order  $1/\log(n)$  theoretically ensures a sufficiently large interval. Second, for simplicity, we make the interval symmetric around  $t_l$ . Third, in order to obtain a positive estimate (since  $f(t_l)$  is positive), we symmetrically enlarge the interval satisfying the above two requirements until the values of  $\hat{F}$  at the two ends of the interval become different. More specifically, we first find  $j^*$ , which is the smallest integer such that  $\sum_{i=l-j^*}^{l+j^*} N_i/n \geq 1/\log(n)$ . Finally, we find  $i^*$ , which is the smallest integer larger than  $j^*$  such that  $\hat{F}(t_{l-i^*}) < \hat{F}(t_{l+i^*})$ . After identifying the interval  $[t_{l-i^*}, t_{l+i^*}]$ , we fit a line over this interval by weighted least squares. More specifically, we compute

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0, \beta_1) \in \mathbb{R}^2}{\operatorname{argmax}} \left\{ \sum_{i=l-i^*}^{l+i^*} \left( \hat{F}(t_i) - \beta_0 - \beta_1 t_i \right)^2 N_i \right\},$$

and then estimate  $f(t_l)$  (and  $f(x_0)$ ) by  $\hat{\beta}_1$ . Once these nuisance parameters have been estimated, the practical adaptive procedure can be implemented.

To evaluate its finite sample performance in simulations, we also provide simulated confidence intervals of an idealized adaptive procedure where the true values of the parameters  $F(x_0)$ ,  $g(x_0)$  and  $f(x_0)$  are used, but  $\gamma$  is still practically assumed to be  $1/3$  and  $c$  is taken as the previous  $\hat{c}$ . These confidence intervals can be considered as the best Wald-type confidence intervals based on the adaptive procedure.

The simulation settings are as follows: The sampling interval  $[a, b]$  is  $[0, 1]$ . The design density  $g$  is uniform on  $[a, b]$ . The distribution of  $T$  is the uniform distribution over  $[a, b]$  or the exponential distribution with  $\lambda = 1$  or  $2$ . The point of interest  $x_0$  is  $0.5$ . The pair of two controlling parameters  $(\gamma, c)$  takes values  $(1/6, 1/6)$ ,  $(1/4, 1/4)$ ,  $(1/3, 1/2)$ ,  $(1/2, 1)$ ,  $(2/3, 2)$  or  $(3/4, 3)$ . The sample size  $n$  is from  $100$  to  $1000$  by  $100$ . When generating the quantiles of  $\mathbb{X}_{\hat{c}}(0)$ ,  $K_a$  is set to be  $300$  and the corresponding iteration number  $3000$ . We are interested in constructing  $95\%$  confidence intervals for  $F(t_l)$ . The iteration number for each simulation is  $3000$ .

Denote the simulated coverage rates and average lengths for the practical procedure as  $\text{CR}(\text{P})$  and  $\text{AL}(\text{P})$  and those for the theoretical procedure as  $\text{CR}(\text{T})$  and  $\text{AL}(\text{T})$ . Figure 2.2 contains the plots of  $\text{CR}(\text{P})$ ,  $\text{CR}(\text{T})$ ,  $\text{AL}(\text{P})$  and  $\text{AL}(\text{T})$  and Table 2.1 contains the corresponding numerical values for  $n$  being  $n_1 = 100$ ,  $n_2 = 300$  or  $n_3 = 500$ . The first plot of Figure 2.2 shows that both  $\text{CR}(\text{T})$  and  $\text{CR}(\text{P})$  are usually close to the nominal level  $95\%$  from below and  $\text{CR}(\text{T})$  are generally about  $1\%$  better than  $\text{CR}(\text{P})$ . This reflects the price of not knowing the true values of the parameters  $F(x_0)$ ,  $g(x_0)$  and  $f(x_0)$  in the practical procedure. On the other hand, the second plot of Figure 2.2 shows that the  $\text{AL}(\text{P})$ s are usually slightly shorter than  $\text{AL}(\text{T})$ s. This indicates that the practical procedure is slightly more aggressive. As the sample size increases, the coverage rates usually approach the nominal level and the average lengths also become shorter, as expected.

The patterns noted above show up in more extensive simulation studies, not shown here owing to constraints of space. Also, the adaptive procedure is seen to compete well with the asymptotic approximations that one would use for constructing CIs *were*  $\gamma$  known. Of course, for extreme values of  $\gamma$  (close to  $0$  or  $1$ ), the likelihood ratio based confidence intervals using the relevant asymptotic approximations (i.e.  $\chi_1^2$  in the small  $\gamma$  and  $\mathbb{D}$  in the large  $\gamma$  settings) in the known  $\gamma$  case systematically outperform the adaptive ones and also enjoy the advantage of being constructed without nuisance parameter estimation but that is hardly surprising since extreme values of  $\gamma$  correspond to a ‘black and white’ situation, while moderate  $\gamma$ ’s correspond to the ‘grey’ area and pose greater challenges to estimation and it is here that the adaptive procedure is most useful.

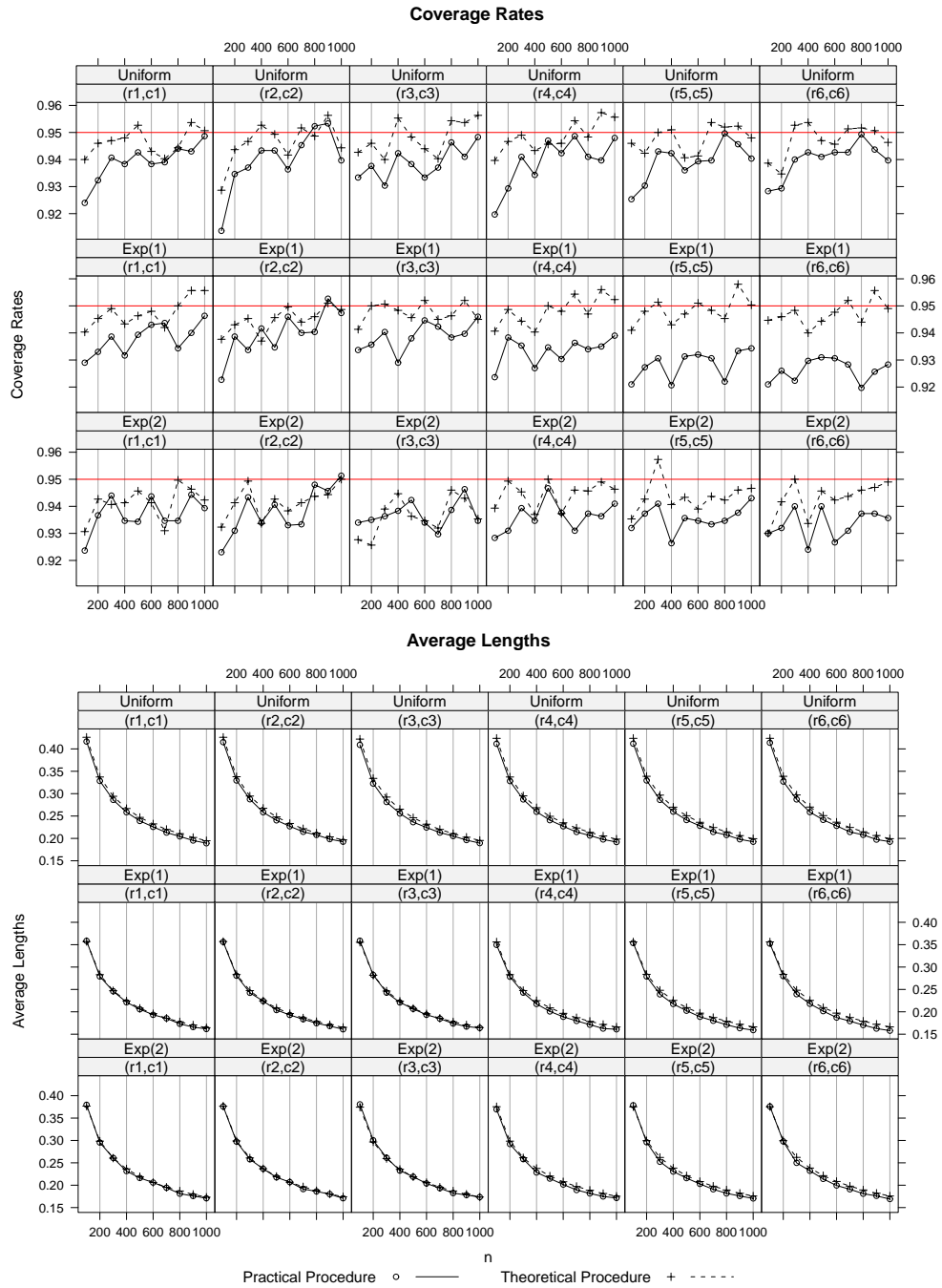


Figure 2.2: A comparison of the coverage rates and average lengths of the practical and theoretical procedures, where  $(r_i, c_i)$  for  $i = 1, \dots, 6$  are  $(1/6, 1/6)$ ,  $(1/4, 1/4)$ ,  $(1/3, 1/2)$ ,  $(1/2, 1)$ ,  $(2/3, 2)$  or  $(3/4, 3)$ , respectively. The sample size  $n$  varies from 100 to 1000 by 100.

Table 2.1: A comparison of the coverage rates and average lengths of the practical procedure with those of the theoretical procedure, where  $U[0, 1]$  and  $\exp(\lambda)$  stand for the uniform distribution over  $[0, 1]$  and the exponential distributions with the parameter  $\lambda$ , and  $n_1, n_2$  and  $n_3$  are 100, 300, and 500, respectively.

Coverage Rates									
CR(P)	$U[0, 1]$			$\exp(1)$			$\exp(2)$		
$(\gamma, c)$	$n_1$	$n_2$	$n_3$	$n_1$	$n_2$	$n_3$	$n_1$	$n_2$	$n_3$
(1/6, 1/6)	.924	.941	.943	.929	.939	.939	.924	.944	.934
(1/4, 1/4)	.914	.937	.943	.923	.934	.935	.923	.943	.941
(1/3, 1/2)	.933	.930	.938	.934	.940	.938	.934	.936	.942
(1/2, 1)	.920	.941	.947	.924	.935	.935	.928	.939	.947
(2/3, 2)	.925	.943	.936	.921	.931	.931	.932	.941	.936
(3/4, 3)	.928	.940	.941	.921	.922	.931	.930	.940	.940
CR(T)	$U[0, 1]$			$\exp(1)$			$\exp(2)$		
$(\gamma, c)$	$n_1$	$n_2$	$n_3$	$n_1$	$n_2$	$n_3$	$n_1$	$n_2$	$n_3$
(1/6, 1/6)	.940	.947	.953	.940	.949	.946	.931	.941	.946
(1/4, 1/4)	.929	.947	.949	.938	.945	.946	.932	.949	.943
(1/3, 1/2)	.943	.940	.948	.941	.951	.946	.928	.939	.936
(1/2, 1)	.940	.949	.946	.941	.944	.950	.939	.945	.950
(2/3, 2)	.946	.950	.941	.941	.951	.947	.935	.957	.943
(3/4, 3)	.939	.953	.947	.945	.948	.944	.930	.950	.946
Average Lengths									
AL(P)	$U[0, 1]$			$\exp(1)$			$\exp(2)$		
$(\gamma, c)$	$n_1$	$n_2$	$n_3$	$n_1$	$n_2$	$n_3$	$n_1$	$n_2$	$n_3$
(1/6, 1/6)	.417	.286	.239	.358	.246	.206	.380	.261	.216
(1/4, 1/4)	.415	.287	.240	.356	.242	.204	.376	.258	.218
(1/3, 1/2)	.409	.281	.236	.359	.243	.207	.381	.258	.219
(1/2, 1)	.411	.287	.241	.350	.243	.201	.370	.258	.215
(2/3, 2)	.411	.286	.241	.354	.239	.202	.379	.253	.216
(3/4, 3)	.414	.287	.241	.352	.239	.202	.376	.250	.214
AL(T)	$U[0, 1]$			$\exp(1)$			$\exp(2)$		
$(\gamma, c)$	$n_1$	$n_2$	$n_3$	$n_1$	$n_2$	$n_3$	$n_1$	$n_2$	$n_3$
(1/6, 1/6)	.426	.294	.247	.357	.247	.208	.377	.260	.219
(1/4, 1/4)	.426	.295	.248	.357	.247	.208	.377	.261	.220
(1/3, 1/2)	.422	.292	.246	.355	.246	.208	.374	.260	.219
(1/2, 1)	.424	.295	.249	.356	.247	.209	.375	.261	.220
(2/3, 2)	.424	.297	.251	.356	.248	.209	.375	.262	.221
(3/4, 3)	.424	.297	.251	.356	.248	.209	.375	.262	.221

## 2.7 Conclusions

In this paper, we have considered isotonic nonparametric estimation and hypothesis testing for the survival function at a point in the current status model with i.i.d. data. The design density for the covariate is assumed to be an equally spaced grid distribution with the grid resolution being  $\delta = cn^{-\gamma}$  for  $c > 0$  and  $\gamma > 0$  and incorporates situations where there are systematic ties in the observation times of the entities involved and the number of distinct observation times can increase with the sample size.

The asymptotic properties of the isotonic regression estimator and the likelihood ratio test statistic depend critically on the order of the grid resolution  $\gamma$ . For  $\gamma \in (0, 1/3)$ , the asymptotic distributions are normal and chi-squared, which are the standard limit distributions in parametric problems with finite number of unknown parameters; for  $\gamma \in (1/3, \infty)$ , the asymptotic distributions are Chernoff and the so-called  $\mathbb{D}$ , which are the standard limit distributions in isotonic regression problems with continuous design densities. Thus, when  $\gamma \in (0, 1/3)$ , the grid is so sparse that the nonparametric problem is essentially reduced to a parametric one. On the other hand, when  $\gamma \in (1/3, \infty)$ , the grid is dense enough that the observation time can be viewed as an absolutely continuous random variable. For the most interesting boundary case with  $\gamma = 1/3$ , the grid is, in some sense, neither too sparse nor too dense and the asymptotic distributions are  $\mathcal{S}_c$  and  $\mathbb{M}_c$ , functionals of the unconstrained and constrained GCMs of discrete time stochastic processes which depend on  $c$ , the scaling parameter in the grid resolution.

The limit distributions  $\mathcal{S}_c$  and  $\mathbb{M}_c$  are different from those obtained in the other two cases. However, as  $c$  goes to  $\infty$ ,  $\mathcal{S}_c$  and  $\mathbb{M}_c$  converge in distribution to the normal and chi-squared distributions; as  $c$  goes to 0,  $\mathcal{S}_c$  and  $\mathbb{M}_c$  converge in distribution to the Chernoff's and  $\mathbb{D}$  distributions. These weak convergence results allow the approximation of the extreme distributions with the boundary ones, by adjusting the value of the scaler  $c$ , and lead to an adaptive procedure for statistical inferences which obviates the need to estimate or specify the order  $\gamma$ , and therefore provides a powerful inferential tool. Note that while we have considered grid resolutions of the order  $n^{-\gamma}$ , the derivations in this paper are easily generalizable to arbitrary grid resolutions that converge to 0 with  $n$ . So long as the resolution  $r_n$  satisfies  $n^{-1/3} = o(r_n)$ , standard asymptotics prevail while Chernoff-type asymptotics are obtained when  $r_n = o(n^{-1/3})$ . However the class of grid resolutions of order  $n^{-\gamma}$  for  $\gamma \in (0, \infty)$  is rich enough that

almost any setting with tied observation times can be viewed as coming from such a scenario.

The results in this paper reveal some new directions for future research. As touched upon in the introduction, some recent related work by [Maathuis and Hudgens \(2010\)](#) deals with the estimation of *competing risks current status data* with discrete (or grouped) observation times. A natural question of interest, then, is what happens if the observation times in their paper are supported on grids of increasing size as considered in this paper for simple current status data. Does  $\gamma = 1/3$  also form the boundary between normal and non-normal asymptotics as with the simpler current status model? Can an adaptive procedure of similar vein be devised for current status data with competing risks? One could also consider the problem of grouped current status data (with and without the element of competing risks), where the observation times are not exactly known but grouped into bins. Based on communications with us and preliminary versions of this paper, [Maathuis and Hudgens \(2010\)](#) conjecture that for grouped current status data *without competing risks*, one may expect findings similar to those in this paper, depending on whether the number of groups increases at rate  $n^{1/3}$  or at a faster/slower rate. Whether similar phenomena would arise for grouped current status data with competing risks is again, unclear, though  $\gamma = 1/3$  is certainly not an un-natural candidate for the transition from normality to non-normality.

Viewed as a regression model, the current status model is a monotone binary regression model and the results obtained here translate almost directly to results on monotone binary regression. In fact, it is fairly clear that the adaptive inference scheme proposed in this paper will apply to monotone regression models with discrete covariates in general. In particular, the very general conditionally parametric response models studied in [Banerjee \(2007\)](#) under the assumption of a continuous covariate can be handled for the discrete covariate case as well by adapting the methods of this paper and the adaptive procedure can be made to work similarly. Furthermore, similar adaptive inference in more complex forms of interval censoring, like Case-2 censoring or mixed-case censoring (see, for example, [Sen and Banerjee \(2006\)](#) and [Schick and Yu \(2000\)](#)), should also be possible in situations where the multiple observation times are discrete-valued. Finally, we conjecture that phenomena similar to those revealed in this paper will appear in nonparametric regression problems with grid-supported covariates under more complex shape constraints (like convexity, for example), though

the boundary value of  $\gamma$  as well as the nature of the non-standard limits will be different and will depend on the ‘order’ of the shape constraint.

## 2.8 Appendix

In this appendix, we provide proofs of some of the technical results.

### 2.8.1 Supplementary proofs for Case One $\gamma \in (0, 1/3)$

First, we establish two useful lemmas.

*Lemma 2.8.1.* If  $\gamma \in (0, 1)$  and **(A1.2)** holds, for every  $\eta \in (0, 1)$ , we have

$$\mathbb{P} \left( \bigcap_{i=1}^K \{N_i \geq \eta m_l\} \right) \rightarrow 1.$$

*Proof.* Denote the event of interest as  $A_n$ . Then,  $A_n^c = \bigcup_{i=1}^K \{N_i < \eta m_l\}$  and it is sufficient to show  $\mathbb{P}(A_n^c) \rightarrow 0$ . Let  $t = -\log(\eta) > 0$ , we have

$$\begin{aligned} \mathbb{P}(N_i < \eta m_l) &= \mathbb{P} \left( \sum_{j=1}^n \{X_j = t_i\} < \eta m_l \right) \leq e^{t\eta m_l} \mathbb{P} e^{-t \sum_{j=1}^n \{X_j = t_i\}} \\ &\leq e^{t\eta m_l} [1 - (1 - e^{-t})g_l \delta]^n \leq e^{t\eta m_l} e^{-(1-e^{-t})g_l \delta n} \\ &\leq e^{m_l(\eta t - 1 + e^{-t})} = e^{m_l(-\eta \log(\eta) - 1 + \eta)}. \end{aligned}$$

The third inequality above exploits the fact  $1 - x \leq e^{-x}$  for  $x > 0$ . Note that  $-\eta \log(\eta) - 1 + \eta < 0$  for every  $\eta \in (0, 1)$ ,  $K \sim (b-a)c^{-1}n^\gamma$  and  $m_l = g_l c n^{1-\gamma}$ . Then, we have

$$\mathbb{P}(A_n^c) \leq \sum_{i=1}^K \mathbb{P}(N_i < \eta m_l) \leq K e^{m_l(-\eta \log(\eta) - 1 + \eta)} \rightarrow 0,$$

which completes the proof.  $\square$

Next, we introduce two related binary regression models which yield sufficient statistics whose distributions are identical to the sufficient statistics in the current status model. The first model is as follows. Suppose  $\{t_i\}_{i=1}^K$  and  $X$  are defined as before. Let  $\{X_j\}_{j=1}^n$  be i.i.d. copies of  $X$  and  $N_i = \sum_{j=1}^n \{X_j = t_i\}$  for  $i = 1, 2, \dots, K$ . Given  $\{N_i\}_{i=1}^K$ , for each  $i$  draw an i.i.d. sample  $\{Y_{ij}\}_{j=1}^{N_i}$  from  $Bernoulli(1, F_i)$ . Denote  $\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} Y_{ij}$ , for each  $i$ . The second model is as follows. Suppose  $\{t_i\}_{i=1}^K$ ,  $X$ ,  $\{X_j\}_{j=1}^n$  and  $\{N_i\}_{i=1}^K$  are defined as before. Let  $\{Y'_{ij} : 1 \leq i \leq K, 1 \leq j \leq n\}$  be a

family of mutually independent random variables, distributed independently of the variables in the previous sentence, such that for each  $i$ ,  $Y'_{ij}$  follows  $Bernoulli(1, F_i)$  for  $1 \leq j \leq n$ . Denote  $\bar{Y}'_i = N_i^{-1} \sum_{j=1}^n Y'_{ij} \{X_j = t_i\}$  for each  $i$ . Then, we have the following equalities in distribution, which will be used in other proofs. Its proof can be shown by straightforward comparing distributions and omitted here.

$$\text{Lemma 2.8.2. } (\{N_i\}, \{\bar{Z}_i\}) \stackrel{d}{=} (\{N_i\}, \{\bar{Y}_i\}) \stackrel{d}{=} (\{N_i\}, \{\bar{Y}'_i\})$$

*Proof of Proposition 2.4.9.* It suffices to show  $\mathbb{P}(\bar{Y}_1 \leq \bar{Y}_2 \leq \dots \leq \bar{Y}_K) \rightarrow 1$  by Lemma 2.8.2, which is equivalent to show  $\mathbb{P}(\cup_{i=1}^{K-1} \{\bar{Y}_i > \bar{Y}_{i+1}\}) \rightarrow 0$ . Denote this probability as  $T$ . We then have  $T = T_1 + T_2$ , where

$$T_1 = \mathbb{P}(\cup_{i=1}^{K-1} \{\bar{Y}_i > \bar{Y}_{i+1}\}, \cap_{i=1}^K \{N_i \geq \eta m_l\}), \text{ with } \eta > 0$$

and  $T_2 = T - T_1$ . Since, by Lemma 2.8.1,  $T_2 \leq \mathbb{P}(\cup_{i=1}^K \{N_i < \eta m_l\}) \rightarrow 0$ , it remains to show  $T_1 \rightarrow 0$ .

On one hand, we have  $T_1 \leq \sum_{i=1}^{K-1} \sum_A S_i \mathbb{P}(N_i = n_i, i = 1, 2, \dots, K)$ , where

$$S_i = \mathbb{P}\left(\frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} > \frac{1}{n_{i+1}} \sum_{j=1}^{n_{i+1}} Y_{(i+1)j}\right)$$

and  $A = \{(n_1, n_2, \dots, n_K) \in \mathbb{N}^K : \sum_{i=1}^K n_i = n, n_i \geq \eta m_l \text{ for } i = 1, 2, \dots, K\}$ . On the other hand, we will show that, for each  $i = 1, \dots, K-1$ ,

$$S_i \leq 2 \exp\left\{-\frac{\eta f_l^2}{16} \delta^2 m_l\right\}. \quad (2.30)$$

Then, the result follows by noticing that  $\delta^2 m_l \sim c^3 g_l n^{1-3\gamma} \rightarrow \infty$  for  $\gamma \in (0, 1/3)$ .

Next, we show (2.30). Denote  $Z_{ij} = Y_{ij} - F_i$ ,  $Z'_{ij} = -Z_{ij}$  for  $i = 1, \dots, K$  and  $j = 1, \dots, N_i$  and  $\Delta_i = F_{i+1} - F_i$  for  $i = 1, \dots, K-1$ . Then we have

$$S_i \leq \mathbb{P}\left(\frac{1}{n_i} \sum_{j=1}^{n_i} Z_{ij} > \frac{\Delta_i}{2}\right) + P\left(\frac{1}{n_{i+1}} \sum_{j=1}^{n_{i+1}} Z'_{(i+1)j} > \frac{\Delta_i}{2}\right) =: S_{i1} + S_{i2}.$$

In a similar way, it can be shown that both  $S_{i1}$  and  $S_{i2}$  are less than or equal to  $\exp\{-\eta f_l^2 \delta^2 m_l / 16\}$ . Next, we only show the inequality on  $S_{i1}$ .



For every  $t > 0$ , we have

$$S_{i1} \leq e^{-\frac{\Delta_i}{2}n_it}(\mathbb{P}e^{tZ_{i1}})^{n_i} \leq \exp\{-\phi(t)n_i\},$$

where  $\phi(t) = t(F_i + F_{i+1})/2 - F_i(e^t - 1)$ . Note that the maximizer of  $\phi(t)$  is  $t_i^* = \log((F_i + F_{i+1})/2F_i) > 0$  and that the maximum of  $\phi$  is

$$\phi(t_i^*) = \frac{F_i + F_{i+1}}{2} \log \frac{F_i + F_{i+1}}{2F_i} - \frac{F_{i+1} - F_i}{2}.$$

Thus, we have  $S_{i1} \leq \exp\{-\phi(t_i^*)n_i\}$ . For  $x \in [t_1, t_{K-1}]$ , define

$$h(x) = \frac{F(x) + F(x + \delta)}{2} \log \frac{F(x) + F(x + \delta)}{2F(x)} - \frac{F(x + \delta) - F(x)}{2}.$$

Then, we have  $\phi(t^*) \geq \inf_{x \in [t_1, t_{K-1}]} h(x)$ . Since  $n_i \geq \eta m_l$ , it is sufficient to show

$$\inf_{x \in [t_1, t_{K-1}]} h(x) \geq (f_l \delta)^2 / 16. \quad (2.31)$$

By the assumption **(A1.1)**, there exists  $\xi \in (x, x + \delta)$  such that  $F(x + \delta) = F(x) + f(\xi)\delta$ . Then, we have

$$\begin{aligned} h(x) &= \left( F(x) + \frac{f(\xi)\delta}{2} \right) \log \left( 1 + \frac{f(\xi)\delta}{2F(x)} \right) - \frac{f(\xi)\delta}{2} \\ &= \left( F(x) + \frac{f(\xi)\delta}{2} \right) \left( \frac{f(\xi)\delta}{2F(x)} - \frac{1}{2(1+\zeta)^2} \frac{f(\xi)^2}{4F(x)^2} \delta^2 \right) - \frac{f(\xi)\delta}{2} \\ &= \frac{f^2(\xi)}{4F(x)} \delta^2 - \frac{1}{(1+\zeta)^2} \frac{f^2(\xi)}{8F(x)} \delta^2 - \frac{1}{(1+\zeta)^2} \frac{f^3(\xi)}{16F^2(x)} \delta^3 \\ &\geq \frac{f^2(\xi)}{4F(x)} \delta^2 - \frac{f^2(\xi)}{8F(x)} \delta^2 - \frac{f^3(\xi)}{16F^2(x)} \delta^3 \geq \frac{f^2(\xi)}{16F(x)} \delta^2 \geq \frac{f_l^2}{16} \delta^2, \end{aligned}$$

where the Taylor's expansion of  $\log(1+x)$  around 0 is utilized in the second equality and  $\zeta \in (0, f(\xi)\delta/(2F(x)))$ ; the assumptions **(A1.1)** and **(A1.3)** and  $F(x) \leq 1$  are exploited for the last two inequalities.  $\square$

*Proof of Proposition 2.4.10.* The proof follows very similar lines as that of Proposition 2.4.9 and is therefore skipped.  $\square$

*Proof of Proposition 2.4.11.* By Proposition 2.8.2, it suffices to show the results with  $\bar{Y}'_l$  and  $\bar{Y}'_r$  replacing  $\bar{Z}_l$  and  $\bar{Z}_r$ . We have  $\bar{Y}'_l - F_l = T_1/T_2$ , where  $T_1 = n^{-1} \sum_{j=1}^n p_l^{-1} 1(X_j =$

$t_l)(Y'_{lj} - F_l)$  and  $T_2 = n^{-1} \sum_{j=1}^n p_l^{-1} 1(X_j = t_l)$ . By Chebyshev's Inequality, it is easy to check that  $T_1$  and  $T_2$  converge to 0 and 1 in probability, respectively. Then  $\bar{Z}_l - F_l$  converges to 0 in probability by Slutsky's Lemma. Similarly,  $\bar{Z}_r - F_r$  converges to 0 in probability.

Next, we show weak convergence. Denote  $Z_{nj} = (Z_{njl}, Z_{njr})$ , where

$$\begin{aligned} Z_{njl} &= p_l^{-1/2} 1(X_j = t_l)(Y'_{lj} - F_l) / \sqrt{F_l(1 - F_l)}, \\ Z_{njr} &= p_r^{-1/2} 1(X_j = t_r)(Y'_{rj} - F_r) / \sqrt{F_r(1 - F_r)}, \end{aligned}$$

for  $j = 1, 2, \dots, n$ . Then  $\{Z_{nj}\}$  are independent and identically distributed. It is easy to check that the means of  $Z_{njl}$  and  $Z_{njr}$  are 0, the variances are 1 and the covariance is 0. Then, by a triangular array version of multivariate central limit theorem (see Proposition 2.27 in [van der Vaart \(1998\)](#)), in order to show  $n^{-1/2} \sum_{i=1}^n Z_{nj} \xrightarrow{d} N(0, I_2)$ , it is sufficient to check the Lindeberg condition: for each  $\epsilon > 0$ ,

$$\sum_{j=1}^n E \|n^{-1/2} Z_{nj}\|^2 \{ \|n^{-1/2} Z_{nj}\| > \epsilon \} \rightarrow 0.$$

Since

$$\sum_{j=1}^n E \|n^{-1/2} Z_{nj}\|^2 \{ \|n^{-1/2} Z_{nj}\| > \epsilon \} \leq \frac{1}{n\epsilon^2} E \|Z_{n1}\|^4 = \frac{1}{n\epsilon^2} E[Z_{n1l}^4 + Z_{n1r}^4],$$

$$E[Z_{n1l}^4] = \frac{[F_l^3 + (1 - F_l)^3]}{p_l F_l(1 - F_l)} \leq \frac{4}{g_l \delta F(x_0)(1 - F(x_0))},$$

and similar inequality holds for  $E[Z_{n1r}^4]$ , we have

$$\sum_{j=1}^n E \|n^{-1/2} Z_{nj}\|^2 \{ \|n^{-1/2} Z_{nj}\| > \epsilon \} \leq \frac{8}{\epsilon^2 g_l F(x_0)(1 - F(x_0))} \cdot \frac{1}{cn^{1-\gamma}} \rightarrow 0.$$

Thus, the Lindeberg condition holds.

Denote  $T_n = (T_{nl}, T_{nr})$ , where  $T_{nl} = N_l^{1/2}(\bar{Y}'_l - F_l)$  and  $T_{nr} = N_r^{1/2}(\bar{Y}'_r - F_r)$ . Then, we have  $T_n = \beta(x_0)n^{-1/2} \sum_{i=1}^n Z_{nj} + R_n$ , where

$$\begin{aligned} R_n &= \left( (\eta_{nl} - \beta(x_0)) \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{njl}, (\alpha_{nr} - \beta(x_0)) \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{njr} \right), \\ \beta(x_0) &= \sqrt{F(x_0)(1 - F(x_0))}, \quad \alpha_{nl} = [F_l(1 - F_l)]^{1/2} / [N_l / (np_l)]^{1/2}, \end{aligned}$$

and  $\alpha_{nr}$  is similarly defined. Then, by Slutsky's Lemma, it suffices to show both  $\alpha_{nl} - \beta(x_0)$  and  $\alpha_{nr} - \beta(x_0)$  converge to 0 in probability. Here we only show the former and the latter can be shown in a similar manner. From the continuity of  $F$  at  $x_0$ ,  $[F_l(1 - F_l)]^{1/2} \rightarrow \beta(x_0)$ . Notice that  $N_l/(np_l)$  is exactly  $T_2$  and  $N_l/(np_l) \rightarrow 1$  in probability. This completes the proof.  $\square$

*Proof of Lemma 2.4.6.* By the definition of  $\tilde{F}(x_0)$ , we have

$$\begin{aligned}\tilde{F}(x_0) - F(x_0) &= \frac{t_r - x_0}{t_r - t_l}(F_l^* - F_l) + \frac{x_0 - t_l}{t_r - t_l}(F_r^* - F_r) \\ &+ \frac{t_r - x_0}{t_r - t_l}(F_l - F(x_0)) + \frac{x_0 - t_l}{t_r - t_l}(F_r - F(x_0)).\end{aligned}$$

By Lemma 2.4.1, both  $F_l^* - F_l$  and  $F_r^* - F_r$  converge to 0 in probability. For the continuity of  $f$  at  $x_0$ , both  $F_l - F(x_0)$  and  $F_r - F(x_0)$  converge to 0. Note that the absolute values of the coefficients are less than 1. Therefore,  $\tilde{F}(x_0) - F(x_0)$  converges to 0 in probability.  $\square$

*Proof of Theorem 2.4.7.* Denote  $\tilde{p}_n = p_n/(p_n^2 + q_n^2)^{1/2}$ ,  $\tilde{q}_n = p_n/(p_n^2 + q_n^2)^{1/2}$ ,  $\alpha_l = [p_l F_l(1 - F_l)]^{1/2}$ ,  $\alpha_r = [p_r F_r(1 - F_r)]^{1/2}$ ,  $\xi_{lj} = \tilde{q}_n(Y_{lj} - F_l)\{X_j = t_l\}/\alpha_l$ ,  $\xi_{rj} = \tilde{p}_n(Y_{rj} - F_r)\{X_j = t_r\}/\alpha_r$ . Then, we have

$$\frac{1}{\sqrt{p_n^2 + q_n^2}}(\tilde{F}(x_0) - F(x_0)) = [\alpha_l \frac{n^{1/2}}{N_l} T_1 + (R_{n1} + R_{n2})] A_n + \tilde{I} A_n^c + \frac{1}{\sqrt{p_n^2 + q_n^2}} II,$$

where  $\tilde{I} = \tilde{q}_n(F_l^* - F_l) + \tilde{p}_n(F_r^* - F_r)$ ,  $II = q_n F_l + p_n F_r - F(x_0)$ ,  $T_1 = n^{-1/2} \sum_{j=1}^n (\xi_{lj} + \xi_{rj})$ ,  $R_{n1} = (\alpha_r - \alpha_l) N_r^{-1} \sum_{j=1}^n \xi_{lj}$  and  $R_{n2} = \alpha_l (N_r^{-1} - N_l^{-1}) \sum_{j=1}^n \xi_{rj}$ . When  $\gamma \in (0, 1/3)$  and **(A1)** hold, by regular argument,  $T_1$  converges to  $N(0, 1)$  in distribution;  $n^{(1-\gamma)/2} \alpha_l n^{1/2}/N_l$  and  $n^{(1-\gamma)/2} (R_{n1} + R_{n2})$  converge to  $[F(x_0)(1 - F(x_0))/(cg(x_0))]^{1/2}$  and 0 in probability, respectively.

Since  $f''$  is bounded in a neighborhood of  $x_0$ , we have

$$\begin{aligned}F_l &= F(x_0) + f(x_0)(t_l - x_0) + \frac{1}{2} f'(x_0)(t_l - x_0)^2 + o((t_l - x_0)^2), \\ F_r &= F(x_0) + f(x_0)(t_r - x_0) + \frac{1}{2} f'(x_0)(t_r - x_0)^2 + o((t_r - x_0)^2).\end{aligned}$$

Denote  $T_2 = (p_n^2 + q_n^2)^{-1/2} n^{(1-\gamma)/2} II$ . Then,

$$\begin{aligned} T_2 &= \frac{1}{2\sqrt{p_n^2 + q_n^2}} f'(x_0) n^{(1-\gamma)/2} (t_r - x_0)(x_0 - t_l) \\ &\quad + o(q_n n^{(1-\gamma)/2} (t_l - x_0)^2 \vee p_n n^{(1-\gamma)/2} (t_r - x_0)^2). \end{aligned}$$

Thus, for  $\gamma \in (1/5, 1/3)$ , we have  $T_2 \rightarrow 0$ ; for  $\gamma = 1/5$  and  $p_n$  converging to some  $p \in (0, 1)$ , we have  $T_2 \rightarrow (1/2)pq(p_n^2 + q_n^2)^{-1/2} c^2 f'(x_0)$ ; Further, for  $\gamma \in (0, 1/5)$  and  $f'(x_0) \neq 0$ , we have  $T_2 \rightarrow \text{Sign}(f'(x_0))\infty$ . Finally, noticing that both  $np_l/N_l$  and  $np_r/N_r$  converge to 1 in probability and that both  $p_l$  and  $p_r$  are asymptotically equivalent to  $g(x_0)cn^{-\gamma}$  completes the proof.  $\square$

### 2.8.2 Proofs for Case Three $\gamma = 1/3$

In what follows we will denote the distribution of the vector  $(X_{1n}, Y_{1n})$  by  $\mathbb{P}$  (suppressing the dependence on  $n$ ) and  $\mathbb{P}_n$  will denote the empirical measure of the sample, i.e.  $n$  i.i.d. observations drawn from  $\mathbb{P}$ .

*Lemma 2.8.3.* Fact 1 in the proof of Lemma 2.4.22 holds.

*Proof.* For  $\beta = 0$ , Fact 1 holds obviously. Next, fix a small interval  $\mathcal{N}_0$  about 0 such that  $f, g, f', g'$  are all uniformly bounded in this neighborhood. Let  $\epsilon > 0$ . Consider  $\beta > 0, \beta \in \mathcal{N}_0$ . We have

$$\begin{aligned} \mathbb{P}g_{n1}(x, y; \beta) &= \mathbb{P}_X \mathbb{P}_{Y|X} (Y - F(t_{l-M})) \{t_{l-M} < X \leq t_{l-M} + \beta\} \\ &= \sum_{j=1}^{\lfloor \beta/\delta \rfloor} (F(t_{l-M+j}) - F(t_{l-M})) (G(t_{l-M+j}) - G(t_{l-M+j-1})) \\ &= \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j f(t_{l-M+j}^{1*}) g(t_{l-M+j}^{2*}) = T_1 + T_2 \end{aligned}$$

where the mean value theorem is applied twice at the third equality,  $t_{l-M+j}^{1*}$  lies in  $[t_{l-M}, t_{l-M+j}]$ ,  $t_{l-M+j}^{2*}$  lies in  $[t_{l-M+j-1}, t_{l-M+j}]$ ,  $T_1 = \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j f(x_0) g(x_0)$  and  $T_2 = \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j [f(t_{l-M+j}^{1*}) g(t_{l-M+j}^{2*}) - f(x_0) g(x_0)]$ . We will show that for  $\beta$  sufficiently small (depending on  $\epsilon$ ) both  $|T_1 - (1/2)f(x_0)g(x_0)\beta^2|$  and  $|T_2|$  are dominated by  $\epsilon\beta^2 + O(n^{-2/3})$  when  $n$  is sufficiently large (depending on  $\epsilon$ ), where the  $O(n^{-2/3})$  terms can again depend on  $\epsilon$ . Then the result follows, as  $\epsilon$  is arbitrary.

First consider  $T_1$ . We have  $T_1 - (1/2)f(x_0)g(x_0)\beta^2 = (1/2)f(x_0)g(x_0)T_{11}$  where

$T_{11} = \delta^2 \lfloor \beta/\delta \rfloor (1 + \lfloor \beta/\delta \rfloor) - \beta^2$ . Since  $\lfloor x \rfloor \in (x - 1, x]$  for  $x \in \mathbb{R}$ , simple algebra gives  $|T_{11}| \leq \beta\delta$  and consequently

$$|T_1 - (1/2)f(x_0)g(x_0)\beta^2| \leq (f(x_0)g(x_0)/2)\beta\delta \leq \epsilon\beta^2 + \tilde{\delta}^2/\epsilon = \epsilon\beta^2 + O(n^{-2/3}),$$

where  $\tilde{\delta} = (f(x_0)g(x_0)/2)\delta$  and we use the fact that  $\delta = cn^{-1/3}$  to get the last expression in the above display. Note that the  $O(n^{-2/3})$  term depends on  $\epsilon$ .

Next consider  $T_2$ . We have  $|T_2| \leq T_{21} + T_{22}$ , where

$$T_{21} = \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j |f(t_{i-M+j}^{1*}) - f(x_0)| g(t_{i-M+j}^{2*}),$$

$$T_{22} = \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j |g(t_{i-M+j}^{2*}) - g(x_0)| f(x_0).$$

Here we only show  $T_{21} \leq \epsilon\beta^2 + O(n^{-2/3})$  and  $T_{22} \leq \epsilon\beta^2 + O(n^{-2/3})$  can be shown similarly. We have

$$\begin{aligned} T_{21} &\lesssim \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j |f(t_{i-M+j}^{1*}) - f(x_0)| = \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j |f'(t_{i-M+j}^{1**})| |t_{i-M+j}^{1*} - x_0| \\ &\lesssim \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j |t_{i-M+j}^{1*} - x_0| \lesssim \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j \lfloor \beta/\delta \rfloor \delta + \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j(M+1)\delta. \end{aligned}$$

Denote the above last two terms as  $T_{211}$  and  $T_{212}$ . The mean value theorem is applied at the second step and  $t_{i-M+j}^{1**}$  lies between  $t_{i-M+j}^{1*}$  and  $x_0$ . In the first and third steps, the assumption that  $g$  and  $f'$  are bounded around  $x_0$  is utilized. Denote the constant associated with the  $\lesssim$  by  $\tilde{K}$ . Then  $T_{21} \leq T_{211} + T_{212}$  with  $T_{211} \leq \tilde{K}\delta^3 \lfloor \beta/\delta \rfloor^2 (1 + \lfloor \beta/\delta \rfloor) \leq \tilde{K}\delta^3 (\beta/\delta)^2 (\beta/\delta + 1) = \tilde{K}\beta^3 + \tilde{K}\beta^2\delta \leq \epsilon\beta^2$ , by choosing  $\beta < \epsilon/\tilde{K}$  and  $n$  sufficiently large (depending on  $\epsilon$ ), and  $T_{212} \leq \tilde{K}\delta^3 \lfloor \beta/\delta \rfloor (1 + \lfloor \beta/\delta \rfloor) \leq \tilde{K}\beta^2\delta + \tilde{K}\beta\delta^2 \leq \epsilon\beta^2 + O(n^{-2/3})$ , again for  $n$  sufficiently large (depending only on  $\epsilon$ ). Thus,  $T_{21} \leq \epsilon\beta^2 + O(n^{-2/3})$ , which completes the proof.  $\square$

*Lemma 2.8.4.* Fact 2 in the proof of Lemma 2.4.22 holds.

*Proof.* For  $\beta = 0$ , Fact 2 holds obviously. Next, suppose  $\beta > 0$  and is restricted to the neighborhood  $\mathcal{N}_0$  from the proof of the previous lemma. The case  $\beta < 0$  can be

handled in the same way. We have

$$\begin{aligned}\mathbb{P}g_{n2}(x; \beta) &= \mathbb{P}\{t_{l-M} < X \leq t_{l-M} + \beta\} = \sum_{j=1}^{\lfloor \beta/\delta \rfloor} (G(t_{l-M+j}) - G(t_{l-M+j-1})) \\ &= \sum_{j=1}^{\lfloor \beta/\delta \rfloor} g(x_0)\delta + \sum_{j=1}^{\lfloor \beta/\delta \rfloor} (g(t_{l-M+j}^*) - g(x_0))\delta,\end{aligned}$$

where the mean value theorem is applied at the last step and  $t_{l-M+j}^*$  lies between  $[t_{l-M+j-1}$  and  $t_{l-M+j}]$ . Denote the above last two terms as  $T_1$  and  $T_2$ .

We have  $|T_1 - g(x_0)\beta| \leq g(x_0)|\delta\lfloor \beta/\delta \rfloor - \beta| \leq g(x_0)\delta = O(n^{-1/3})$ . Next, consider  $T_2$ . Similar to the argument involving  $T_{21}$  in the proof of Lemma 2.8.3, for a constant  $K'$  depending only on  $\mathcal{N}_0$ , we have: we have

$$|T_2| \leq K'\delta \sum_{j=1}^{\lfloor \beta/\delta \rfloor} |t_{l-M+j}^* - x_0| \leq K'\delta \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j\delta + K'\delta \sum_{j=1}^{\lfloor \beta/\delta \rfloor} (M+1)\delta.$$

Denote the above last two terms as  $S_1$  and  $S_2$ . We have  $S_1 \leq K'\delta^2\lfloor \beta/\delta \rfloor(1 + \lfloor \beta/\delta \rfloor) \leq K'\beta^2 + K'\beta\delta \leq K'\beta^2 + O(n^{-1/3})$  and  $S_2 \leq K'\delta^2\lfloor \beta/\delta \rfloor \leq K'\beta\delta \leq O(n^{-1/3})$ . Thus,  $T_2 \leq K'\beta^2 + O(n^{-1/3})$  and Fact 2 now follows.  $\square$

The proof of the next lemma follows the lines of the proof of Lemma 4.1 in [Kim and Pollard \(1990\)](#).

*Lemma 2.8.5.* Fact 3 in the proof of Lemma 2.4.22 holds.

*Proof.* Denote  $R_0 > 0$  as a small constant. Let, for each  $m > 0$ ,

$$\begin{aligned}M_n &= \sup_{|\beta| \leq R_0} n^{2/3} (|(\mathbb{P}_n - \mathbb{P})g_{n1}(x, y; \beta)| - \epsilon\beta^2), \\ E &= \{\exists \beta \in [\pm R_0] \text{ s.t. } |(\mathbb{P}_n - \mathbb{P})g_{n1}(x, y; \beta)| > \epsilon\beta^2 + n^{-2/3}m^2\}, \\ A_{n,j} &= \{|\beta| \leq R_0 : (j-1)n^{-1/3} \leq |\beta| \leq jn^{-1/3}\}, \\ B_j &= \{\exists \beta \in A_{n,j} \text{ s.t. } n^{2/3}|(\mathbb{P}_n - \mathbb{P})g_{n1}(x, y; \beta)| > \epsilon(j-1)^2 + m^2\}.\end{aligned}$$

Then, we have  $\mathbb{P}(M_n > m) \leq \mathbb{P}E \leq \sum_{j=1}^{\infty} \mathbb{P}B_j$ . By Chebyshev's Inequality,

$$\mathbb{P}B_j \leq n^{4/3} \mathbb{P} \sup_{|\beta| < jn^{-1/3}} |(\mathbb{P}_n - \mathbb{P})g_{n1}(x, y; \beta)|^2 / [\epsilon(j-1)^2 + m^2]^2.$$

Denote  $\mathcal{G}_{n1}(R) = \{g_{n1}(x, y; \beta) : |\beta| \leq R\}$  for  $0 < R \leq R_0$ . Then one envelope

function is  $F_{n_1}(x, y; R) = \{x \in [t_{l-M} \pm R]\}$  by noticing  $|y - F(t_{l-M})| \leq 1$ . Straight forward calculation gives  $\mathbb{P}F_{n_1}^2(R) \leq C_1 R$  for  $n$  large enough, where  $C_1$  can be any constant greater than  $2g(x_0)$ , say  $3g(x_0)$ . Note that, for every  $R \in (0, R_0]$ ,  $\mathcal{G}_{n_1}(R)$  is a bounded VC-class of functions with VC-dimension bounded by a constant (independent of  $n$ ). It follows readily that  $J(1, \mathcal{G}_{n_1}(R))$  is finite and uniformly bounded in  $n$  (see Page 239 of *van der Vaart and Wellner (1996)*). Theorem 2.14.1 of *van der Vaart and Wellner (1996)* with  $\mathcal{F}$  taken to be  $\mathcal{G}_{n_1}(jn^{-1/3})$  and  $p = 2$  now yields that for some constant  $C_2$  and  $C = C_1 C_2$

$$\mathbb{P} \sup_{|\beta| < jn^{-1/3}} |(\mathbb{P}_n - \mathbb{P}) g_{n_1}(x, y; \beta)|^2 \leq n^{-1} C_2 \mathbb{P}F_{n_1}^2(jn^{-1/3}) \leq C j n^{-4/3}.$$

Thus, we have  $\mathbb{P}(M_n > m) \leq \sum_{j=1}^{\infty} \mathbb{P}B_j \leq \sum_{j=1}^{\infty} Cj / [\epsilon(j-1)^2 + m^2]^2$ . The last term converges for each  $m > 0$  and goes to 0 as  $m \rightarrow \infty$  by the Dominated Convergence Theorem. Therefore  $M_n = O_P(1)$ , which completes the proof.  $\square$

*Lemma 2.8.6.* Fact 4 in the proof of Lemma 2.4.22 holds.

*Proof.* The proof is the same to that of Lemma 2.8.5 except changing subscripts  $n_1$  to  $n_2$ . Note that the same envelope function is used.  $\square$

*Lemma 2.8.7.* Claim 2 in the proof of Theorem 2.4.17 holds.

*Proof.* Without loss of generality, here we only prove for the first equality. Denote  $\mathbb{K}(i) = GCM\{\mathcal{P}_c(k), k \in \mathbb{Z}\}(ci)$  for  $i \in \mathbb{Z}$ . Let  $L$  and  $U$  be the largest integer less than  $-M$  and the smallest integer larger than  $M$  such that  $\mathbb{K}$  change slopes at  $cL$  and  $cU$ . Then, the first equality holds. Thus, it is sufficient to show both  $L$  and  $U$  are  $O_P(1)$ . We show the latter and the former can be shown in the same way.

Denote  $A_j = \{\mathbb{K}(M) + \mathbb{X}(M)(j - M) = \mathcal{P}_{2,c}(j)\}$  for  $j \geq M$ . Then,

$$\{U = +\infty\} \subset \{A_j, i.o. \text{ for } j \geq M\} =: B.$$

On the other hand, we have

$$B \subset \left\{ \limsup_{t \rightarrow +\infty} \frac{\alpha W(t) + \beta t(c+t)}{t} = +\infty \right\}^c =: D^c.$$

From  $\mathbb{P}(\lim_{t \rightarrow +\infty} W(t)/t = 0) = 1$ , the law of large number for a standard Brownian motion, we have  $\mathbb{P}(D) = 1$  by noticing  $\beta > 0$ . Therefore,  $\mathbb{P}(U = +\infty) = 0$ , which

completes the proof. □

*Lemma 2.8.8.* The convergence in probability (2.25) in the proof of Lemma 2.4.23 holds.

The proof is fairly straightforward and therefore skipped.

*Lemma 2.8.9.* The weak convergence (2.26) in the proof of Lemma 2.4.23 holds.

*Proof.* It suffices to show that for each integer  $C > 0$ ,

$$\{V_n^*(ck), k \in [-C, C]\} \xrightarrow{d} \{\mathcal{P}_{2,c}, k \in [-C, C]\}.$$

We have  $V_n^*(ck) = T_1(ck) + T_2(ck)$  for  $k \in [-C, C]$ , where

$$\begin{aligned} T_1(ck) &= g(x_0)^{-1}n^{2/3}(\mathbb{P}_n - \mathbb{P})(y - F(t_l)) (\{x \leq t_l + kcn^{-1/3}\} - \{x \leq t_l\}), \\ T_2(ck) &= g(x_0)^{-1}n^{2/3}\mathbb{P}(y - F(t_l)) (\{x \leq t_l + kcn^{-1/3}\} - \{x \leq t_l\}). \end{aligned}$$

Then it is sufficient to show

$$\begin{aligned} \{T_1(ck), k \in [-C, C]\} &\xrightarrow{d} \{\alpha W(ck), k \in [-C, C]\}, \\ \{T_2(ck), k \in [-C, C]\} &\rightarrow \{\beta c^2 k(1+k), k \in [-C, C]\}. \end{aligned}$$

It is easy to show that  $T_2(ck) \rightarrow \beta c^2 k(1+k)$  uniformly for  $k \in [\pm C]$ . We, therefore, consider  $T_1$ . Set  $\xi_{i,n} = (\xi_{i,n,-C}, \dots, \xi_{i,n,0}, \dots, \xi_{i,n,C})$  with  $\xi_{i,n,k} = g(x_0)^{-1}n^{-1/3}(Y_{i,n} - F(t_l)) (\{X_{i,n} \leq t_l + kcn^{-1/3}\} - \{X_{i,n} \leq t_l\})$  for each  $i \in [1, n]$  and  $k \in [\pm C]$ . Then  $\{T_1(ck), k \in [-C, C]\} = \sum_{i=1}^n (\xi_{i,n} - \mathbb{P}\xi_{i,n})$ . Note that  $\{\xi_{i,n}\}$  is a row independent triangular array and each element is a  $2C + 1$  dimensional vector. By the triangular version of the central limit theorem (see, for example, Proposition 2.27 of [van der Vaart \(1998\)](#)), it is sufficient to show that  $\xi_{i,n}$  has a finite variance for each  $i$ ,  $\sum_{i=1}^n \mathbb{P}\|\xi_{i,n}\|^2 \{ \|\xi_{i,n}\| > \epsilon \} \rightarrow 0$  for each  $\epsilon > 0$ , and  $\sum_{i=1}^n \text{Cov}(\xi_{i,n})$  converges to the covariance matrix of  $\{\alpha W(ck), k \in [\pm C]\}$ . Since  $\mathbb{P}\xi_{i,n,k}^2 \leq g(x_0)^{-2}n^{-2/3} \leq g(x_0)^{-2}$  by noting both  $Y_{i,n}$  and  $F(t_l)$  are bounded by 1, we know  $\xi_{i,n}$  has finite variance.



Further, we have

$$\begin{aligned} \sum_{i=1}^n \mathbb{P} \|\xi_{i,n}\|^2 \{ \|\xi_{i,n}\| > \epsilon \} &\leq \frac{1}{\epsilon^2} \sum_{i=1}^n \mathbb{P} \|\xi_{i,n}\|^4 \\ &= \frac{1}{\epsilon^2} \sum_{i=1}^n \mathbb{P} \left( \sum_{k=-C}^C \xi_{i,n,k}^2 \right)^2 \leq \frac{1}{\epsilon^2 g(x_0)^4} (2C+1)^2 n^{-1/3} \rightarrow 0. \end{aligned}$$

That  $\sum_{i=1}^n \text{Cov}(\xi_{i,n,k_1}, \xi_{i,n,k_2})$  converges to  $\text{Cov}(\alpha W(ck_1), \alpha W(ck_2))$  for  $k_1, k_2 \in [\pm C]$  follows by direct calculation. This completes the proof.  $\square$

*Remark 2.8.10.* In the proofs of Lemma 2.8.8 and Lemma 2.8.9 we exploit the fact that, for the boundary case with  $\gamma = 1/3$ , we only need to consider finite dimensional random vectors  $\{G_n^*(ck), k \in [\pm C]\}$  and  $\{V_n^*(ck), k \in [\pm C]\}$  for every integer  $C > 0$ . However, for the case with  $\gamma \in (1/3, 1)$ , the dimension goes to  $\infty$  as  $n$  goes to  $\infty$ . Then the usual Chebyshev's inequality and the central limit theorem for a triangular array of random vectors will not work and more powerful tools from empirical processes need to be used.

*Lemma 2.8.11.* In the proof of Theorem 2.4.20,  $S_n \rightsquigarrow S$  holds.

*Proof.* The same truncation technique illustrated in the proof of Theorem 2.4.17 is utilized here again. More specifically, We will show that, in the following Lemmas 2.8.12 and 2.8.13, the following two claims hold:

**Claim 1:** Both (nonnegative)  $L_n$  and  $U_n$  are  $O_P(1)$ .

**Claim 2:** Both (nonnegative)  $L$  and  $U$  are  $O_P(1)$ .

Then, for each small  $\epsilon > 0$ , there exists (integer)  $M_\epsilon$  large enough such that  $P(M_\epsilon > \max\{L_n, U_n, L, U\}) > 1 - \epsilon$ . Denote

$$S_n^{M_\epsilon} = \sum_{j \in [\pm M_\epsilon]} (\mathbb{X}_n^2(cj) - \mathbb{Y}_n^2(cj)), \quad S^{M_\epsilon} = \sum_{j \in [\pm M_\epsilon]} (\mathbb{X}^2(cj) - \mathbb{Y}^2(cj)).$$

Then, we will show the following facts:

**Fact 1:**  $\lim_{\epsilon \rightarrow 0} \overline{\lim}_{n \rightarrow \infty} \mathbb{P}(S_n^{M_\epsilon} \neq S_n) = 0$ .

**Fact 2:**  $\lim_{\epsilon \rightarrow 0} \mathbb{P}(S^{M_\epsilon} \neq S) = 0$ .

**Fact 3:**  $S_n^{M_\epsilon} \rightsquigarrow S^{M_\epsilon}$ , as  $n \rightarrow \infty$  for each  $\epsilon > 0$ .

Fact 1 and Fact 2 hold since that  $\{M_\epsilon > \max\{|L_n|, U_n, |L|, U\}\}$  is a subset of both  $\{S_n^{M_\epsilon} = S_n\}$  and  $\{S^{M_\epsilon} = S\}$ . By Theorem 2.4.17 and the Continuous Mapping Theorem, Fact 3 follows. Therefore, by Lemma 2.4.18, we have  $S_n \rightsquigarrow S$ .  $\square$

*Lemma 2.8.12.* Claim 1 in Lemma 2.8.11 holds.

*Proof.* Let  $-cL'_n$  be the smallest grid point less than 0 such that  $\mathbb{X}_n(-cL'_n)$  is equal to  $\mathbb{X}_n(0)$  and  $cR'_n$  the largest grid point larger than or equal to 0 such that  $\mathbb{X}_n(cR'_n)$  is equal to  $\mathbb{X}_n(0)$ . Let  $-cL''_n$  be the smallest grid point less than 0 such that  $\mathbb{X}_n(-cL''_n) > 0$  and  $cR''_n$  the largest grid point greater than 0 such that  $\mathbb{X}_n(cR''_n) < 0$ . If there does not exist such a grid point  $-cL''_n$  or  $cR''_n$ , let  $L''_n$  or  $R''_n$  be any nonnegative integer, say 0. Note that  $L'_n, R'_n, L''_n$  and  $R''_n$  are all nonnegative.

Then,  $\mathbb{X}_n$  and  $\mathbb{Y}_n$  differ at most over  $[-cL'_n - cL''_n, cR'_n + cR''_n]$ . From the relationship between  $\mathbb{X}_n$  and  $\hat{F}$  and that between  $\mathbb{Y}_n$  and  $\hat{F}^\circ$ , it is clear that  $[-L_n, R_n]$  is contained in  $[-L'_n - L''_n, R'_n + R''_n]$ . Thus, it suffices to show  $L'_n + L''_n = O_P(1)$  and  $R'_n + R''_n = O_P(1)$ .

From the proof of Lemma 2.4.22, we have already shown  $L'_n$  and  $R'_n$  are  $O_P(1)$ , just by letting  $M$  there be 0. Thus, it remains to show  $L''_n$  and  $R''_n$  are  $O_P(1)$ . We next show the former and the latter can be derived in the same way.

For each integer  $M > 0$ , by Theorem 2.4.17,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(L''_n > M) \leq \limsup_{n \rightarrow \infty} \mathbb{P}(\mathbb{X}_n(-cM) > 0) = \mathbb{P}(\mathbb{X}(-cM) > 0).$$

On the other hand, we will show  $\lim_{M \rightarrow \infty} \mathbb{P}(\mathbb{X}(-cM) > 0) = 0$  in the proof of the next Lemma 2.8.13. Therefore,  $L''_n$  is  $O_P(1)$ , which completes the proof.  $\square$

*Lemma 2.8.13.* Claim 2 in Lemma 2.8.11 holds.

*Proof.* Define  $L', R', L''$  and  $R''$  similar to those  $L'_n, R'_n, L''_n$  and  $R''_n$  in the proof of Lemma 2.8.12. More specifically, let  $-cL'$  be the smallest grid point less than 0 such that  $\mathbb{X}(-cL')$  is equal to  $\mathbb{X}(0)$  and  $cR'$  the largest grid point larger than or equal to 0 such that  $\mathbb{X}(-cR')$  is equal to  $\mathbb{X}(0)$ . Let  $-cL''$  be the smallest grid point less than 0 such that  $\mathbb{X}(-cL'') > 0$  and  $cR''$  the largest grid point greater than 0 such that  $\mathbb{X}(cR'') < 0$ . If there does not exist such a grid point  $-cL''$  or  $cR''$ , let  $L''$  or  $R''$  be any nonnegative integer, say 0. Note that  $L', R', L''$  and  $R''$  are nonnegative.

Then,  $\mathbb{X}$  and  $\mathbb{Y}$  differ at most over  $[-cL' - cL'', cR' + cR'']$ . Since that  $[-L, R]$  is in  $[-L' - L'', R' + R'']$ , it suffices to show  $L' + L'' = O_P(1)$  and  $R' + R'' = O_P(1)$ .

From the proof of Lemma 2.8.7, we have already known  $L'$  and  $R'$  are  $O_P(1)$ , just by letting  $M$  there be 0. Thus, it remains to show  $L''$  and  $R''$  are  $O_P(1)$ . We next show the former, and the latter can be done in the same way.

For each integer  $M > 0$ , we have

$$\mathbb{P}\left(L'' > M\right) \leq \mathbb{P}\left(\bigcup_{m=M}^{\infty} \left\{\frac{\mathcal{P}_{2,c}(-m)}{-cm} > 0\right\}\right).$$

On the other hand, we have

$$\mathbb{P}\left\{\limsup_{t \rightarrow -\infty} \frac{\alpha W(t) + \beta t(c+t)}{t} = -\infty\right\} = 1,$$

which gives

$$\mathbb{P}\left\{\limsup_{t \rightarrow -\infty} \frac{\alpha W(t) + \beta t(c+t)}{t} > 0\right\} = 0.$$

Thus, we have

$$\lim_{M \rightarrow \infty} \mathbb{P}\left(\bigcup_{m=M}^{\infty} \left\{\frac{\mathcal{P}_{2,c}(-m)}{-cm} > 0\right\}\right) = \mathbb{P}\left(\limsup_{m \rightarrow \infty} \frac{\mathcal{P}_{2,c}(-m)}{-cm} > 0\right) = 0,$$

Therefore,  $L''$  is  $O_P(1)$ . □

*Lemma 2.8.14.* In the proof of Theorem 2.4.20, the remainder terms  $R_1, R_2$  and  $R_3$  are all  $o_P(1)$  holds.

*Proof.* We will only show  $R_1 = o_P(1)$  and  $R_2 = o_P(1)$  can be established in the same way. Recall that

$$\begin{aligned} R_1 &= -\frac{1}{2} \left( \frac{1}{F_l^2} - \frac{1}{(1-F_l)^2} \right) \sum_{i \in J_n} \left( \hat{F}(X_i) - F_l \right)^3 \\ &\quad + \frac{1}{3} \sum_{i \in J_n} \left( \frac{Y_i}{F_{li}^{*3}} - \frac{1-Y_i}{(1-F_{li}^{**})^3} \right) \left( \hat{F}(X_i) - F_l \right)^3. \end{aligned}$$

First, we have

$$n^{1/3} \sum_{i \in J_n} \left| \hat{F}(X_i) - F_l \right|^3 = \sum_{j \in [L_n, U_n]} |\mathbb{X}_n(cj)|^3 W_j \rightsquigarrow c \sum_{j \in [L, U]} |\mathbb{X}(cj)|^3,$$

where the weak convergence holds by the same argument as in Lemma 2.8.11.

Second, it is easy to see that there exists  $C > 0$  such that  $1/F_l^2 + 1/(1 - F_l)^2 < C$  for large  $n$ . For example, take  $C = 2/F(x_0)^2 + 2/(1 - F(x_0))^2$ .

Third, recall that  $F_{l_i}^*$  and  $F_{l_i}^{**}$  lie between  $F_l$  and  $\hat{F}(X_i)$  for  $i \in J_n$ . Note that  $X_i$  with  $i \in J_n$  can only be  $t_j$  for some  $j \in D_n$ . That is, for each  $i \in J_n$ , there exists  $j \in D_n$  such that  $\hat{F}(X_i) = \hat{F}(t_j)$ . On the other hand, note  $L_n$  and  $U_n$  are  $O_P(1)$ . Then, with arbitrary high probability  $D_n \subset [\pm M]$  for  $M$  large enough. Since  $\hat{F}(t_j)$  converges to  $F(x_0)$  in probability for each fixed  $j$  and  $F_l$  always converges to  $F(x_0)$ , both  $F_{l_i}^*$  and  $F_{l_i}^{**}$  converge to  $F(x_0)$  in probability for  $i \in [\pm M]$  with a fixed  $M$ . Thus, we have  $\sup_{i \in J_n} (1/F_{l_i}^{*3} + 1/(1 - F_{l_i}^{**})^3)$  converges to  $1/F(x_0)^3 + 1/(1 - F(x_0))^3$  in probability.

Therefore, using the triangular inequality for absolute value, combining the above three results and using Slutsky's Lemma gives  $R_1 = o_P(1)$ .

Next, consider  $R_3$ . We have:  $R_3 = g(x_0)(F_l(1 - F_l))^{-1}T_n$ , where

$$T_n = \sum_{j \in [-L_n, U_n]} (\mathbb{X}_n^2(cj) - \mathbb{Y}_n^2(cj)) [(G_n^*(c(j)) - G_n^*(c(j-1))) - c].$$

Then  $|T_n| \leq S_n \xi_n$  by noticing that  $\mathbb{X}_n^2(cj) \geq \mathbb{Y}_n^2(cj)$  for each  $j$ , where

$$\xi_n = \sup_{j \in [-L_n, U_n]} |(G_n^*(c(j)) - G_n^*(c(j-1))) - c|.$$

By Lemma 2.8.11,  $S_n$  converges weakly. Thus, it suffices to show the nonnegative quantity  $\xi_n = o_P(1)$ . Denote  $D_n = \{[-L_n, U_n] \subset [\pm C]\}$  for  $C > 0$ . By Lemma 2.8.12, the probability of  $D_n$  can be made arbitrarily close to 1 as  $C$  becomes large enough. Then, it suffices to show  $\xi_n D_n = o_P(1)$ . We have

$$\begin{aligned} \xi_n D_n &\leq \sup_{j \in [\pm C]} |G_n^*(cj) - cj| + \sup_{j \in [\pm C]} |G_n^*(c(j-1)) - c(j-1)| \\ &\leq 2 \sup_{j \in [\pm(C+1)]} |G_n^*(cj) - cj|. \end{aligned}$$

The last term converges to 0 in probability by Lemma 2.8.8, which completes the proof.  $\square$

## CHAPTER III

# A Two-Stage Hybrid Procedure for Estimating an Inverse Regression Function

### 3.1 Introduction

The problem of estimating an *inverse* regression function has a long history in Statistics, due to its importance in diverse areas including toxicology, drug development and engineering. The canonical formulation of the problem is as follows. Let

$$Y = f(x) + \epsilon,$$

where  $f$  is a *monotone* function establishing the relationship between the design variable  $x$  and the response  $Y$ , and  $\epsilon$  an error term with zero mean and finite variance  $\sigma^2$ . Further, without loss of generality it is assumed that  $f$  is isotonic and  $x \in [0, 1]$ . It is of interest to estimate  $d_0 = f^{-1}(\theta_0)$  for some  $\theta_0$  in the interior of the range of  $f$ , given  $f'(d_0) > 0$ .

Depending on the nature of the problem, one usually first obtains an estimate of  $f$  and subsequently of  $d_0$ , either from observational data or from design studies [Morgan \(1992\)](#). In the latter case, one specifies a number of values for the design variable, and obtains the corresponding responses, which are then used to get the estimates.

Motivated by an engineering application, fully described in Section [3.5](#), we introduce a two-stage design for estimating  $d_0$ . Specifically, we consider a complex queueing system operating in discrete time under a throughput (average number of customers processed per unit of time) maximizing resource allocation policy (for details see [Bambos and Michailidis \(2004\)](#)). Unfortunately the customers' average delay, which is an important "quality-of-service" metric of the performance of the system,

is not analytically tractable and can only be obtained via expensive simulations. The average delay as a function of the system's loading (number of customers arriving per unit of time) is depicted in Figure 3.1. The relationship between system loading and average delay can not be easily captured by a simple parametric model; hence, a nonparametric estimator might be more useful. In addition, given that the responses are obtained through simulation, only a relatively small number of simulation runs can be performed. It is of great interest for the system's operator to obtain accurate estimates of the loading corresponding to prespecified delay thresholds (e.g. 10 and 15 time units), so as to be able to decide whether to upgrade the available resources.

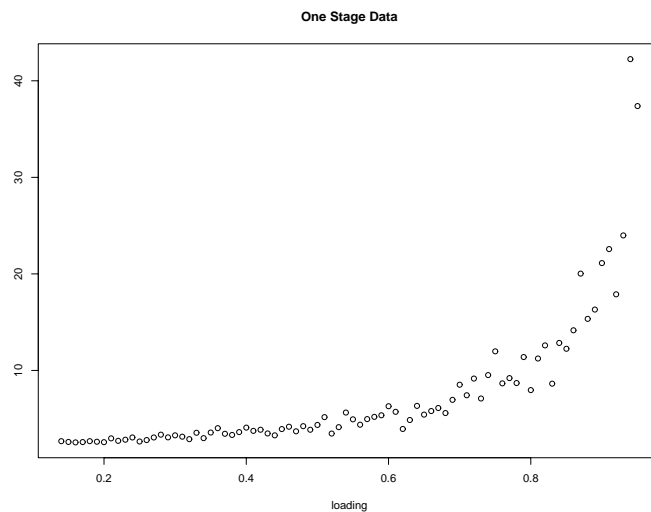


Figure 3.1: The average delay as a function of system's loading.

The main idea of the proposed two-stage approach is summarized next: at stage one, an initial set of design points and their corresponding responses are generated. Then a first-stage nonparametric estimate of  $f$  is obtained and subsequently a first-stage estimate of  $d_0$ . Next, a second-stage sampling interval covering  $d_0$  with high probability is specified and all new design points are laid down at the two boundary points of this interval and their responses obtained. Finally, a linear regression model is fitted to the second-stage data by least squares and a second-stage estimate of  $d_0$  computed as the inverse of the locally approximating line of  $f$  at  $\theta_0$ . As we will see, the employment of a local linear approximation at stage two allows the second-stage estimate of  $d_0$  to attain a  $\sqrt{n}$  parametric rate of convergence, despite the nonparametric nature of the problem. To overcome estimation of several tuning parameters

required by the second-stage estimate, a bootstrapped variant is introduced and its consistency properties established. To clinch the asymptotic results of the proposed two-stage estimate and its bootstrapped counterpart, a number of subtle technical issues need to be addressed and these are resolved in subsequent sections. Before proceeding further, it is important to draw attention to the fact that our proposed two-stage method relies critically on the *reproducibility* of the experiment: i.e. at any stage, it is possible to sample responses from any pre-specified covariate value. While reproducibility in this sense is guaranteed for our motivating application, the two-stage procedure above is not applicable in the absence of adequate degree of control on the covariate. For example, if the covariate is time, the implementation of a two-stage procedure would require one to go back and sample from the past, which is impossible.

Isotonic regression is a conceptually natural and computationally efficient estimation method for shape-restricted problems [Barlow et al. \(1972\)](#); [Robertson et al. \(1988\)](#). In the framework of regression, the asymptotic distribution for the isotonic regression estimator at a fixed point was first derived in [Brunk \(1970\)](#), and then extended in [Wright \(1981\)](#) and [Leurgans \(1982\)](#). The asymptotic distribution for the  $\mathbb{L}_1$ -distance between the isotonic estimator and the regression function was obtained in [Durot \(2002\)](#), paralleling [Groeneboom et al. \(1999\)](#) on a unimodal density, and then extended in [Durot \(2007, 2008\)](#). [Banerjee and Wellner \(2005\)](#) derive the asymptotic distribution for the inverse of the distribution function of the survival time at a given point in the current status model; the regression version of this result will be used to derive the asymptotics for the two stage procedures.

The inverse regression problem has been extensively studied in the context of different applications. For example, in statistical calibration, the goal is to estimate a scalar quantity  $d_0$  from a model  $Z = f(d_0) + \epsilon$ , with  $Z$  observed. The information about the underlying function  $f$  comes from experimental data  $\{Y_i, X_i\}$  that follow the same regression model; namely,  $Y_i = f(X_i) + \epsilon_i$ . [Osborne \(1991\)](#) gives a comprehensive review of this topic and [Gruet \(1996\)](#) provides a kernel based direct nonparametric estimator of  $d_0$ . It is clear that when  $\epsilon = 0$ , the calibration problem becomes the canonical problem described above.

Another active area is provided by the model-based dose-finding problems in toxicology and drug development, where  $d_0$  corresponds to either the maximal tolerated dose or the effective dose with respect to a given maximal toxicity or an efficacy level.

Possible dose levels are often prespecified. The dose-response relationship is usually assumed to be monotone and described either by parametric models (e.g. probit, logit [Morgan \(1992\)](#), multihit [Rai and Ryzin \(1981\)](#), cubic logistic [Morgan \(1985\)](#)), or by nonparametric models, for which kernel estimates [Staniswalis and Cooper \(1988\)](#) and isotonic regression [Stylianou and Flournoy \(2002\)](#) are employed. On the other hand, due to ethical and budget considerations, most studies encompass sequential designs, so that more subjects (e.g. patients) receive doses close to the target  $d_0$  (see [Rosenberger \(1996\)](#) and [Rosenberger and Haines \(2002\)](#) for comprehensive reviews on the subject). [Stylianou and Flournoy \(2002\)](#) compare parametric estimators using maximum likelihood and weighted least squares based on the logit model and nonparametric ones using sample mean and isotonic regression with a sequential up-and-down biased coin design, and show that a linearly interpolated isotonic regression estimator performs best in most simulated scenarios. Further, [Ivanova et al. \(2003\)](#) claim that the isotonic regression estimator still performs best for small to moderate sample sizes with several sequential designs from a family of up-and-down designs; [Gezmu and Flournoy \(2006\)](#) recommend using smoothed isotonic regression with their group up-and down designs. All these partially motivate the use of isotonic regression in our two-stage procedure, though it should be noted that our approach is markedly different from the ones discussed above, owing to the different nature of the motivating application; in particular, ethical constraints that prevent administration of high dose-levels are absent in our situation.

In a nonparametric setting, one could also employ a fully sequential Robbins-Monro procedure [Robbins and Monro \(1951\)](#) for finding  $d_0$ . This corresponds to a stochastic version of Newton’s scheme for root finding problems. [Anbar \(1977\)](#) considered a modified Robbins-Monro type procedure approximating the root from one side. A good review of this area is provided in [Lai \(2003\)](#), in which it is also pointed out that the procedure usually exhibits an “unsatisfactory finite-sample performance except for linear problems” especially in noisy settings, due to the fact that it does not incorporate modeling for (re)using all the available –up to that instance– data. Another downside of a sequential design, as opposed to the *batch* design employed in this study, is the time and effort required to collect the data points [Müller and Schmitt \(1990\)](#).

The remainder of the paper is organized as follows: Section [3.2](#) describes the two-stage procedures. The asymptotic properties of the two-stage estimators are obtained



in Section 3.3. Simulation studies and data analysis are presented in Sections 3.4 and 3.5, respectively. We close with a discussion in Section 3.6, which is followed by an appendix containing technical details.

## 3.2 Two-Stage Procedures

In this Section, we review some necessary background material and introduce the proposed two-stage estimation procedures.

### 3.2.1 Preliminaries: A Single-Stage Procedure

We review some material on estimating the parameter of interest  $d_0$  by using isotonic regression combined with a single-stage design. The procedure is outlined next:

1. Choose  $n$  increasing design points  $\{x_{in}\}_{i=1}^n \in [0, 1]$  and obtain the corresponding responses that are generated according to  $Y_{in} = f(x_{in}) + \epsilon_{in}$ ,  $i = 1, 2, \dots, n$ , where  $f$  is in  $\mathcal{F}_0$ , a class of increasing real functions on  $[0, 1]$  with positive and continuous first derivatives in a neighborhood of  $d_0$  and  $\epsilon_{in}$  are independently and identically distributed (iid) random errors with mean zero and constant variance  $\sigma^2$ . Note that the subscript  $n$  will be suppressed from now on for simplicity of notation.
2. Obtain the isotonic regression estimate  $\hat{f}$  of  $f$  from the data  $\{(x_i, Y_i)\}_{i=1}^n$ . (For details see Chapter 1 of [Robertson et al. \(1988\)](#)).
3. Estimate  $d_0$  by  $\hat{d}_n^{(1)} = \hat{f}^{-1}(\theta_0) = \inf\{x \in [0, 1] : \hat{f}(x) \geq \theta_0\}$ , where  $\theta_0 = f(d_0)$ .

In order to study the properties of  $\hat{f}$  and  $\hat{d}_n^{(1)}$ , we consider the following further assumption on the design points.

**(A1)** There exists a distribution function  $G$ , whose Lebesgue density  $g$  is positive at  $d_0$ , such that  $\sup_{x \in [0, 1]} |F_n(x) - G(x)| = o(n^{-1/3})$ , where  $F_n$  is the empirical function of  $\{x_i\}_{i=1}^n$ .

For example, the discrete uniform design  $x_i = i/n$  for  $i = 1, 2, \dots, n$  satisfies (A1) with  $G$  being the uniform distribution on  $[0, 1]$  and  $g(d_0) = 1 > 0$ .

The following basic result provides the asymptotic distribution of  $\hat{d}_n^{(1)}$ .

*Theorem 3.2.1.* If  $f \in \mathcal{F}_0$  and (A1) holds,

$$n^{1/3}(\hat{d}_n^{(1)} - d_0) \xrightarrow{d} CZ,$$

where  $C = [4\sigma^2/(f'(d_0)^2g(d_0))]^{1/3}$  and  $Z$  follows Chernoff's distribution.

*Remark 3.2.2.* Chernoff's distribution is the distribution of the almost sure unique maximizer of  $B(t) - t^2$  on  $\mathbb{R}$ , where  $B(t)$  denotes a two-sided standard Brownian motion starting at the origin ( $B(0) = 0$ ). It is symmetric around zero, with tails dwindling faster than those of the Gaussian and its quantiles have been tabled in [Groeneboom and Wellner \(2001\)](#).

The proof of [Theorem 3.2.1](#) follows by adaptations of the arguments from [Theorem 1](#) in [Banerjee and Wellner \(2005\)](#) to the current regression setting. Hence, an approximate confidence interval for  $d_0$  with significance level  $1 - 2\alpha$  can be constructed as follows

$$[\hat{d}_n^{(1)} - n^{-1/3}\hat{C}q_\alpha, \hat{d}_n^{(1)} + n^{-1/3}\hat{C}q_\alpha] \cap (0, 1), \quad (3.1)$$

where  $q_\alpha$  denotes the upper  $\alpha$  quantile of Chernoff's distribution and  $\hat{C}$  is a consistent estimate of  $C$ .

In the presence of relatively small budgets for design points, the slow convergence rate and the need to estimate  $f'(d_0)$  adversely impact the performance of this procedure. In order to accelerate the convergence rate, we propose next an alternative that is based on a two-stage sampling design and uses local linear approximation for  $f$  in stage two.

### 3.2.2 Procedures Based On Two-Stage Sampling Designs

We describe next a hybrid estimation procedure for estimating  $d_0$  based on a two-stage sampling design. Suppose that the total budget consists of  $n$  doses that are going to be allocated in two stages.

1. Allocate  $n_1 = np$ ,  $p \in (0, 1)$  design points and obtain the first-stage data  $\{(x_i, Y_i)\}_{i=1}^{n_1}$ , the isotonic regression estimate of  $f$  and the estimate  $\hat{d}_{n_1}^{(1)}$  of  $d_0$  as outlined in [Section 3.2.1](#). Note that by  $np$ , we denote by  $\lfloor np \rfloor$  or  $\lfloor np \rfloor + 1$ , depending on whether  $n - \lfloor np \rfloor$  is even or not. Also, recall that the additional subscript  $n$  is suppressed.
2. Determine two second-stage sampling points  $L$  and  $U$  symmetrically around

$\hat{d}_{n_1}^{(1)}$ , where  $L = \hat{d}_{n_1}^{(1)} - Kn_1^{-\gamma}$  and  $U = \hat{d}_{n_1}^{(1)} + Kn_1^{-\gamma}$ , for some constants  $\gamma > 0$  and  $K > 0$ .

3. Allocate the remaining  $n - n_1$  design points *equally* to  $L$  and  $U$  and generate the responses as  $Y_i' = f(L) + \epsilon_i'$  and  $Y_i'' = f(U) + \epsilon_i''$  for  $i = 1, 2, \dots, n_2$ , with  $\{\epsilon_i'\}$  and  $\{\epsilon_i''\}$  being iid random errors with mean zero and constant variance  $\sigma^2$ , mutually independent and also independent of  $\{\epsilon_i\}$ .
4. Fit the second-stage data  $\{(L, Y_i'), (U, Y_i'')\}$  with the linear model  $y = \beta_0 + \beta_1 x$  using least squares. Denote the resulting intercept and slope estimates by  $(\hat{\beta}_0, \hat{\beta}_1)$ , respectively. Then, the second-stage (or two-stage) estimator of  $d_0$  is given by  $\tilde{d}_n^{(2)} = (\theta_0 - \hat{\beta}_0)/\hat{\beta}_1$ .

Asymptotic properties of  $\tilde{d}_n^{(2)}$  will be established in Subsection 3.3.1. For example, when  $f$  is in a subset of  $\mathcal{F}_0$ , denoted as  $\mathcal{F}$ , the third derivatives of whose elements are uniformly bounded around  $d_0$ , and  $\gamma \in (1/4, 1/3)$ , we have

$$n^{1/2}(\tilde{d}_n^{(2)} - d_0) \xrightarrow{d} \frac{\sigma}{f'(d_0)(1-p)^{1/2}} N(0, 1), \quad (3.2)$$

where  $\xrightarrow{d}$  denotes convergence in distribution. Thus, the convergence rate of the two-stage estimator of  $d_0$  becomes  $n^{1/2}$ , the standard parametric convergence rate, which is faster than the  $n^{1/3}$  convergence rate of the one-stage isotonic regression estimator.

However, when constructing confidence intervals from asymptotic results like (3.2), we face two difficulties. One is that the limiting distributions of interest still depend on  $f'(d_0)$ , accurate estimation of which is difficult for small to moderate sample sizes. The other one, which is less obvious but perhaps with more serious practical implications, is that the asymptotic results of interest suffer slow speed of convergence in distribution. Therefore, a bootstrap variant of the two-stage procedure that avoids direct estimation of  $f'(d_0)$  is introduced and is seen to relieve the slow convergence problem.

### 3.2.3 Bootstrapping The Two-Stage Estimator

The steps of the bootstrapped two-stage procedure are outlined next.

1. Follow steps 1–4 to obtain the second stage design points  $L$  and  $U$ , responses  $\{Y_i'\}$  and  $\{Y_i''\}$  and  $\tilde{d}_n^{(2)}$ .

2. Sample with replacement, responses  $\{Y_i'^*\}_{i=1}^{n_2}$  and  $\{Y_i''^*\}_{i=1}^{n_2}$ , from  $\{Y_i'\}_{i=1}^{n_2}$  and  $\{Y_i''\}_{i=1}^{n_2}$ , respectively. Construct the corresponding bootstrapped second-stage (or two-stage) estimator  $\tilde{d}_n^{(2)*}$ , and calculate the root  $R_n^* = n^{1/2}(\tilde{d}_n^{(2)*} - \tilde{d}_n^{(2)})$ .
3. Repeat the previous step  $B$  times to obtain  $\{R_n^{*b}\}_{b=1}^B$ . Subsequently, calculate the lower and upper  $\alpha$  quantiles,  $q_l^*$  and  $q_u^*$ , of  $\{R_n^{*b}\}_{b=1}^B$ . Finally, construct a  $1 - 2\alpha$  bootstrapped Wald-type confidence interval for  $d_0$  as

$$[\tilde{d}_n^{(2)} - n^{-1/2}q_u^*, \tilde{d}_n^{(2)} - n^{-1/2}q_l^*]. \quad (3.3)$$

Note that the procedure does not require estimation of  $f'(d_0)$ .

The asymptotic properties of the bootstrapped two-stage estimator are established in Subsection 3.3.2. For example, when  $f \in \mathcal{F}$ ,  $\gamma \in (0, 1/3)$  and all the absolute moments of the random error are finite, we have

$$n^{1/2}(\tilde{d}_n^{(2)*} - \tilde{d}_n^{(2)}) \xrightarrow{d^*} \frac{\sigma}{f'(d_0)(1-p)^{1/2}} N(0, 1), \quad (P - a.s.), \quad (3.4)$$

where  $\xrightarrow{d^*}$  implies convergence in distribution conditional on the data obtained from the employed two-stage design.

From (3.2) and (3.4), the strong consistency of the bootstrapped estimator  $\tilde{d}_n^{(2)*}$  is ensured for  $f \in \mathcal{F}$  and  $\gamma \in (1/4, 1/3)$ . In fact, the strong assumption on the random error can be replaced by a mild one that the sixth moment of the random error is finite, at the price of replacing strong consistency with weak consistency. Therefore, the bootstrapped procedure is theoretically validated under certain conditions.

*Remark 3.2.3.* Both the two-stage estimator and its bootstrapped variant rely on the choice of a number of tuning parameters:  $p$ ,  $\gamma$  and  $K$ . Practical procedures for their selection will be discussed in Section 3.4.

### 3.3 Asymptotic Properties of Two-Stage Estimators

In this Section, we establish the asymptotic properties of both the two-stage estimator and its bootstrapped variant for  $d_0$ . We start by discussing the two-stage estimator  $\tilde{d}_n^{(2)}$ .

### 3.3.1 Two-Stage Estimator

All results in this subsection are derived under the assumption (A1). According to the two-stage procedure,

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1 \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n_2} [(Y_i' - \beta_0 - \beta_1 L)^2 + (Y_i'' - \beta_0 - \beta_1 U)^2].$$

Denote  $Y_i^+ = Y_i'' + Y_i'$  and  $Y_i^- = Y_i'' - Y_i'$ . Then,

$$\hat{\beta}_0 = (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^+ - \hat{d}_{n_1}^{(1)} \hat{\beta}_1, \quad \hat{\beta}_1 = (2Kn_1^{-\gamma}n_2)^{-1} \sum_{i=1}^{n_2} Y_i^-. \quad (3.5)$$

Setting  $\theta_0 = \hat{\beta}_0 + \hat{\beta}_1 \tilde{d}_n^{(2)}$  gives

$$\tilde{d}_n^{(2)} = (1/\hat{\beta}_1)(\theta_0 - \hat{\beta}_0) = (1/\hat{\beta}_1)[\theta_0 - (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^+] + \hat{d}_{n_1}^{(1)}. \quad (3.6)$$

In order to analyze  $\tilde{d}_n^{(2)}$ , additional assumptions about the smoothness of the underlying function  $f$  around  $d_0$  are required. We consider the following three classes of underlying functions:

$$\begin{aligned} \mathcal{F} &= \{f \in \mathcal{F}_0 : f'''(x) \text{ is } UBN(d_0)\}, \\ \mathcal{F}_1 &= \{f \in \mathcal{F}_0 : f''(d_0) \neq 0, f'''(x) \text{ is } UBN(d_0)\}, \\ \mathcal{F}_2 &= \{f \in \mathcal{F}_0 : f''(d_0) = 0, f'''(d_0) \neq 0, f^{(4)}(x) \text{ is } UBN(d_0)\} \end{aligned}$$

where  $UBN(d_0)$  means “uniformly bounded in a neighborhood of  $d_0$ ”. Then, the mutually exclusive  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are subsets of  $\mathcal{F}$ .

*Remark 3.3.1.* A function in  $\mathcal{F}_2$  is exactly locally linear at  $d_0$  while that in  $\mathcal{F}_1$  is not. Notice that both  $\mathcal{F}_2$  and  $\mathcal{F}_1$  depend on  $d_0$ . For example, consider the sigmoid function  $f(x) = \exp\{a(x - b)\}/(1 + \exp\{a(x - b)\})$  for some constants  $a > 0$  and  $b \in (0, 1)$ . It belongs to  $\mathcal{F}_2$  if  $d_0 = b$  and to  $\mathcal{F}_1$  otherwise. Obviously, the size of  $\mathcal{F}_2$  is much smaller than that of  $\mathcal{F}_1$ . However, the asymptotic results for  $f \in \mathcal{F}_2$  should also provide good approximations for functions that are approximately linear in the vicinity of  $d_0$ . Hence, the class  $\mathcal{F}_2$  is also of interest.

We consider next the asymptotic properties of  $\tilde{d}_n^{(2)}$ , starting with the consistency of the two-stage estimator.

*Lemma 3.3.2.* For  $f \in \mathcal{F}$  and  $\gamma \in (0, 1/2)$ , we have:

$$\hat{\beta}_0 \xrightarrow{P} f(d_0) - f'(d_0)d_0, \quad \hat{\beta}_1 \xrightarrow{P} f'(d_0), \quad \text{and} \quad \tilde{d}_n^{(2)} \xrightarrow{P} d_0.$$

Based on Lemma 3.3.2, we obtain the asymptotic distribution of  $\tilde{d}_n^{(2)}$  in the next theorem. It turns out that the asymptotic results with  $f \in \mathcal{F}_1$  and  $\mathcal{F}_2$  are the same for  $\gamma > 1/4$ . This implies that the nonlinearity of  $f$  at  $d_0$  becomes asymptotically ignorable as the length of the neighborhood of  $d_0$  shrinks fast enough.

*Theorem 3.3.3.* For  $f \in \mathcal{F}$  and  $\gamma \in (1/4, 1/2)$ ,

$$\begin{aligned} n^{1/2}(\tilde{d}_n^{(2)} - d_0) &\xrightarrow{d} C_2 Z_1, \quad \text{for } \gamma \in (1/4, 1/3), \\ n^{1/2}(\tilde{d}_n^{(2)} - d_0) &\xrightarrow{d} C_2 Z_1 + C_3 \mathbb{Z} Z_2, \quad \text{for } \gamma = 1/3, \\ n^{(5/6-\gamma)}(\tilde{d}_n^{(2)} - d_0) &\xrightarrow{d} C_3 \mathbb{Z} Z_2, \quad \text{for } \gamma \in (1/3, 1/2); \end{aligned}$$

for  $f \in \mathcal{F}_1$  and  $\gamma \in (0, 1/4]$ ,

$$\begin{aligned} n^{2\gamma}(\tilde{d}_n^{(2)} - d_0) &\xrightarrow{d} C_1, \quad \text{for } \gamma \in (0, 1/4), \\ n^{1/2}(\tilde{d}_n^{(2)} - d_0) &\xrightarrow{d} C_1 + C_2 Z_1, \quad \text{for } \gamma = 1/4; \end{aligned}$$

for  $f \in \mathcal{F}_2$  and  $\gamma \in (1/8, 1/4]$ ,

$$n^{1/2}(\tilde{d}_n^{(2)} - d_0) \xrightarrow{d} C_2 Z_1, \quad \text{for } \gamma \in (1/8, 1/4];$$

where  $C_1 = -K^2 p^{-2\gamma} f''(d_0) / [2f'(d_0)]$ ,  $C_2 = \sigma / [f'(d_0)(1-p)^{1/2}]$ ,  $C_3 = CC_2/K$ ,  $C$  is as given in Theorem 3.2.1,  $Z_1$  and  $Z_2$  are standard normal,  $\mathbb{Z}$  follows Chernoff's distribution and  $\mathbb{Z}, Z_1, Z_2$  are mutually independent.

*Remark 3.3.4.* Theorem 3.3.3 characterizes the convergence rate of the estimator in terms of the size of the shrinking neighborhood. It shows that for  $\gamma \in [1/4, 1/3]$  the parametric rate of  $n^{1/2}$  is achieved given  $f \in \mathcal{F}$ . On the other hand, for the boundary values of  $\gamma = 1/4$  and  $1/3$ , there exists asymptotic bias in the former case (for  $f \in \mathcal{F}_1$ ), while in the latter case the asymptotic variance increases. For  $\gamma > 1/3$ , the rate of convergence falls below  $\sqrt{n}$ , while for  $\gamma < 1/4$  and  $f \in \mathcal{F}_1$  the limit distribution of the two-stage estimate is degenerate and thus not conducive to inference. Hence, these results suggest selecting  $\gamma$  in the  $(1/4, 1/3)$  range. Note that, the function class  $\mathcal{F}_2$  achieves the  $n^{1/2}$  rate of convergence for a slightly larger range

of values for  $\gamma$  than  $\mathcal{F}_1$ . This is a consequence of the near linearity of  $f$  in the vicinity of  $d_0$ , which allows a good linear approximation of  $f$  with a relatively long interval  $[L, U]$ .

*Remark 3.3.5.* The case of  $\gamma < 1/8$  has been omitted for  $f \in \mathcal{F}_2$ , since it involves a Taylor expansion of  $f$  up to its fifth derivative. Nevertheless, in principle no other technical challenges are in play.

### 3.3.2 Bootstrapped Two-Stage Estimator

We consider next the asymptotic properties of the bootstrapped two-stage estimator, which is:

$$\tilde{d}_n^{(2)\star} = (1/\hat{\beta}_1^\star)(\theta_0 - \hat{\beta}_0^\star) = (1/\hat{\beta}_1^\star)[f(d_0) - (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^{\star+}] + \hat{d}_{n_1}^{(1)}, \quad (3.7)$$

where  $Y_i^{\star+} = Y_i^{\prime\prime\star} + Y_i^{\prime\star}$ ,  $Y_i^{\star-} = Y_i^{\prime\prime\star} - Y_i^{\prime\star}$  and

$$\hat{\beta}_0^\star = (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^{\star+} - \hat{d}_{n_1}^{(1)} \hat{\beta}_1^\star, \quad \hat{\beta}_1^\star = (2Kn_1^{-\gamma}n_2)^{-1} \sum_{i=1}^{n_2} Y_i^{\star-}. \quad (3.8)$$

We now present a probabilistic framework needed to clearly establish the asymptotic properties of the bootstrapped estimator rigorously. The point is that the design points and random errors involved in the sampling mechanism are assumed to come from triangular arrays but not necessarily from sequences.

Let  $\{\{x_{in}\}_{i=1}^n\}_{n=1}^\infty$  be a triangular array of distinct design points in  $[0, 1]$  and  $\epsilon$  a continuous random variable in  $\mathbb{R}$  with mean 0 and finite variance  $\sigma^2 > 0$ . Now, there exists, on some probability space  $(\Omega, \mathcal{A}, P)$ , a set of random errors

$\{\{\epsilon_{in}\}_{i=1}^n, \{\epsilon'_{in}\}_{i=1}^n, \{\epsilon''_{in}\}_{i=1}^n\}_{n=1}^\infty$  which are iid copies of  $\epsilon$ . Then, suppressing the subscript  $n$ ,  $\{\{x_i\}_{i=1}^n, \{\epsilon_i(\omega)\}_{i=1}^{n_1}, \{\epsilon'_i(\omega)\}_{i=1}^{n_2}, \{\epsilon''_i(\omega)\}_{i=1}^{n_2}\}_{n=1}^\infty$  represents a fixed triangular array of real numbers for a fixed  $\omega \in \Omega$ , where  $n = n_1 + 2n_2$  with  $n_1$  and  $2n_2$  denoting the first and second stage sample sizes.

Given  $\omega \in \Omega$ , according to the sampling mechanism used in the bootstrapped procedure, the data obtained from the first stage are given by  $\{(x_i, Y_i(\omega))\}_{i=1}^{n_1}$ , which are subsequently used to obtain  $\hat{d}_{n_1}^{(1)}(\omega)$  and the lower and upper boundary points  $L(\omega)$  and  $U(\omega)$  to be used in the second stage. Hence, the second-stage data are given by  $\{L(\omega), Y'_i(\omega)\}$  and  $\{U(\omega), Y''_i(\omega)\}$  and the resulting estimate by  $\tilde{d}_n^{(2)}(\omega)$ . The procedure then requires bootstrapping  $\{Y'_i(\omega)\}_{i=1}^{n_2}$  and  $\{Y''_i(\omega)\}_{i=1}^{n_2}$ , which is conceptually

equivalent to bootstrapping  $\{\epsilon'_i(\omega)\}_{i=1}^{n_2}$  and  $\{\epsilon''_i(\omega)\}_{i=1}^{n_2}$  to get  $\{\epsilon_i'^*\}_{i=1}^{n_2}$  and  $\{\epsilon_i''^*\}_{i=1}^{n_2}$ , so that  $Y_i'^* = f(L(\omega)) + \epsilon_i'^*$  and  $Y_i''^* = f(U(\omega)) + \epsilon_i''^*$  for  $i = 1, 2, \dots, n_2$ . Note that given  $\omega$  and  $n$ , the bootstrapped second-stage random errors  $\{\epsilon_i'^*\}_{i=1}^{n_2}$  and  $\{\epsilon_i''^*\}_{i=1}^{n_2}$  are iid uniform random variables supported on  $\{\epsilon'_i(\omega)\}_{i=1}^{n_2}$  and  $\{\epsilon''_i(\omega)\}_{i=1}^{n_2}$ , respectively. Finally, the bootstrapped estimate  $\tilde{d}_n^{(2)*}$  is calculated from  $\{(L(\omega), Y_i'^*), (U(\omega), Y_i''^*)\}_{i=1}^{n_2}$ .

Thus, given  $\omega$  and with  $n$  increasing, the design points and random errors are sampled as rows from the fixed triangular array. Then, the bootstrapped random errors  $\{\epsilon_i'^*\}_{i=1}^{n_2}$  and  $\{\epsilon_i''^*\}_{i=1}^{n_2}$  also form triangular arrays as  $n$  varies. Given  $\omega$  and  $n$ , the randomness of  $\tilde{d}_n^{(2)*}$  comes from the bootstrapping step.

Under the above theoretical setting, in order to obtain the strong consistency of the bootstrapped estimator, we consider the following strong assumptions on the design points, the regression function and the random errors.

**(A2)** There exists a distribution function  $G$ , whose Lebesgue density  $g$  is positive and has a bounded first derivative on  $[0, 1]$ , such that  $\sup_{x \in [0, 1]} |F_n(x) - G(x)| \lesssim n^{-1/2}$ , where  $F_n$  is the empirical function of  $\{x_i\}_{i=1}^n$  and “ $\lesssim$ ” denotes that the left side is less than a constant times the right side.

**(A3)** The regression function  $f \in \mathcal{F}_0$  is differentiable on  $[0, 1]$  with  $\inf_{x \in [0, 1]} f'(x)$  and  $\sup_{x \in [0, 1]} f'(x)$  both positive and finite.

**(A4)** All the absolute moments of  $\epsilon$  are finite, i.e.  $\mathbb{E}|\epsilon|^q < \infty$  for all  $q \in \mathbb{N}$ .

*Remark 3.3.6.* There exist triangular arrays of design points satisfying (A2). For example, with  $x_i = i/n$  for  $i = 1, 2, \dots, n$  and all  $n$ , we have an array of discrete uniform designs on  $[0, 1]$ . Let  $G$  be the uniform distribution function on  $[0, 1]$ . Then, for this special array  $\sup_{x \in [0, 1]} |F_n(x) - G(x)| \leq 1/n$ . Note that (A2) is stronger than (A1). A random variable with finite moment generating function in a small neighborhood of 0 satisfies (A4), such as a normal random variable. The assumptions (A2) to (A4) are essentially the fixed design versions of the assumptions for Lemma 1 of [Durot \(2008\)](#), a modification of which enables us to identify a crucial boundary rate for the almost sure convergence of the isotonic regression estimator of  $d_0$ . This boundary rate plays a central role in the strong consistency of the bootstrapped estimator.

Next, we state results on the strong consistency of  $\hat{\beta}_1$  and the conditional weak consistency of  $\hat{\beta}_1^*$  and then on strong consistency of the bootstrapped estimator. Note



that  $P^*$  denotes the probability of the bootstrapped data conditional on the original data.

*Lemma 3.3.7.* If  $f \in \mathcal{F}$ ,  $\gamma \in (0, 1/2)$  and (A2) to (A4) hold,

$$\hat{\beta}_1 \rightarrow f'(d_0), \quad (P - a.s.), \quad \hat{\beta}_1^* \xrightarrow{P^*} f'(d_0), \quad (P - a.s.),$$

where  $\xrightarrow{P^*}$  denotes convergence in probability conditional on a given  $\omega$ .

*Theorem 3.3.8.* If  $f \in \mathcal{F}$ ,  $\gamma \in (0, 1/3)$  and (A2) to (A4) hold,

$$n^{1/2}(\tilde{d}_n^{(2)*} - \tilde{d}_n^{(2)}) \xrightarrow{d^*} C_2 Z_1, \quad (P - a.s.),$$

where  $C_2$  and  $Z_1$  are as in Theorem 3.3.3. That is,

$$\sup_{t \in \mathbb{R}} |P^* \left( n^{1/2}(\tilde{d}_n^{(2)*} - \tilde{d}_n^{(2)}) \leq t \right) - P(C_2 Z_1 \leq t)| \xrightarrow{a.s.} 0.$$

From the above strong consistency, the corresponding weak consistency follows under the same set of conditions. However, weak consistency can be obtained with the following weaker requirement on the random error:

**(A5)** The sixth moment of  $\epsilon$  is finite, i.e.  $\mathbb{E}\epsilon^6 < \infty$ .

*Theorem 3.3.9.* If  $f \in \mathcal{F}$ ,  $\gamma \in (0, 1/3)$  and (A1) and (A5) hold, for  $t \in \mathbb{R}$ ,

$$\sup_{t \in \mathbb{R}} |P^* \left( n^{1/2}(\tilde{d}_n^{(2)*} - \tilde{d}_n^{(2)}) \leq t \right) - P(C_2 Z_1 \leq t)| \xrightarrow{P} 0,$$

where  $C_2$  and  $Z_1$  are as in Theorem 3.3.3.

*Remark 3.3.10.* Comparing Theorem 3.3.8 with Theorem 3.3.3, we see that, under the strong assumption (A5) on the random errors, the bootstrapped estimator is strongly consistent for  $f \in \mathcal{F}$  and  $\gamma \in (1/4, 1/3)$ , which is exactly the  $\gamma$ -range of most interest. Further, if  $f$  is locally linear at  $d_0$ , i.e.  $f \in \mathcal{F}_2$ , the strong consistency continues to hold for  $\gamma \in (1/8, 1/4]$ . Similar conclusions on weak consistency hold by comparing Theorem 3.3.9 with Theorem 3.3.3, but under the milder assumption (A5) on the random errors.

### 3.4 Performance Evaluation

In this section, through an extensive simulation study we investigate the finite sample performance of the One-Stage Procedure (henceforth, OSP), the proposed Two-Stage Procedure (TSP) and its bootstrapped variant (BTSP).

Notice that for practically implementing the OSP, as well as the two-stage procedures, estimates of  $f'(d_0)$  and  $\sigma^2$  need to be obtained; the resulting procedures are called POSP, PTSP and PBTSP, respectively (Practical OSP, TSP and BTSP). For  $\sigma^2$ , we employ the nonparametric estimator proposed by [Gasser et al. \(1986\)](#), which is based on local linear fitting. Suppose the data  $\{(x_i, Y_i)\}_{i=1}^n$  are already sorted in ascending order of  $x_i$ 's. Then, we calculate

$$S^2 = (n_1 - 2)^{-1} \sum_{i=2}^{n-1} c_i^2 \tilde{\epsilon}_i^2,$$

where  $\tilde{\epsilon}_i = a_i Y_{i-1} + b_i Y_{i+1} - Y_i$ ,  $c_i^2 = (a_i^2 + b_i^2 + 1)^{-1}$ ,  $a_i = (x_{i+1} - x_i)/(x_{i+1} - x_{i-1})$  and  $b_i = (x_i - x_{i-1})/(x_{i+1} - x_{i-1})$ , for  $i = 2, 3, \dots, n-1$ . An estimate of  $f'(d_0)$  is obtained through the local quadratic regression estimator proposed by [Fan and Gijbels \(1996\)](#), at the estimate  $\hat{d}_n^{(1)}$ . Specifically, let  $K(\cdot)$  denote the Epanechnikov kernel function and  $h > 0$  the bandwidth, so that  $K_h(\cdot) = (1/h)K(\cdot/h)$ . Further, let  $\hat{\eta} = (\hat{\eta}_0, \hat{\eta}_1, \hat{\eta}_2)$  be given by

$$\hat{\eta} = \underset{\eta \in \mathbb{R}^3}{\operatorname{argmin}} \sum_{i=1}^n \left[ Y_i - \sum_{j=0}^2 \eta_j (x_i - \hat{d}_n^{(1)})^j \right]^2 K_h(x_i - \hat{d}_n^{(1)}).$$

Then, the local quadratic regression estimator of  $f'(\hat{d}_n^{(1)})$  is given by  $\hat{\eta}_1$ . The bandwidth  $h$  is chosen by first fitting a fifth order polynomial function to the data to obtain  $\hat{f}_{pol}(x) = \sum_{j=0}^5 \hat{\alpha}_j x^j$ . Next, the estimate of the third order derivative of  $f$  at  $\hat{d}_n^{(1)}$  is obtained by  $\hat{f}_{pol}^{(3)}(\hat{d}_n^{(1)}) = 6\hat{\alpha}_3 + 24\hat{\alpha}_4 \hat{d}_n^{(1)} + 60\hat{\alpha}_5 (\hat{d}_n^{(1)})^2$ . Finally the bandwidth  $h$  is calculated as

$$\hat{h}_{opt} = C_{1,2}(K) \left[ S^2 / (\hat{f}_{pol}^{(3)}(\hat{d}_n^{(1)}))^2 \right]^{1/7} n^{-1/7},$$

where  $C_{1,2}(K) = 2.275$ .

For the two-stage procedures, the tuning parameters  $\gamma$  and  $K$  need to be specified for obtaining the second-stage sampling points  $L$  and  $U$ . We select them as the end points of a *high level Wald-type confidence interval* calculated from the first-stage

data; that is,  $\gamma$  and  $K$  satisfy

$$Kn_1^{-\gamma} = Cq_\beta n_1^{-1/3}, \quad (3.9)$$

where  $q_\beta$  is the upper  $\beta$  quantile of  $\mathbb{Z}$ . On the other hand, a good quantitative rule for selecting the first-stage sample proportion  $p$  is not available; nevertheless, a practical qualitative rule of thumb dictates that  $p$  should decrease, while  $np$  should increase as the sample size increases. In our simulation study a number of different values for  $p$  are considered.

Finally, due to presence of small sample sizes the following modification of the second-stage estimator is adopted:

$$\tilde{d}_n^{(2)} = \begin{cases} \min(\max((\theta_0 - \hat{\beta}_0)/\hat{\beta}_1, 0), 1) & \text{if } \hat{\beta}_1 > 0, \\ \hat{d}_{n_1}^{(1)} & \text{otherwise.} \end{cases}$$

The same modification applies to the bootstrapped second-stage estimator in BTSP.

*Remark 3.4.1.* Note that our method for choosing the tuning parameters  $\gamma, K$  brings in another subjective parameter  $\beta$ . However, the choice of  $\beta$  is guided by a rational principle, namely the requirement that the chosen interval contain the truth with high probability. The magnitude of  $\beta$  is related to how conservative the experimenter wants to be in the construction of the second stage interval which is fundamentally subjective. Also, our rule of thumb regarding the choice of  $p$  is based on the idea that with larger budgets smaller  $p$ 's at stage one will still lead to reasonably precise sampling intervals at stage two, leaving a larger proportion of points for stage two and the possibility of more accurate conclusions.

The basic settings of the simulation study are as follows: two regression functions are considered,  $f_1(x) = x^2 + x/5$  and  $f_2(x) = e^{4(x-0.5)}/(1 + e^{4(x-0.5)})$  for  $x \in [0, 1]$ . The first-stage design points are drawn from a discrete uniform distribution on  $[0, 1]$ , i.e.  $x_i = i/(n_1 + 1)$ . Further, the target is set to  $d_0 = 0.5$ , the standard deviation of the random error  $\sigma$  to 0.1, 0.3 and 0.5, the sample size  $n$  ranges from 50 to 500 in increments of 50, while the first-stage sample proportion  $p$  ranges from 0.2 to 0.8 in increments of 0.1. Finally, the levels of significance  $\alpha$  and  $\beta$  are set to 0.025. Note that  $\beta$  is only required to be small and the specific choice of 0.025 is somewhat arbitrary. The following quantities are computed: coverage rates and average lengths of confidence intervals, and mean squared errors of estimators. The

simulation programs and more results can be found on the first author’s webpage: [www.stat.lsa.umich.edu/~rltang](http://www.stat.lsa.umich.edu/~rltang). In this paper, we show part of the results for saving space.

*Remark 3.4.2.* Choosing  $\gamma$  and  $K$  via equation (3.9) is theoretically equivalent to having  $\gamma = 1/3$  and  $K = Cq_\beta$ . Notice that strictly speaking, neither strong nor weak consistency for  $\gamma = 1/3$  is expected to hold for the bootstrapped estimator. However, it is reasonable to expect that for realistic sample sizes, the performance of the bootstrap would be satisfactory, since  $\gamma = 1/3$  is at the boundary of consistency. The obtained simulation results certainly vindicate this expectation. We would like to note that there are other bootstrap methods that could have been used, like the wild or residual bootstrap or the  $m$  out of  $n$  bootstrap, but it is not clear whether they would yield consistency at  $\gamma = 1/3$ . It would be interesting to explore some of these issues in future work.

### 3.4.1 Comparison of Two-Stage Procedures

By Theorem 3.2.1, from the first-stage data, an asymptotic  $(1 - 2\beta)$  confidence interval for  $d_0$  with the true parameter is given by:

$$[\hat{d}_{n_1}^{(1)} - Cq_\beta n_1^{-1/3}, \hat{d}_{n_1}^{(1)} + Cq_\beta n_1^{-1/3}] \cap [0, 1].$$

We consider the above confidence interval as the sampling interval  $[L, U]$  with  $\gamma = 1/3$  and  $K = Cq_\beta$ . Then, by Theorem 3.3.3, for  $f \in \mathcal{F}$  and  $\gamma = 1/3$ ,

$$n^{1/2}(\tilde{d}_n^{(2)} - d_0) \xrightarrow{d} C_2 Z_1 + C_3 \mathbb{Z} Z_2.$$

Hence, the corresponding asymptotic  $(1 - 2\alpha)$  confidence interval of  $d_0$  is given by:

$$[\tilde{d}_n^{(2)} - \tilde{q}_\alpha n^{-1/2}, \tilde{d}_n^{(2)} + \tilde{q}_\alpha n^{-1/2}] \cap [0, 1], \quad (3.10)$$

where  $\tilde{q}_\alpha$  is the upper  $\alpha$  quantile of  $C_2 Z_1 + C_3 \mathbb{Z} Z_2$ .

Next we compare the two-stage procedures, focusing on the coverage rates. In the first row of Figure 3.2, the coverage rates of the (3.10) confidence intervals for combinations of  $f, n$  and  $\sigma$  are shown based on 5000 replications, using the *true* parameters  $f'(d_0)$  and  $\sigma$  (i.e. the true  $C, C_2$  and  $C_3$  in constructing the confidence

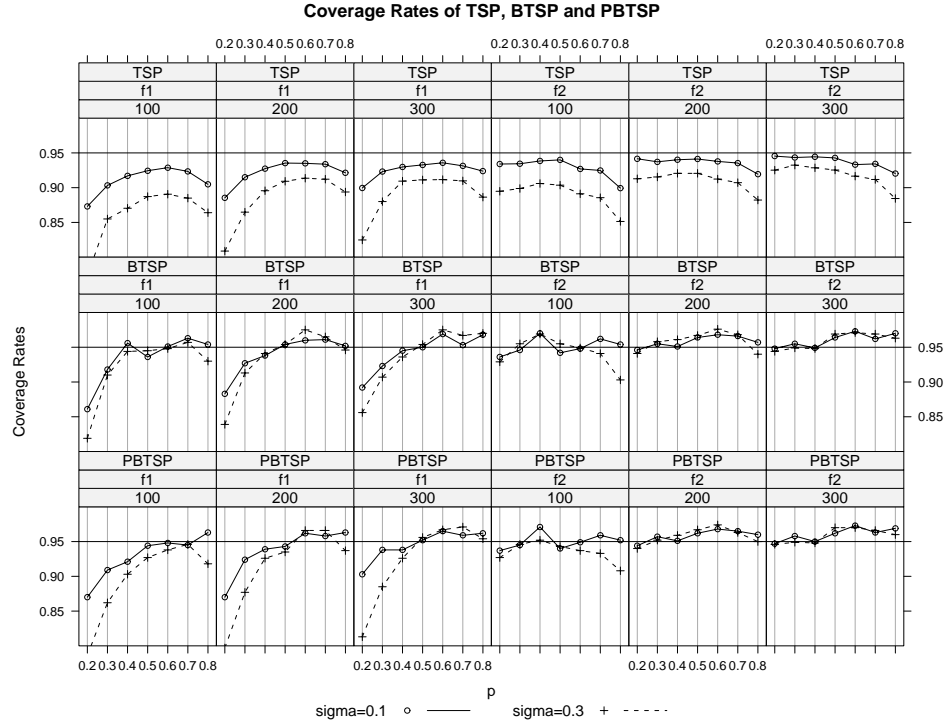


Figure 3.2: Coverage Rate plot grouped with different  $\sigma$ 's.

intervals). It can be seen that in general, coverage rates are below the nominal level 0.95, which is depicted by a solid horizontal line in each subplot. This reflects that  $\tilde{d}_n^{(2)}$  usually has slow speed of convergence in distribution. As expected, the results improve for small noise levels, larger sample sizes and functions closer to linearity in the vicinity of  $d_0$ .

The second row in Figure 3.2 shows the coverage rates of the bootstrapped procedure, based on 1000 replicates and 3000 bootstrap samples per replicate, using the true parameters  $f'(d_0)$  and  $\sigma$  at stage one. It can be seen that the coverage rates achieve the nominal level with proper first-stage sample proportions, smaller values of which are preferred since both average lengths and mean square errors are usually increasing with  $p$  from simulation results not shown in this paper. It can be concluded that the BTSP exhibits superior performance to the TSP in terms of coverage rates, especially for settings with  $f_1$ , moderate noise and relatively small sample sizes.

Finally, the third row in Figure 3.2 depicts the coverage rates of the bootstrapped procedure, when both  $f'(d_0)$  and  $\sigma$  are estimated from the first-stage data. The results based on 1000 replicates and 3000 bootstrap samples per replicate indicate a high level of agreement with those of the BTSP, which in turn suggests that the

PBTSP is reliable in applications.

Our findings suggest that  $p = 0.4$  is a good conservative choice for functions exhibiting a strong linear trend in the vicinity of  $d_0$ , while  $p = 0.5$  is preferable otherwise.

### 3.4.2 Comparison of One- and Two-Stage Procedures

We compare next the POSP and the PBTSP, in terms of coverage rates and average lengths of confidence intervals. We also compare the mean squared errors of the first- and second-stage estimates of  $d_0$ . The results for POSP are based on 5000 replications, while those of PBTSP on 1000 replications and 3000 bootstrap samples per replication, due to its computational intensity. It can be seen from Table 3.1 that both procedures usually perform well in terms of coverage rates. Further, under the PBTSP, confidence intervals usually have shorter average lengths, and the estimates for  $d_0$  have smaller mean squared errors, with slightly more gains accruing in the  $f_2$  case. However, it needs to be pointed out that both procedures suffer in the case with large noise and small to moderate sample sizes, especially for  $f_1$ .

Table 3.1: CR, AL and MSE stand for coverage rates, average lengths and mean squared errors of PBTSP while CR1, AL1 and MSE1 stand for those of POSP. ALR and MSER are the ratios of CR over CR1 and MSE over MSE1, respectively.

$f$	$p$	$\sigma$	$n$	CR	CR1	AL	AL1	ALR	MSE	MSE1	MSER
$f_1$	0.5	0.1	100	<b>0.944</b>	<b>0.955</b>	0.06	0.13	<b>0.43</b>	2e-04	1e-03	<b>0.21</b>
			200	<b>0.943</b>	<b>0.953</b>	0.04	0.10	<b>0.37</b>	1e-04	7e-04	<b>0.15</b>
			300	<b>0.952</b>	<b>0.956</b>	0.03	0.09	<b>0.35</b>	7e-05	5e-04	<b>0.14</b>
		0.3	100	<b>0.927</b>	<b>0.942</b>	0.21	0.27	<b>0.79</b>	3e-03	5e-03	<b>0.58</b>
			200	<b>0.935</b>	<b>0.947</b>	0.14	0.21	<b>0.63</b>	1e-03	3e-03	<b>0.39</b>
			300	<b>0.956</b>	<b>0.947</b>	0.11	0.19	<b>0.58</b>	8e-04	2e-03	<b>0.33</b>
$f_2$	0.4	0.1	100	<b>0.971</b>	<b>0.966</b>	0.06	0.16	<b>0.40</b>	2e-04	1e-03	<b>0.16</b>
			200	<b>0.951</b>	<b>0.964</b>	0.04	0.12	<b>0.34</b>	1e-04	9e-04	<b>0.13</b>
			300	<b>0.950</b>	<b>0.966</b>	0.03	0.11	<b>0.31</b>	7e-05	7e-04	<b>0.11</b>
		0.3	100	<b>0.952</b>	<b>0.948</b>	0.24	0.32	<b>0.76</b>	5e-03	6e-03	<b>0.79</b>
			200	<b>0.959</b>	<b>0.956</b>	0.16	0.25	<b>0.62</b>	2e-03	4e-03	<b>0.46</b>
			300	<b>0.948</b>	<b>0.955</b>	0.12	0.22	<b>0.53</b>	8e-04	3e-03	<b>0.25</b>

*Remark 3.4.3.* One of the advantages of the bootstrap procedure, as pointed out in Subsection 3.2.3, is that its implementation does not require knowledge of  $f'(d_0)$ . One might feel that the practical implementation of the bootstrap procedure defeats this advantage, since  $f'(d_0)$  is estimated from the first-stage data to construct the second

stage sampling interval. However, note that only a rough and ready estimate of  $f'(d_0)$  would suffice for the purpose of setting the sampling interval. On the contrary, to set a confidence interval directly from the asymptotic distribution of the second-stage estimate requires a much more precise estimate of  $f'(d_0)$ . Thus, the really crucial advantage with the bootstrap is that it obviates the need for a precise estimate of  $f'(d_0)$ .

*Remark 3.4.4.* Notice that the sigmoid function  $f_2$  belongs to class  $\mathcal{F}_2$  for the case  $d_0 = 0.5$ , since its second-derivative vanishes at that point. It is of practical interest to investigate the performance of the PBTSP for the case where the regression function at the target point is close to, but not exactly, linear. We have examined the case for  $f_2$  and  $d_0 = 0.4$  and  $0.6$  under the previously considered settings. The curvatures (i.e. second derivatives) of the regression functions at these two points are about  $0.76$  and  $-0.76$ , respectively. The results are very close to those obtained for  $d_0 = 0.5$ .

*Remark 3.4.5.* In PBTSP, the second stage sampling points  $L$  and  $U$  are identified through a Wald-type confidence interval constructed via estimating  $f'(d_0)$  and  $\sigma^2$ , with  $\hat{d}_{n_1}^{(1)}$  at the center of  $[L, U]$ . An alternative, albeit ad-hoc way of obtaining an interval centered at  $\hat{d}_{n_1}^{(1)}$  is to set  $L = \hat{d}_{n_1}^{(1)} - L_n/2$  and  $U = \hat{d}_{n_1}^{(1)} + L_n/2$ , where  $L_n$  is the length of a testing-based confidence interval for  $d_0$  obtained from the first-stage data. This testing-based interval is obtained as follows: consider testing the hypothesis  $H_{0,d} : f^{-1}(\theta_0) = d$  vs  $H_{1,d} : f^{-1}(\theta_0) \neq d$ . Let  $\hat{f}^{(1)}$  denote the usual isotonic estimator of  $f$  from the stage one data and  $\hat{f}_d^{(1)}$  the constrained isotonic estimator under  $H_{0,d}$ . The residual sum of squares based test statistic is given by

$$RSS(d) = \frac{\sum_{i=1}^{n_1} (Y_i - \hat{f}_d^{(1)}(x_i))^2 - \sum_{i=1}^{n_1} (Y_i - \hat{f}^{(1)}(x_i))^2}{\hat{\sigma}^2},$$

where  $\hat{\sigma}^2$  is a consistent estimate of  $\sigma^2$ . The inversion procedure assigns  $d$  to the confidence set if  $RSS(d)$  falls below an appropriate threshold determined by a pre-specified quantile of its limit distribution  $\mathbb{D}$  (when  $d = d_0$  holds true), which is completely parameter-free and therefore enables the construction of the confidence set without the need for nuisance parameter estimation. The limit distribution of  $RSS(d^0)$  can be derived by adapting Theorem 2 of [Banerjee and Wellner \(2005\)](#) (where a likelihood ratio statistic is dealt with) to the residual sum of squares statistic in the nonparametric regression setting, but see also [Banerjee \(2007\)](#) and [Banerjee \(2009\)](#) for a unified treatment of likelihood ratio and residual sum of squares statistics

in monotone function problems.

Alternatively, we can use the extremities of the testing-based confidence interval themselves as the sampling points for the second stage. For both cases, simulations show that their results are very similar to those of PBTSP using the Wald-type confidence interval, thus implying that the procedure is not particularly sensitive to the exact specification of  $L$  and  $U$ . Note that although this testing-based approach has the merit of completely avoiding the estimation of  $f'(d_0)$ , the asymptotic properties of the corresponding two-stage estimator and its bootstrapped variant become intractable since neither the testing-based confidence interval nor the length  $L_n$  admits an easy analytical characterization, unlike the analytically simple Wald-type confidence intervals used in this paper. To conform to the theoretical development and to save space, we only present simulation results for such Wald-type stage two sampling intervals.

*Remark 3.4.6.* In the case of  $f \in \mathcal{F}_1$ , one may question the use of a linear working model for approximating  $f$  around  $d_0$ . Instead, fitting a higher order polynomial working model may seem more appropriate. We examined the case of  $f_1$  using a quadratic working model. The results show that this model improves the mean squared error of the estimates when the noise is large, but leads to substantial undercoverage.

*Remark 3.4.7.* Our simulation results indicate that good choices for  $p$  are 0.5 for  $f_1$  and 0.4 for  $f_2$ , respectively. Our practical recommendation is  $p = 0.5$ , whenever no prior information about the linearity of  $f$  around  $d_0$  is available.

### 3.5 Data Application

We apply our methods to the engineering problem introduced at the beginning of this paper. We briefly describe the underlying system next: consider a complex queueing system comprising  $N$  first-in-first-out infinite capacity queues holding different classes of customers and a set of service resources. These resources are externally modulated by a stochastic process. The main issue is to allocate the available resources to the queue in an appropriate manner so as to maximize the system's throughput. This system represents a canonical model for wireless data/voice transmissions, in flexible manufacturing and in call centers (for more details see [Bambos and Michailidis \(2004\)](#)).

An important quality of service metric is the average delay of jobs (over all classes).



This quantity can only be obtained through simulation of the system, due to its analytical intractability. The average delay of the jobs in a two-class system as a function of its loading under the optimal throughput policy introduced in [Bambos and Michailidis \(2004\)](#) is shown in Figure 3.1. It can be seen that delay is, in general, an increasing function of the loading. The response was obtained by a discrete event simulation of the system for each loading, based on 2,000 events. Notice that our ability to simulate the system at any loading in order to obtain the response, allows us to easily implement the proposed two-stage procedure.

It is of interest to estimate  $d_0 = f^{-1}(\theta_0)$  for  $\theta_0 = 10$  and 15 units of delay, since around loadings corresponding to those levels the quality of service provided by the system exhibits a significant deterioration. For comparing the one- and two-stage procedures we fix a budget of  $n = 82$ . A fixed design with spacing 0.01 was used in the interval  $[0.14, 0.95]$  to obtain the one-stage data shown in Figure 3.1 (also in the left-panel plots of Figure 3.4). It can be seen that the response is heteroskedastic, but this does not affect the isotonic regression based estimation of  $f$  and thus of  $d_0$ . However, it impacts the construction of confidence intervals through the estimation of the variance at  $d_0$ . To overcome this issue, the variance function is estimated locally by the method proposed in [Müller and Stadtmüller \(1987\)](#). More specifically, we compute the initial local variance estimates with the weights  $(1/\sqrt{2}, -1/\sqrt{2})$  and the smoothed variance function by using *gkerns* in the R package *lokern* with an adaptive bandwidth, shown in the left panel of Figure 3.3.

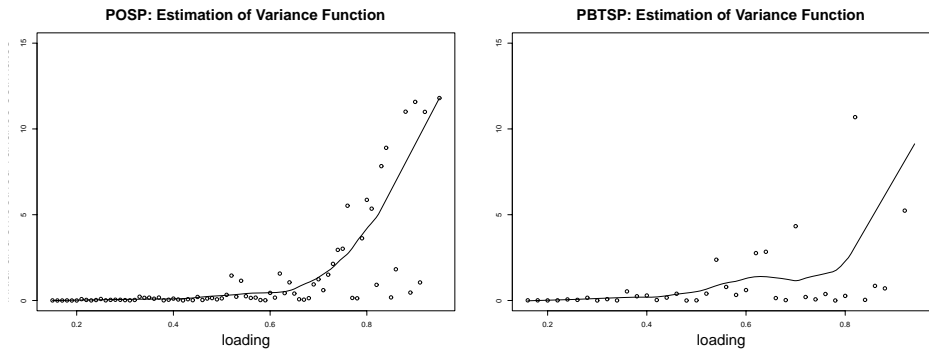


Figure 3.3: Estimation of the Variance Function in POSP and PBTSP

When implementing the two-stage procedure, we selected every other point from those used in the one-stage procedure ( $p = 0.5$ ), thus resulting in a fixed design with spacing 0.02 on the interval  $[0.14, 0.94]$ . The initial local variance estimates and

smoothed variance function with the first-stage data are shown in the right panel of Figure 3.3. After obtaining the  $40 = 2 \times 20$  second-stage responses, the second-stage estimator of  $d_0$  was computed using weighted least squares, with weights being the reciprocals of the estimated local variances at the corresponding sampling points.

The point estimates and the associated 95% confidence intervals from the POSP and the PBTSP are given in Table 3.2 and plotted in Figure 3.4. It can be seen that the point estimates are fairly similar. More significantly, the confidence intervals from PBTSP are much shorter than those from POSP, especially for the case  $\theta_0 = 10$ . This can be attributed to two factors: (i) the applicability of the linear model locally and (ii) the presence of a strong signal (small noise) for the design points around 0.8.

Table 3.2: Comparing POSP and PBTSP

		POSP $n = 82$	PBTSP $n = 81 = 41 + 2 \times 20$
$\theta = 10$	estimates of $d_0$	$\hat{d}_n^{(1)} = 0.803$	$\tilde{d}_n^{(2)} = 0.799$
	95% CI	[0.764, 0.841]	[0.794, 0.804]
$\theta = 15$	estimates of $d_0$	$\hat{d}_n^{(1)} = 0.863$	$\tilde{d}_n^{(2)} = 0.857$
	95% CI	[0.839, 0.887]	[0.845, 0.875]

### 3.6 Conclusions

In this study, a two-stage hybrid procedure for estimating an inverse regression function at a given point was introduced. The proposed procedure, by first obtaining a non-parametric estimate of the regression function and subsequently fitting a parametric linear model in an appropriately shrinking neighborhood of the parameter of interest, achieves a  $\sqrt{n}$  rate of convergence for the corresponding estimator. Note that isotonic regression was used in the first stage as it works with minimal assumptions on the underlying monotone regression function; nevertheless, other non-parametric procedures could be used. Further, the local approximation was primarily based on a linear model, although quadratic and suitable higher-order approximations could be used, especially in the presence of a small budget of design points, since the first stage sampling interval may not be short enough.

A bootstrapped version of the two-stage procedure is provided to overcome the difficulties posed by the requirement of estimating the derivative of the regression

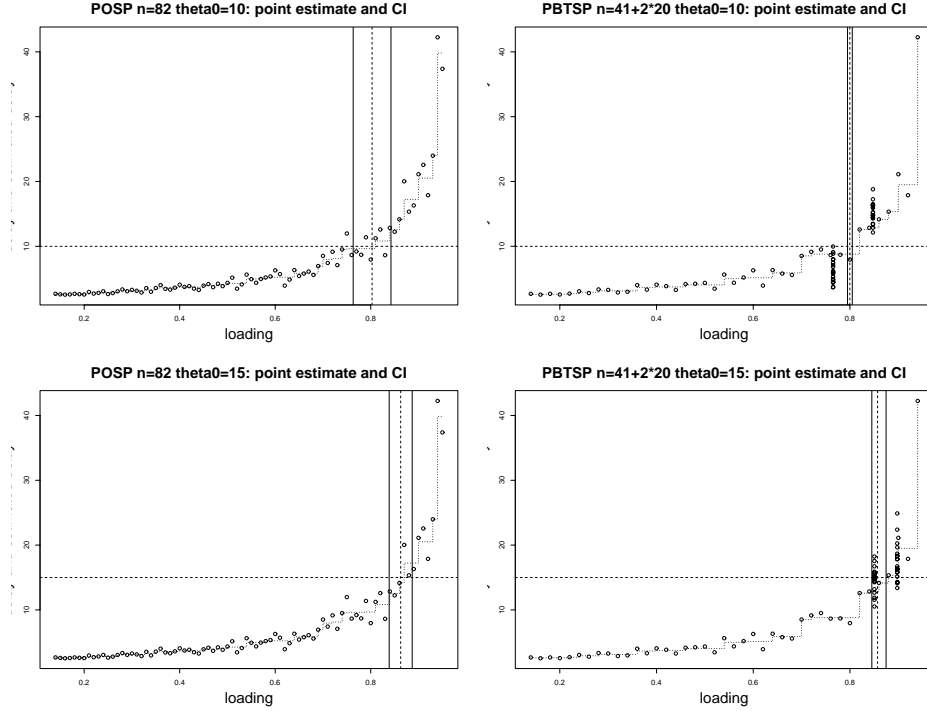


Figure 3.4: Comparing POSP and PBTSP

function at the unknown target point and the slow speed of convergence, especially with moderate sample sizes. Its asymptotic properties are also investigated and its strong consistency established (on this point see also Remark 3.7.4).

Our simulation results indicate that the practical bootstrapped procedure performs well in a variety of settings. Note that all the plans can be equipped with random designs for generating the first-stage data and similar asymptotic results follow. Nevertheless, for relatively small budgets, fixed designs (e.g. quantile based) usually yield improved performance.

Finally, we note that the main results generalize readily to heteroskedastic models of the form  $Y = f(x) + \sigma(x)\epsilon$ , where  $\sigma(x)$  is a scaling function that determines the error variance. Further, the proposed procedure should also work for discrete response models; for example, univariate binary and Poisson regression models with a monotone mean function. Qualitatively, the results are expected to be analogous to those established in this study; namely, a  $\sqrt{n}$  rate of convergence would be obtained for the estimator of the parameter of interest. However, the asymptotic behavior of the second-stage estimator and its bootstrap counterpart would be different and depend in an explicit manner on the specific model under consideration.

## 3.7 Appendix

### 3.7.1 Important Lemmas

In order to establish the strong consistency of the bootstrapped two-stage estimator, we need a rate of the almost sure convergence for the one-stage isotonic regression estimator  $\hat{d}_n^{(1)}$  of  $d_0$ . The following lemma, which is the fixed design version of Lemma 1 in [Durot \(2008\)](#), provides a useful tail probability for  $\hat{d}_n^{(1)}$ .

*Lemma 3.7.1.* Suppose  $\mathbb{E}|\epsilon|^q < \infty$  for some  $q \geq 2$  and (A2) and (A3) hold. Then, there exists  $K > 1$ , depending on  $q$ , such that for every  $\theta \in \mathbb{R}$  and  $\eta > 0$ ,

$$P(|\hat{d}_n^{(1)} - d_0| \geq \eta) \leq K(n\eta^3)^{-q/2}.$$

*Proof.* It will be shown that (A2) implies

$$\sup_{u \in [0,1]} |F_n^{-1}(u) - G^{-1}(x)| \lesssim n^{-1/2}. \quad (3.11)$$

Recall that “ $\lesssim$ ” denotes that the left side is less than a constant times the right side. Then, reworking the proof of Lemma 1 in [Durot \(2008\)](#) for our fixed design setting and an increasing function, and replacing expression (13) in that Lemma with (3.11) ensures that all subsequent steps go through yielding the desired conclusion. To show (3.11) note that from (A2), we get  $|G^{-1}(u) - G^{-1}(v)| \lesssim |u - v|$  for every  $u, v \in [0, 1]$ . Then,

$$\begin{aligned} & \sup_{u \in [0,1]} |F_n^{-1}(u) - G^{-1}(u)| \\ &= \max\{|G^{-1}(G(x_i)) - G^{-1}(i/n)|, |G^{-1}(G(x_{i+1})) - G^{-1}(i/n)|, \\ & \quad \text{for } i = 1, 2, \dots, n-1, |G^{-1}(G(x_1))|, |G^{-1}(G(x_n)) - 1|\} \\ &\lesssim \max\{|G(x_i) - i/n|, |G(x_{i+1}) - i/n|, \text{ for } i = 1, 2, \dots, n-1, \\ & \quad |G(x_1) - 0|, |G(x_n) - 1|\} \\ &= \sup_{x \in [0,1]} |F_n(x) - G(x)| \end{aligned}$$

gives (3.11) again by (A2). □

With the help of Lemma 3.7.1, next we show that  $n^{1/3}$  is a boundary rate of almost sure convergence.

*Lemma 3.7.2.* If (A2) to (A4) hold, for each  $\alpha > 0$ ,

$$P\left(\lim_{n \rightarrow \infty} n^{1/3-\alpha} |\hat{d}_n^{(1)} - d_0| = 0\right) = 1.$$

Thus, for every  $r < 1/3$ ,  $\lim_{n \rightarrow \infty} n^r (\hat{d}_n^{(1)} - d_0) = 0$ , ( $P - a.s.$ ).

*Proof.* Use the notations  $K$ ,  $q$  and  $\eta$  in Lemma 3.7.1. Denote  $K' = K\eta^{-3q/2}$  and  $A_n = \{n^{1/3-\alpha} |\hat{d}_n^{(1)} - d_0| \geq \eta\}$ . By Lemma 3.7.1,  $P(A_n) \leq K'n^{-3\alpha q/2}$  for each  $\alpha > 0$ . On the other hand, (A4) allows  $q$  to be arbitrarily large. Choosing  $q > 2/(3\alpha)$  gives  $\sum_{n=1}^{\infty} P(A_n) \leq K' \sum_{n=1}^{\infty} n^{-3\alpha q/2} < \infty$ . Note that  $\eta > 0$  is arbitrary. Therefore,  $n^{1/3-\alpha} |\hat{d}_n^{(1)} - d_0|$  converges to 0 almost surely (see Corollary on Page 254–255 in [Shiryayev \(1995\)](#)), which completes the proof.  $\square$

*Remark 3.7.3.* Note that Lemmas 3.7.1 and 3.7.2 hold for not only sequences, but also triangular arrays of design points and random errors.

*Remark 3.7.4.* The proof of Lemma 3.7.2 implies  $n^{1/3-\alpha} (\hat{d}_n^{(1)} - d_0) \xrightarrow{a.s.} 0$  for each  $\alpha \in (0, 1/3)$  given  $q > 2/(3\alpha)$ . Then,  $\mathbb{E}|\epsilon|^8 < \infty$  ensures  $n^\beta (\hat{d}_n^{(1)} - d_0) \xrightarrow{a.s.} 0$  for each  $\beta < 1/4$ . However, this almost sure convergence result actually holds under a weaker condition  $\mathbb{E}|\epsilon|^3 < \infty$  by Theorem in [Makowski \(1975\)](#) and Remark 4 in [Makowski \(1973\)](#). This shows that it might be possible to weaken the assumption (A4) a little. Essentially, it means that it might be possible to weaken the condition on the random error in Lemma 3.7.1. In fact, this possibility has been mentioned in Durot's papers on isotonic regression [Durot \(2002, 2007, 2008\)](#).

### 3.7.2 Proofs for Results in Subsection 3.3.1

For the simplicity of notation, from now on denote  $\delta_d = \hat{d}_{n_1}^{(1)} - d_0$ ,  $\epsilon_i^+ = \epsilon_i'' + \epsilon_i'$ ,  $\epsilon_i^- = \epsilon_i'' - \epsilon_i'$ ,  $f_{UL}^+ = f(U) + f(L)$ ,  $f_{UL}^- = f(U) - f(L)$ ,  $R_{UL}^+ = R_U + R_L$ ,  $R_{UL}^- = R_U - R_L$ ,  $R_{UL}'^+ = R_U' + R_L'$ ,  $R_{UL}'^- = R_U' - R_L'$ . Recall  $Y_i^+ = Y_i'' + Y_i'$  and  $Y_i^- = Y_i'' - Y_i'$ .

*Proof of Lemma 3.3.2.* Consider the following Taylor's expansions:

$$\begin{aligned} f(U) = f(\hat{d}_{n_1}^{(1)} + Kn_1^{-\gamma}) &= f(d_0) + f'(d_0)(\delta_d + Kn_1^{-\gamma}) \\ &+ (1/2)f''(d_0)(\delta_d + Kn_1^{-\gamma})^2 + R_U, \end{aligned} \quad (3.12)$$

$$\begin{aligned}
f(L) = f(\hat{d}_{n_1}^{(1)} - Kn_1^{-\gamma}) &= f(d_0) + f'(d_0)(\delta_d - Kn_1^{-\gamma}) \\
&+ (1/2)f''(d_0)(\delta_d - Kn_1^{-\gamma})^2 + R_L, \tag{3.13}
\end{aligned}$$

where  $R_U = f'''(\xi_1)(\delta_d + Kn_1^{-\gamma})^3/6$ ,  $R_L = f'''(\xi_2)(\delta_d - Kn_1^{-\gamma})^3/6$ ,  $\xi_1$  lies between  $d_0$  and  $\hat{d}_{n_1}^{(1)} + Kn_1^{-\gamma}$  and  $\xi_2$  lies between  $d_0$  and  $\hat{d}_{n_1}^{(1)} - Kn_1^{-\gamma}$ . Since  $\hat{d}_{n_1}^{(1)}$  converges to  $d_0$  in probability by Theorem 3.2.1, so do  $\xi_1$  and  $\xi_2$ .

Then, from (3.5), the definitions of  $Y_i'$  and  $Y_i''$  and the Taylor expansions (3.12) and (3.13), we get

$$\begin{aligned}
\hat{\beta}_1 &= (2Kn_1^{-\gamma}n_2)^{-1} \sum_{i=1}^{n_2} Y_i^- = (2Kn_1^{-\gamma})^{-1} f_{UL}^- + (2Kn_1^{-\gamma}n_2)^{-1} \sum_{i=1}^{n_2} \epsilon_i^- \\
&= f'(d_0) + f''(d_0)\delta_d + (2Kn_1^{-\gamma})^{-1} R_{UL}^- + (2Kn_1^{-\gamma}n_2)^{-1} \sum_{i=1}^{n_2} \epsilon_i^-.
\end{aligned}$$

From Theorem 3.2.1,  $\delta_d \xrightarrow{P} 0$ ; and by the Lindeberg-Feller CLT for triangular arrays, for  $\gamma \in (0, 1/2)$ ,  $(n_1^\gamma/n_2) \sum_{i=1}^{n_2} \epsilon_i^- \xrightarrow{P} 0$ . Next we show that  $R_{UL}^-/(2Kn_1^{-\gamma}) \xrightarrow{P} 0$  for  $\gamma \in (0, 1)$ . Hence, for  $\gamma \in (0, 1/2)$  we get  $\hat{\beta}_1 \xrightarrow{P} f'(d_0)$ . It suffices to show both  $n_1^\gamma R_U$  and  $n_1^\gamma R_L$  converge to 0 in probability for  $\gamma \in (0, 1)$ . We only show the former; the latter follows in an analogous manner.

From the definition of  $R_U$ , we have

$$\begin{aligned}
n_1^\gamma R_U &= (1/6)n_1^\gamma f'''(\xi_1)(\delta_d + Kn_1^{-\gamma})^3 \\
&= (1/6)f'''(\xi_1) [n_1^\gamma \delta_d^3 + 3K\delta_d^2 + 3K^2n_1^{-\gamma}\delta_d + K^3n_1^{-2\gamma}]. \tag{3.14}
\end{aligned}$$

Theorem 3.2.1 coupled with Slutsky's Lemma, shows that the sum of the four terms within the square bracket in (3.14) is  $o_P(1)$  for  $\gamma \in (0, 1)$ . Thus, we have  $n_1^\gamma R_U = f'''(\xi_1)o_P(1)$ . Since  $f'''(\cdot)$  is uniformly bounded around  $d_0$  and  $\xi_1 \rightarrow d_0$  in probability,  $f'''(\xi_1)o_P(1) = o_P(1)$ . This shows that  $n_1^\gamma R_U$  converges to 0 in probability for  $\gamma \in (0, 1)$ . Obviously,  $R_U = o_P(1)$ .

Then, for  $\gamma \in (0, 1/2)$ ,

$$\begin{aligned}
\hat{\beta}_0 &= (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^+ - \hat{d}_{n_1}^{(1)} \hat{\beta}_1 = f(d_0) + (1/2)f''(d_0)[\delta_d^2 + K^2n_1^{-2\gamma}] \\
&+ f'(d_0)\delta_d + (1/2)R_{UL}^+ + (2n_2)^{-1} \sum_{i=1}^{n_2} \epsilon_i^+ - \hat{d}_{n_1}^{(1)} \hat{\beta}_1 \xrightarrow{P} f(d_0) - d_0 f'(d_0).
\end{aligned}$$

Finally, for  $\gamma \in (0, 1/2)$ , the weak consistency of  $\hat{\beta}_1$  and  $\hat{\beta}_0$  gives  $\tilde{d}_n^{(2)} = (\theta_0 - \hat{\beta}_0)/(\hat{\beta}_1) \xrightarrow{P} d_0$ .  $\square$

*Proof of Theorem 3.3.3.* First, suppose  $f \in \mathcal{F}_1$ . From (3.6), the definitions of  $Y_i'$  and  $Y_i''$  and the Taylor expansions (3.12) and (3.13), we get

$$\begin{aligned} \tilde{d}_n^{(2)} - d_0 &= (1/\hat{\beta}_1)[f(d_0) - (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^+] + \delta_d \\ &= (1/f'(d_0))[f(d_0) - (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^+] + \delta_d \\ &\quad + (f'(d_0)\hat{\beta}_1)^{-1}(f'(d_0) - \hat{\beta}_1) \left[ f(d_0) - (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^+ \right] \\ &= S_1 + S_2 \times S_3 \end{aligned}$$

where

$$\begin{aligned} S_1 &= -f''(d_0)(2f'(d_0))^{-1} (\delta_d^2 + K^2 n_1^{-2\gamma}) \\ &\quad - (2f'(d_0))^{-1} R_{UL}^+ - (2f'(d_0)n_2)^{-1} \sum_{i=1}^{n_2} \epsilon_i^+, \\ S_2 &= (f'(d_0)\hat{\beta}_1)^{-1} \left[ f''(d_0)\delta_d + (2Kn_1^{-\gamma})^{-1} R_{UL}^- + (2Kn_1^{-\gamma}n_2)^{-1} \sum_{i=1}^{n_2} \epsilon_i^- \right], \\ S_3 &= f'(d_0)\delta_d + (1/2)f''(d_0) (\delta_d^2 + K^2 n_1^{-2\gamma}) + (1/2)R_{UL}^+ + (2n_2)^{-1} \sum_{i=1}^{n_2} \epsilon_i^+. \end{aligned}$$

Next consider the exact stochastic orders of the terms  $S_1$ ,  $S_2$  and  $S_3$ . We start with  $S_1$ . From Theorem 3.2.1,  $\delta_d^2 = O_P(n^{-2/3})$ ; for  $\gamma > 0$ ,  $n_1^{-2\gamma} = O_P(n^{-2\gamma})$ ,  $R_U = O_P(n^{-1}) + O_P(n^{-3\gamma})$ ,  $R_L = O_P(n^{-1}) + O_P(n^{-3\gamma})$ , and  $n_2^{-1} \sum_{i=1}^{n_2} \epsilon_i^+ = O_P(n^{-1/2})$ . Note that these are the exact rates of weak convergence. Then, for  $\gamma \in (0, 1/2)$ ,  $S_1 = T_1 + T_2 + o_P(n^{-2\gamma} \vee n^{-1/2})$ , where

$$T_1 = -(2f'(d_0))^{-1} f''(d_0) K^2 n_1^{-2\gamma}, \quad T_2 = -(2f'(d_0)n_2)^{-1} \sum_{i=1}^{n_2} \epsilon_i^+.$$

Thus, the possible main terms of  $S_1$  are  $T_1$  and  $T_2$ . In the same way, we can obtain the main terms of  $S_2$  and  $S_3$  and then those of  $S_2 \times S_3$ . Finally we have  $S_1 + S_2 \times S_3 =$

$T_1 + T_2 + T_3 + R$ , where

$$T_3 = (2K\hat{\beta}_1 n_1^{-\gamma} n_2)^{-1} \delta_d \sum_{i=1}^{n_2} \epsilon_i^-, \quad R = o_P(n^{-2\gamma} \vee n^{-1/2} \vee n^{\gamma-5/6}).$$

It is easy to see that among the three rates  $n^{-2\gamma}$ ,  $n^{-1/2}$  and  $n^{\gamma-5/6}$ , the first, second or last one is slowest according as  $\gamma$  belongs to the interval  $(1, 1/4)$ ,  $(1/4, 1/3)$ , or  $(1/3, 1/2)$ , respectively; the first and the second are the slowest for  $\gamma = 1/4$ ; while the second and the last ones are the slowest for  $\gamma = 1/3$ . In other words,  $T_1$ ,  $T_2$  or  $T_3$  becomes the main term according as  $\gamma \in (0, 1/4)$ ,  $\gamma \in (1/4, 1/3)$  or  $\gamma \in (1/3, 1/2)$ , respectively. When  $\gamma = 1/4$ , both  $T_1$  and  $T_2$  become the main terms and when  $\gamma = 1/3$ , both  $T_2$  and  $T_3$  become the main terms.

Then, by Theorem 3.2.1, the Lindeberg-Feller CLT for triangular arrays, Slutsky's Lemma and the Continuous Mapping Theorem, and noting that  $n_1^{1/3} \delta_d$  is independent of  $n_2^{-1/2} \sum_{i=1}^{n_2} \epsilon_i^+$  and  $n_2^{-1/2} \sum_{i=1}^{n_2} \epsilon_i^-$  and that  $\epsilon_i^+$  is uncorrelated with  $\epsilon_i^-$ , we obtain the results of the five cases for  $f \in \mathcal{F}_1$  defined by the different ranges of  $\gamma$  in the statement of the theorem.

For the purpose of illustration, we outline the case  $\gamma = 1/3$ , for which  $T_2 + T_3$  is the main term with exact stochastic order  $O_P(n^{-1/2})$ . Thus  $n^{1/2}(\tilde{d}_n^{(2)} - d_0)$  and  $n^{1/2}(T_2 + T_3)$  have the same asymptotic distribution. Since

$$\left( n_1^{1/3} \delta_d, n_2^{-1/2} \sum_{i=1}^{n_2} \epsilon_i^+, n_2^{-1/2} \sum_{i=1}^{n_2} \epsilon_i^- \right) \xrightarrow{d} (C\mathbb{Z}, cZ_1, cZ_2),$$

where  $\mathbb{Z}$  follows Chernoff distribution, independent of  $Z_1$ ,  $Z_2$  which are iid  $N(0, 1)$ , and  $c = \sqrt{2}\sigma$ , by Continuous Mapping Theorem, we have

$$n^{1/2}(T_2 + T_3) \xrightarrow{d} -C_2 Z_1 + (1/K)C_2 C\mathbb{Z} Z_2.$$

Note that  $-C_2 Z_1$  can be replaced by  $C_2 Z_1$  since  $N(0, 1)$  and  $-N(0, 1)$  have the same distribution. In similar fashion, we obtain the asymptotic results for the other four cases.

Carefully examining the above proof reveals that the conclusions with  $\gamma \in (1/4, 1/2)$  also hold for  $f \in \mathcal{F}$ . Thus, it remains to show the cases  $f \in \mathcal{F}_2$  and  $\gamma \in (1/8, 1/4]$ .



For  $f \in \mathcal{F}_2$ , consider the following Taylor's expansions:

$$\begin{aligned} f(U) = f(\hat{d}_{n_1}^{(1)} + Kn_1^{-\gamma}) &= f(d_0) + f'(d_0)(\delta_d + Kn_1^{-\gamma}) \\ &+ (1/6)f'''(d_0)(\delta_d + Kn_1^{-\gamma})^3 + R'_U, \end{aligned} \quad (3.15)$$

$$\begin{aligned} f(L) = f(\hat{d}_{n_1}^{(1)} - Kn_1^{-\gamma}) &= f(d_0) + f'(d_0)(\delta_d - Kn_1^{-\gamma}) \\ &+ (1/6)f'''(d_0)(\delta_d - Kn_1^{-\gamma})^3 + R'_L, \end{aligned} \quad (3.16)$$

where  $R'_U = f^{(4)}(\xi_1)(\delta_d + Kn_1^{-\gamma})^4/24$ ,  $R'_L = f^{(4)}(\xi_2)(\delta_d - Kn_1^{-\gamma})^4/24$ ,  $\xi_1$  lies between  $d_0$  and  $\hat{d}_{n_1}^{(1)} + Kn_1^{-\gamma}$  and  $\xi_2$  lies between  $d_0$  and  $\hat{d}_{n_1}^{(1)} - Kn_1^{-\gamma}$ .

Then, for  $\gamma \in (1/8, 1/2)$ ,

$$\begin{aligned} \tilde{d}_n^{(2)} - d_0 &= (1/\hat{\beta}_1)[f(d_0) - (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^+] + \delta_d \\ &= (1/f'(d_0)) \left[ f(d_0) - (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^+ \right] + \delta_d \\ &+ (f'(d_0)\hat{\beta}_1)^{-1}(f'(d_0) - \hat{\beta}_1) \left[ f(d_0) - (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^+ \right] \\ &= S_1 + S_2 \times S_3, \end{aligned}$$

where

$$\begin{aligned} S_1 &= -(6f'(d_0))^{-1}f'''(d_0)\delta_d^3 - (2f'(d_0))^{-1}f'''(d_0)\delta_d K^2 n_1^{-2\gamma} \\ &\quad - (2f'(d_0))^{-1}R'_{UL} - (2f'(d_0)n_2)^{-1} \sum_{i=1}^{n_2} \epsilon_i^+, \\ S_2 &= (f'(d_0)\hat{\beta}_1)^{-1}[(1/2)f'''(d_0)\delta_d^2 + (1/6)f'''(d_0)K^2 n_1^{-2\gamma} \\ &\quad + (2Kn_1^{-\gamma})^{-1}R'_{UL} + (2Kn_1^{-\gamma}n_2)^{-1} \sum_{i=1}^{n_2} \epsilon_i^-], \\ S_3 &= \left\{ f'(d_0)\delta_d + \frac{f'''(d_0)}{6}\delta_d^3 + \frac{f'''(d_0)}{2}\delta_d K^2 n_1^{-2\gamma} + \frac{1}{2}R'_{UL} + \frac{1}{2n_2} \sum_{i=1}^{n_2} \epsilon_i^+ \right\}. \end{aligned}$$

Similar to the previous argument on the exact weak convergence rates,  $S_1 + S_2 \times$

$S_3 = T_1 + T_2 + R'$ , where

$$T_1 = -(2f'(d_0)n_2)^{-1} \sum_{i=1}^{n_2} \epsilon_i^+, \quad T_2 = (1/\hat{\beta}_1)\delta_d(2Kn_1^{-\gamma}n_2)^{-1} \sum_{i=1}^{n_2} \epsilon_i^-,$$

and  $R'$  is the sum of the remaining terms which converges to 0 faster than  $T_1$  and  $T_2$ . Then  $T_1$  becomes the main term for  $\gamma \in (1/8, 1/3)$  and the weak convergence result for  $f \in \mathcal{F}_2$  and  $\gamma \in (1/8, 1/4]$  follows easily from the Lindeberg-Feller central limit theorem for triangular arrays and Slutsky's lemma. This completes the proof.  $\square$

### 3.7.3 Proofs for Results in Subsection 3.3.2

To simplify arguments, we introduce a notation on the rate of almost sure convergence. Suppose  $\{\zeta_n\}$  is a sequence of random variables and  $b \in \mathbb{R}$ . Write  $\zeta_n = B_{as}(b)$  if  $n^\alpha \zeta_n$  converges to 0 almost surely for every  $\alpha < b$ . It is easy to verify that  $B_{as}(b_1) + B_{as}(b_2) = B_{as}(b_1)$  and  $B_{as}(b_1)B_{as}(b_2) = B_{as}(b_1 + b_2)$  if  $b_1 \leq b_2 \in \mathbb{R}$ . Note that  $\zeta_n = B_{as}(b)$  for some  $b > 0$  implies  $\zeta_n \rightarrow 0$  almost surely. Denote  $V_i^+ \equiv \epsilon_i^{*+} = \epsilon_i''^* + \epsilon_i'^*$  and  $V_i^- \equiv \epsilon_i^{*-} = \epsilon_i''^* - \epsilon_i'^*$ . Recall  $Y_i^{*+} = Y_i''^* + Y_i'^*$  and  $Y_i^{*-} = Y_i''^* - Y_i'^*$ .

*Proof of Lemma 3.3.7.* The proof of Lemma 3.3.2 establishes the weak consistency of  $\hat{\beta}_1$  for the case  $\gamma \in (0, 1/2)$ . In fact, under the setting of the bootstrapped two-stage procedure, the strong consistency of  $\hat{\beta}_1$  can be obtained.

From the proof of Lemma 3.3.2, it suffices to show  $\delta_d$ ,  $(n_1^\gamma/n_2) \sum_{i=1}^{n_2} \epsilon_i^-$  and  $R_{UL}^-/(2Kn_1^{-\gamma})$  converge to 0 almost surely. Lemma 3.7.2 shows that  $\delta_d$  converges to 0 almost surely, while Lemma 3.7.7 establishes that  $(n_1^\gamma/n_2) \sum_{i=1}^{n_2} \epsilon_i^-$  converges to 0 almost surely for  $\gamma \in (0, 1/2)$ . Thus, it suffices to show that both  $n_1^\gamma R_U$  and  $n_1^\gamma R_L$  converge to 0 almost surely for  $\gamma \in (0, 1)$ . Next, we show the former; the latter follows analogously.

Since  $\xi_1$  lies between  $d_0$  and  $\hat{d}_{n_1}^{(1)} + Kn_1^{-\gamma}$  and the latter converges to  $d_0$  almost surely, we know  $\xi_1$  converges to  $d_0$  almost surely. On the other hand,  $f'''(\cdot)$  is uniformly bounded around  $d_0$ ; thus,  $f'''(\xi_1)$  is almost surely bounded. Further, by Lemma 3.7.2, the four terms within square brackets on the right-side of (3.14) are  $B_{as}(1 - \gamma)$ ,  $B_{as}(2/3)$ ,  $B_{as}(1/3 + \gamma)$  and  $B_{as}(2\gamma)$ . Thus,  $n_1^\gamma R_U$  almost surely converges to 0 for  $\gamma \in (0, 1)$ .

So, for  $\gamma \in (0, 1/2)$ , we have  $\hat{\beta}_1 \rightarrow f'(d_0)$ , ( $P - a.s.$ ).

Next, we establish the conditional weak consistency of  $\hat{\beta}_1^*$  for  $f \in \mathcal{F}$ . From (3.8),

we get

$$\hat{\beta}_1^* = (2Kn_1^{-\gamma}n_2)^{-1} \sum_{i=1}^{n_2} Y_i^{*-} = T_1 + T_2,$$

where

$$T_1 = (2Kn_1^{-\gamma}n_2)^{-1} \sum_{i=1}^{n_2} \epsilon_i^{*-}, \quad T_2 = (2Kn_1^{-\gamma})^{-1} f_{UL}^-.$$

Hence, we have  $T_1 = T_{11} + T_{12}$ , where

$$T_{11} = s(2Kn_1^{-\gamma}n_2)^{-1} \sum_{i=1}^{n_2} (V_i^- - \nu^-)/s, \quad T_{12} = (2Kn_1^{-\gamma}n_2)^{-1} \sum_{i=1}^{n_2} \epsilon_i^-,$$

$V_i^- = \epsilon_i^{*-}$ ,  $\nu^- = E_*[V_i^-] = (1/n_2) \sum_{i=1}^{n_2} \epsilon_i^-$ , and

$$s^2 = \text{Var}_*[V_i^-] = \frac{1}{n_2} \sum_{i=1}^{n_2} (\epsilon_i'')^2 - \left( \frac{1}{n_2} \sum_{i=1}^{n_2} \epsilon_i'' \right)^2 + \frac{1}{n_2} \sum_{i=1}^{n_2} (\epsilon_i')^2 - \left( \frac{1}{n_2} \sum_{i=1}^{n_2} \epsilon_i' \right)^2.$$

For  $\gamma \in (0, 1/2)$ , gives that  $T_{12} \rightarrow 0$ , ( $P - a.s.$ ) by Lemma 3.7.7 and  $T_{11} \xrightarrow{P^*} 0$ , ( $P - a.s.$ ) by Lemma 3.7.8 and Slutsky's Lemma. Thus,  $T_1 \xrightarrow{P^*} 0$ , ( $P - a.s.$ ).

Next, we consider  $T_2$ . By the almost sure convergence of  $\delta_d$  and  $R_{UL}^-/(2Kn_1^{-\gamma})$ , we have, for  $\gamma \in (0, 1)$ ,

$$T_2 = f'(d_0) + f''(d_0)\delta_d + (2Kn_1^{-\gamma})^{-1}R_{UL}^- \rightarrow f'(d_0), \quad (P - a.s.).$$

Thus, for  $f \in \mathcal{F}$  and  $\gamma \in (0, 1/2)$ ,  $T_2 \rightarrow f'(d_0)$ , ( $P - a.s.$ ). Therefore, we get  $\hat{\beta}_1^* \xrightarrow{P^*} f'(d_0)$ , ( $P - a.s.$ ).  $\square$

*Proof of Theorem 3.3.8.* From (3.6) and (3.7),

$$n^{1/2}(\tilde{d}_n^{(2)*} - \tilde{d}_n^{(2)}) = -T_1 + T_2,$$

where

$$\begin{aligned}
T_1 &= (f'(d_0)2n_2)^{-1}n^{1/2} \sum_{i=1}^{n_2} (Y_i^{*+} - Y_i^+) \\
T_2 &= n^{1/2} \left[ \left(1/\hat{\beta}_1^* - 1/f'(d_0)\right) \left(f(d_0) - (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^{*+}\right) \right. \\
&\quad \left. - \left(1/\hat{\beta}_1 - 1/f'(d_0)\right) \left(f(d_0) - (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^+\right) \right].
\end{aligned}$$

By the definitions of  $Y_i', Y_i'', Y_i'^*, Y_i''^*$ ,

$$T_1 = n^{1/2}(f'(d_0)2n_2)^{-1} \sum_{i=1}^{n_2} (\epsilon_i^{*+} - \epsilon_i^+) = sn^{1/2}(2f'(d_0)n^{1/2})^{-1} \sum_{i=1}^{n_2} \frac{V_i^+ - \nu^+}{s\sqrt{n_2}},$$

where

$$V_i^+ = \epsilon_i^{*+}, \nu^+ = E_\star[V_i^+] = (1/n_2) \sum_{i=1}^{n_2} \epsilon_i^+,$$

and  $s^2 = \text{Var}_\star[V_i^+]$ , equal to that  $s^2$  in the proof of Lemma 3.3.7.

Lemma 3.7.6 gives  $s^2 \rightarrow 2\sigma^2$ , ( $P - a.s.$ ) and Lemma 3.7.8 gives  $\sum_{i=1}^{n_2} (V_i^+ - \nu^+)/ (s\sqrt{n_2}) \xrightarrow{d^\star} Z_1$ , ( $P - a.s.$ ). Note that  $\sqrt{n}/\sqrt{n_2} \rightarrow \sqrt{2/(1-p)}$ . Thus, Slutsky's lemma implies

$$T_1 \xrightarrow{d^\star} \frac{\sigma}{f'(d_0)(1-p)^{1/2}} Z_1, \quad (P - a.s.).$$

In Lemma 3.7.5 following this proof, we show that for  $\gamma \in (0, 1/3)$ ,  $T_2 \xrightarrow{P^\star} 0$ , ( $P - a.s.$ ). Therefore, another application of Slutsky's Lemma completes the proof. □

*Lemma 3.7.5.* For  $f \in \mathcal{F}$  and  $\gamma \in (0, 1/3)$ ,  $T_2 \xrightarrow{P^\star} 0$ , ( $P - a.s.$ ).

*Proof.* Let

$$\begin{aligned}
I &= \hat{\beta}_1 - f'(d_0), \quad II = f(d_0) - (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^+, \\
A &= \hat{\beta}_1^* - \hat{\beta}_1, \quad B = (2n_2)^{-1} \sum_{i=1}^{n_2} (\epsilon_i^{*+} - \epsilon_i^+), \\
T_{21} &= n^{1/2} A \cdot I \cdot II, \quad T_{22} = n^{1/2} I \cdot B, \\
T_{23} &= n^{1/2} II \cdot A, \quad T_{24} = n^{1/2} A \cdot B.
\end{aligned}$$

Then,

$$\begin{aligned}
T_2 &= n^{1/2} \left\{ -(\hat{\beta}_1^* f'(d_0))^{-1} [I + A] \cdot [II - B] + (\hat{\beta}_1 f'(d_0))^{-1} I \cdot II \right\} \\
&= (\hat{\beta}_1 \hat{\beta}_1^* f'(d_0))^{-1} n^{1/2} A \cdot I \cdot II \\
&\quad - (\hat{\beta}_1^* f'(d_0))^{-1} [-n^{1/2} I \cdot B + n^{1/2} II \cdot A - n^{1/2} A \cdot B] \\
&= (\hat{\beta}_1 \hat{\beta}_1^* f'(d_0))^{-1} T_{21} - (\hat{\beta}_1^* f'(d_0))^{-1} [-T_{22} + T_{23} - T_{24}].
\end{aligned}$$

It will be shown that  $T_{2i} \xrightarrow{P^*} 0$ , ( $P - a.s.$ ),  $i = 1, 2, 3, 4$  for  $\gamma \in (0, 1/3)$ . Thus, by Lemma 3.3.7 and Slutsky's Lemma, the conclusion of this lemma holds.

We establish next the convergence of the terms  $T_{2i}$ . From (3.5), (3.8), the definitions of  $Y_i'$ ,  $Y_i''$ ,  $Y_i'^*$  and  $Y_i''^*$ , and the Taylor's expansions of  $f(L)$  and  $f(U)$  ((3.12) and (3.13)), we have

$$\begin{aligned}
A &= (2Kn_2)^{-1} n_1^\gamma \sum_{i=1}^{n_2} (\epsilon_i^{*-} - \epsilon_i^-) = (2Kn_2^{1/2})^{-1} n_1^\gamma s n_2^{-1/2} \sum_{i=1}^{n_2} (V_i^- - \nu^-) / s, \\
B &= (2n_2)^{-1} \sum_{i=1}^{n_2} (\epsilon_i^{*+} - \epsilon_i^+) = (2n_2^{1/2})^{-1} s n_2^{-1/2} \sum_{i=1}^{n_2} (V_i^+ - \nu^+) / s, \\
I &= \hat{\beta}_1 - f'(d_0) = f''(d_0) \delta_d + (2K)^{-1} n_1^\gamma R_{UL}^- + (2Kn_2)^{-1} n_1^\gamma \sum_{i=1}^{n_2} \epsilon_i^-, \\
II &= f(d_0) - (2n_2)^{-1} \sum_{i=1}^{n_2} Y_i^+ \\
&= -f'(d_0) \delta_d - (1/2) f''(d_0) (\delta_d^2 + K^2 n_1^{-2\gamma}) - (1/2) R_{UL}^+ - (1/n_2) \sum_{i=1}^{n_2} \epsilon_i^+.
\end{aligned}$$

First consider  $T_{21}$ . We have

$$T_{21} = n^{1/2} A \cdot I \cdot II = T'_{21} s n_2^{-1/2} \sum_{i=1}^{n_2} (V_i^- - \nu^-) / s,$$

where  $T'_{21} = C_n \cdot I \cdot II$  and  $C_n = n^{1/2} n_1^\gamma (2Kn_2^{1/2})^{-1}$ . Lemmas 3.7.6 and 3.7.8 give

$$s \rightarrow \sqrt{2}\sigma, \quad (P - a.s.), \quad n_2^{-1/2} \sum_{i=1}^{n_2} (V_i^- - \nu^-) / s \xrightarrow{d^*} Z_2, \quad (P - a.s.).$$

Next, it will be shown that  $T'_{21}$  converges to 0  $P$ -almost surely for  $\gamma \in (0, 5/12)$ .

Then, an application of Slutsky's Lemma gives  $T_{21} \xrightarrow{P^*} 0$ , ( $P - a.s.$ ).

With the notation introduced at the beginning of this subsection and by Lemmas 3.7.7 and 3.7.2, we have, for  $\gamma > 0$ ,  $n_1^\gamma = B_{as}(-\gamma)$ ,  $(\delta_d) = B_{as}(1/3)$ ,  $\sum_{i=1}^{n_2} (\epsilon_i'' + \epsilon_i')/n_2 = B_{as}(1/2)$  and  $\sum_{i=1}^{n_2} (\epsilon_i'' - \epsilon_i')/n_2 = B_{as}(1/2)$ . Both  $R_U$  and  $R_L$  are equal to  $B_{as}(1) + B_{as}(\gamma + 2/3) + B_{as}(2\gamma + 1/3) + B_{as}(3\gamma)$ . Thus we have  $C_n = B_{as}(-\gamma)$ ,  $I = B_{as}(1/3) + B_{as}(-\gamma)[B_{as}(1) + B_{as}(\gamma + 2/3) + B_{as}(2\gamma + 1/3) + B_{as}(3\gamma)] + B_{as}(1/2 - \gamma) = B_{as}(1/3) + B_{as}(2\gamma) + B_{as}(1/2 - \gamma)$  and  $II = B_{as}(1/3) + [B_{as}(2/3) + B_{as}(2\gamma)] + (B_{as}(1) + B_{as}(\gamma + 2/3) + B_{as}(2\gamma + 1/3) + B_{as}(3\gamma)) + B_{as}(1/2) = B_{as}(1/3) + B_{as}(2\gamma)$ . Thus, for  $\gamma \in (0, 1/2)$ ,

$$\begin{aligned} T'_{21} &= C_n \cdot I \cdot II \\ &= B_{as}(-\gamma) \times [B_{as}(1/3) + (B_{as}(2\gamma)) + B_{as}(1/2 - \gamma)] \\ &\quad \times \{B_{as}(1/3) + B_{as}(2\gamma)\} \\ &= B_{as}(2/3 - \gamma) + B_{as}(1/3 + \gamma) + B_{as}(5/6 - 2\gamma) + B_{as}(3\gamma). \end{aligned}$$

It is easy to see that when  $\gamma \in (0, 5/12)$ , the above upper bounds  $1/2 - \gamma$ ,  $1/4 + \gamma$ ,  $3/4 - 2\gamma$ , and  $3\gamma$  are all positive. This implies that  $T'_{21}$  converges to 0  $P$ -almost surely for  $\gamma \in (0, 5/12)$ . Therefore, for  $\gamma \in (0, 5/12)$ ,  $T_{21}$  converges to 0 in probability ( $P - a.s.$ ).

Similarly, we can show that  $T_{2i}$ ,  $i = 2, 3$  or  $4$ , converges to 0 in probability ( $P - a.s.$ ), but with different intervals for  $\gamma$ . We next list these results. For  $\gamma \in (0, 1/2)$ ,  $T_{22}$  and  $T_{24}$  converge to 0 in probability ( $P - a.s.$ ) and for  $\gamma \in (0, 1/3)$ ,  $T_{23}$  converges to 0 in probability ( $P - a.s.$ ). It is worthwhile to note that  $\mathcal{F}$  can be considered directly because the  $B_{as}(1/3 - \gamma)$  term in  $T_{23}$  does not depend on  $f''(d_0)$ . Since  $1/3 < 5/12 < 1/2$ ,  $T_{2i}$  converges to 0 in probability ( $P - a.s.$ ) for  $i = 1, 2, 3, 4$  and  $\gamma \in (0, 1/3)$ . Thus, for  $f \in \mathcal{F}$  and  $\gamma \in (0, 1/3)$ ,  $T_2$  converges to 0 in probability ( $P - a.s.$ ).  $\square$

*Proof of Theorem 3.3.9.* Consider  $0 < \gamma < 1/3$ . Given an arbitrary subsequence  $\{n_k\}_{k=1}^\infty$  of  $\{n\}_{n=1}^\infty$ , let  $n_1 = np$  and  $n_{k,1} = n_k p$ . By Theorem 3.2.1, we know that  $n_1^\gamma (\delta_d) \equiv (np)^\gamma (\hat{d}_{np}^{(1)} - d_0) \xrightarrow{P} 0$ . It follows, by the relationship between convergence in probability and almost sure convergence (for example, see Theorem 20.5 in *Billingsley (1995)*), that there exists  $\{n_{k(i)}\}_{i=1}^\infty$ , a further subsequence of  $\{n_k\}$ , such that  $n_{k(i),1}^\gamma (\hat{d}_{n_{k(i),1}}^{(1)} - d_0) \rightarrow 0$ , ( $P - a.s.$ ). It now suffices to show that

$$n_{k(i)}^{1/2} (\tilde{d}_{n_{k(i)}}^{(2)\star} - \tilde{d}_{n_{k(i)}}^{(2)}) \xrightarrow{d^*} C_2 Z_1, \quad (P - a.s.).$$

Let  $n_{k(i),2} = n_{k(i)}(1-p)/2$ . Write  $\zeta_{n_{k(i)}} = B_{as}(b)$  if  $n_{k(i)}^\alpha \zeta_{n_{k(i)}}$  converges to 0 almost surely for every  $\alpha < b$ . As in the proof of Theorem 3.3.8, write  $n_{k(i)}^{1/2}(\tilde{d}_{n_{k(i)}}^{(2)\star} - \tilde{d}_{n_{k(i)}}^{(2)})$  as  $-T_1 + T_2$ , where both  $T_1$  and  $T_2$  are now indexed by  $n_{k(i)}$ . It is then not difficult to show that the conditional distribution of  $T_1$  converges to that of  $C_2 Z_1$   $P$ -almost-surely by replacing  $n, n_1$  and  $n_2$  by  $n_{k(i)}, n_{k(i),1}$  and  $n_{k(i),2}$  respectively, and essentially repeating the steps in Theorem 3.3.8.

It remains to show that  $T_2 \xrightarrow{P^\star} 0$  ( $P - a.s.$ ). The proof of this follows from that of Lemma 3.7.5 by replacing  $n, n_1$  and  $n_2$  by  $n_{k(i)}, n_{k(i),1}$  and  $n_{k(i),2}$  respectively, and noting that  $\hat{d}_{n_{k(i),1}}^{(1)} - d_0 = B_{as}(1/3)$ .  $\square$

### 3.7.4 Some Auxiliary Lemmas

First we state a special almost sure convergence result on a triangular array of iid mean zero random variables. For the general result, see Proposition in [Hu et al. \(1989\)](#).

*Lemma 3.7.6.* If a triangular array of random variables  $\{X_{ni}\}_{i=1}^{m_n}$  for  $n \in \mathbb{N}$  are iid copies of a mean 0 random variable  $X$  with  $m_n$  increases to  $\infty$  as  $n$  goes to  $\infty$  and  $\mathbb{E}|X|^{2p} < \infty$  for some  $p \in [1, 2)$ ,  $P\left(\lim_{n \rightarrow \infty} m_n^{-1/p} \sum_{i=1}^{m_n} X_{ni} = 0\right) = 1$ .

Suppose a triangular array of random variables  $\{\epsilon_{ni}\}_{i=1}^{m_n}$  for  $n \in \mathbb{N}$  are iid copies of  $\epsilon$  with mean 0, where  $m_n$  increases to  $\infty$  as  $n$  goes to  $\infty$ . Then Lemma 3.7.6 tells that  $\bar{\epsilon}_n = (1/m_n) \sum_{i=1}^{m_n} \epsilon_{ni}$  and  $(1/m_n) \sum_{i=1}^{m_n} \epsilon_{ni}^2$  converge to 0 and  $\sigma^2$  almost surely given  $\mathbb{E}\epsilon^2 < \infty$  and  $\mathbb{E}\epsilon^4 < \infty$ , respectively. Further, the following lemma shows that  $n^{1/2}$  is an upper boundary rate of the almost sure convergence of  $\bar{\epsilon}_n$ .

*Lemma 3.7.7.* If  $\mathbb{E}\epsilon^4 < \infty$ ,  $P(\lim_{n \rightarrow \infty} m_n^\alpha \bar{\epsilon}_n = 0) = 1$  for each  $\alpha < 1/2$ .

*Proof.* A direct application of Lemma 3.7.6 gives that if  $\mathbb{E}|\epsilon|^{2p} < \infty$  for some  $p \in [1, 2)$ ,  $P\left(\lim_{n \rightarrow \infty} m_n^{1-1/p} \bar{\epsilon}_n = 0\right) = 1$ . On the other hand,  $\mathbb{E}\epsilon^4 < \infty$  implies that  $\mathbb{E}|\epsilon|^{2p} < \infty$  for every  $p \in [1, 2)$ . Thus, the conclusion follows.  $\square$

Suppose  $\{\epsilon'_i\}_{i=1}^n, \{\epsilon''_i\}_{i=1}^n, \{\epsilon_i^{\prime\star}\}_{i=1}^n$  and  $\{\epsilon_i^{\prime\prime\star}\}_{i=1}^n$  are the second-stage random errors and the corresponding bootstrapped ones defined in Subsection 3.3.2. Note that the subscripts of these random variables indicating the sample size are suppressed for the simplicity of notation and that here “ $n$ ” is understood as a dummy variable, not the total sample size. Recall  $V_i^+ = \epsilon_i^{\prime\prime\star} + \epsilon_i^{\prime\star}$ ,  $\nu^+ = E_\star[V_i^+]$ ,  $V_i^- = \epsilon_i^{\prime\prime\star} - \epsilon_i^{\prime\star}$  and  $\nu^- = E_\star[V_i^-]$ , where  $E_\star$  means the expectation conditioning on the second-stage data.

Since  $\text{Var}_\star[V_i^+] = \text{Var}_\star[V_i^-]$ , we denote both as  $s^2$ . The following lemma shows that both  $V_i^+$  and  $V_i^-$  are asymptotically normal  $P$ -almost surely.

*Lemma 3.7.8.* If  $\mathbb{E}\epsilon^6 < \infty$ , we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{V_i^+ - \nu^+}{s} \xrightarrow{d^\star} Z, \quad (P - a.s.), \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{V_i^- - \nu^-}{s} \xrightarrow{d^\star} Z, \quad (P - a.s.),$$

where  $Z$  follows a  $N(0, 1)$  distribution.

*Proof.* We only prove the former and the latter can be shown similarly. Let  $\xi_{ni} = (V_i^+ - \nu^+)/(\sqrt{ns})$ , for  $i = 1, 2, \dots, n$ , and  $S_n = \sum_{i=1}^n \xi_{ni}$ . It is easy to see that  $E_\star[\xi_{ni}] = 0$  and  $\text{Var}_\star[S_n] = 1$ . Thus, it suffices to check that the following Lindeberg condition holds for each  $\eta > 0$  (see, for example, Theorem2 on Page 334 of [Shiryayev \(1995\)](#)):  $\sum_i^n E_\star[\xi_{ni}^2 \{|\xi_{ni}| \geq \eta\}] \rightarrow 0$ , ( $P - a.s.$ ). Note that

$$\begin{aligned} \sum_i^n E_\star[\xi_{ni}^2 \{|\xi_{ni}| \geq \eta\}] &= E_\star \left[ \left[ (V_1^+ - \nu^+)/s \right]^2 \{ |(V_1^+ - \nu^+)/s| \geq \sqrt{n}\eta \} \right] \\ &\leq (\sqrt{n}\eta)^{-1} |s|^{-3} E_\star |V_1^+ - \nu^+|^3, \end{aligned}$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (\epsilon_i'')^2 - \left( \frac{1}{n} \sum_{i=1}^n \epsilon_i'' \right)^2 + \frac{1}{n} \sum_{i=1}^n (\epsilon_i')^2 - \left( \frac{1}{n} \sum_{i=1}^n \epsilon_i' \right)^2 \rightarrow 2\sigma^2, \quad (P - a.s.),$$

then it is sufficient to show  $\overline{\lim}_{n \rightarrow \infty} E_\star |V_1^+ - \nu^+|^3 < \infty$ , ( $P - a.s.$ ). Since

$$\begin{aligned} E_\star |V_1^+ - \nu^+|^3 &\leq E_\star [ |V_1^+|^3 + |\nu^+|^3 + 3|V_1^+|^2 |\nu^+| + 3|V_1^+| |\nu^+|^2 ] \\ &= E_\star |V_1^+|^3 + 3|\nu^+| E_\star |V_1^+|^2 + 3|\nu^+|^2 E_\star |V_1^+| + |\nu^+|^3, \end{aligned}$$

and  $\nu^+ = \frac{1}{n} \sum_{i=1}^{n_2} (\epsilon_i'' + \epsilon_i') \rightarrow 0$ , ( $P - a.s.$ ), it suffices to show  $\overline{\lim}_{n \rightarrow \infty} E_\star |V_1^+|^k < \infty$ , ( $P - a.s.$ ), for  $k = 1, 2, 3$ . We only need to show the case where  $k = 3$ . From  $(a + b)^3 \leq 4(a^3 + b^3)$  for nonnegative  $a$  and  $b$ ,

$$E_\star |V_1^+|^3 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |\epsilon_i'' + \epsilon_j'|^3 \leq 4 \left( \frac{1}{n} \sum_{i=1}^n |\epsilon_i''|^3 + \frac{1}{n} \sum_{i=1}^n |\epsilon_i'|^3 \right).$$

By Lemma 3.7.6, both  $(1/n) \sum_{i=1}^n |\epsilon_i''|^3$  and  $(1/n) \sum_{i=1}^n |\epsilon_i'|^3$  converges almost surely under the assumption  $\mathbb{E}\epsilon^6 < \infty$ . Therefore,  $\overline{\lim}_{n \rightarrow \infty} E_\star |V_1^+|^3 < \infty$ , ( $P - a.s.$ ), which completes the proof.  $\square$



## CHAPTER IV

# Two Stage Nonparametric Procedures for Estimating A Threshold Value of A Regressor

### 4.1 Introduction

The basic problem in this chapter is the same as that in the previous chapter. For completeness, we restate it below with slightly different notations. Consider the regression model

$$Y = m(X) + \epsilon, \tag{4.1}$$

where the design point  $X$  takes values in  $[a, b]$ , the regression function  $m$  is non-decreasing and the random error  $\epsilon$  has mean 0 and finite variance  $\sigma^2$ . Suppose there exist  $d_0 \in (a, b)$  and  $\theta_0 \in (m(a), m(b))$  such that  $d_0 = m^{-1}(\theta_0) = \inf\{x \in (a, b) : m(x) \geq \theta_0\}$ . It is of interest to estimate  $d_0$  given  $\theta_0$  in a design setting where design points can be chosen.

Since we have a design setting under consideration, we can first spend part of the sample points to identify a small interval around  $d_0$  and then allocate the remaining sample points within the interval to estimate  $m$  locally. This two-stage idea has been implemented in the previous chapter, where we propose the hybrid two-stage procedure with isotonic regression at stage one and local linear approximation for  $m$  at stage two to accelerate the convergence rate of the resulting estimator to the parametric rate  $\sqrt{n}$ . To fully exploit the advantage of the local linear approximation, the second-stage design points are evenly allocated at the two ends of the small interval around  $d_0$ .

If  $m$  is locally linear in a neighborhood of  $d_0$ , the local linear approximation approach usually works well in practice. However, if  $m$  is locally very nonlinear around

$d_0$  and at the same time the sample size is not large enough, the local linear fitting over a relatively long second-stage sampling interval would bring significant bias into the estimate of  $d_0$ , which would further cause low coverage rates for confidence intervals.

To remedy this situation, our idea is to estimate  $m$  at the second stage purely non-parametrically. More specifically, at stage two, we allocate the second-stage design points within the second-stage sampling interval, for example evenly, and then re-estimate  $m$  over the interval by isotonic regression. Although this two-stage isotonic regression procedure (TSIRP) produces an estimator for  $d_0$  with a slower convergence rate than the parametric one  $\sqrt{n}$ , it usually handles well the case where  $m$  is locally quite nonlinear at  $d_0$  in practice. This TSIRP can be extended to a  $K$ -stage isotonic regression procedure (KSIRP) by repeating the zoom-in step in TSIRP. The convergence rate of the estimator for  $d_0$  from KSIRP strictly ascends to the parametric rate as the number of stages  $K$  increases to infinity. This discovery reveals an essential difference in terms of convergence rates between parametric methods and a special nonparametric method: isotonic regression. Naturally, this discovery drives us to ask a further question: can a  $K$ -stage purely nonparametric procedure achieve or even exceed the parametric rate if a nonparametric method with a one-stage convergence rate faster than  $n^{1/3}$ , the one-stage convergence rate for isotonic regression, is utilized at each stage? We have considered smoothed isotonic regression, a combination of isotonic regression and kernel smoothing, whose one-stage convergence rate can reach  $n^{2/5}$ . Unfortunately, it turns out that such a  $K$ -stage smoothed isotonic regression procedure can not exceed the parametric rate. This, to some extent, indicates that the difference between parametric approaches and nonparametric ones is so fundamental that continuously improving data quality can not make nonparametric approaches outperform parametric ones. However, a heuristic derivation shows that a two-stage smoothed isotonic regression procedure (TSSIRP) can already achieve the parametric rate  $\sqrt{n}$ , just like the hybrid two-stage procedure.

We also introduce variants of practical TSIRP and TSSIRP. A simulation study shows that practical TSSIRP usually produces somewhat better confidence intervals for  $d_0$  than practical TSIRP when regression functions are well behaved. However, practical TSIRP continues to exhibit adequate performance with irregular regression functions while practical TSSIRP suffers significantly. Therefore, practical TSIRP produces a stable choice that, by and large, circumvents the need for model checking, which is of practical interest.

We introduce preliminary one-stage isotonic regression procedure and one-stage smoothed isotonic regression procedure in Section 4.2. In Section 4.3, we propose the two-stage isotonic regression and smoothed isotonic regression procedures. Practical variants of the one-stage and two-stage procedures are introduced in Section 4.4 and a simulation study is carried out in Section 4.5. After a real data analysis in Section 4.6, Section 4.7 concludes with discussion and the Appendix 4.8 contains technical proofs.

## 4.2 Preliminaries

We begin with an introduction of a one-stage isotonic regression plan (OSIRP) and a one-stage smoothed isotonic regression plan (OSSIRP) for estimating  $d_0$ . Although OSIRP has been described in the previous chapter, to make this chapter self-contained, we next introduce it again.

### 4.2.1 One-Stage Isotonic Regression Procedure (OSIRP)

We first consider a one-stage isotonic regression procedure (OSIRP) for estimating  $d_0$  and constructing confidence intervals for it.

Suppose the sample size is  $n$  and the fixed design points  $\{X_i\}_{i=1}^n$  follow a distribution  $G$  defined on  $[a, b]$ . This means that there exists a distribution  $G$  such that  $\sup_{x \in [a, b]} |F_n(x) - G(x)| = o(n^{-1/3})$ , where  $F_n$  is the empirical distribution of the design points. We adopt this fixed design in this work because it is usually more popular than a random design in practice. The case with a random design can be similarly handled. For simplicity of notation, we assume that the design points  $X_i$ 's have already been increasingly sorted. After obtaining the corresponding responses  $\{Y_i\}_{i=1}^n$ , we have the one-stage data  $\{(X_i, Y_i)\}_{i=1}^n$ .

To rigorously state the theoretical results, the following assumptions are needed:

- (A1) The derivative of the regression function  $m'$  is continuous in a neighborhood of  $d_0$  and  $m'(d_0) > 0$
- (A2) The Lebesgue density of  $G$ , denoted by  $g$ , is continuous and  $g(d_0) > 0$ .

We denote the above two assumptions together by **(A)**.

The isotonic regression of  $m$  at  $X_i$ 's is given by

$$\{m_i^*\}_{i=1}^n = \underset{m_1 \leq m_2 \leq \dots \leq m_n}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - m_i)^2. \quad (4.2)$$

The minimizer  $\{m_i^*\}$  exists uniquely. (For example, see Chapter 1 of [Robertson et al. \(1988\)](#).) We define the isotonic regression of  $m$  as the following right continuous function

$$m_I(x) = \begin{cases} m_1^* & \text{if } x \in [a, X_1], \\ m_i^* & \text{if } x \in [X_i, X_{i+1}), i = 1, 2, \dots, n-1, \\ m_n^* & \text{if } x \in [X_n, b]. \end{cases} \quad (4.3)$$

Then, for  $\theta_0 \in (m(a), m(b))$ , the one-stage isotonic regression estimator of  $d_0$  is given by

$$d_I = m_I^{-1}(\theta_0) = \inf\{x \in [a, b] : m_I(x) \geq \theta_0\}, \quad (4.4)$$

where  $\inf\{\emptyset\} = b$ . Its asymptotic distribution is given in the following Theorem.

*Theorem 4.2.1.* Under assumption **(A)**, we have

$$n^{1/3}(d_I - d_0) \xrightarrow{d} C_{d_I} g(d_0)^{-1/3} \mathcal{Z}, \quad (4.5)$$

where  $C_{d_I} = (4\sigma^2/m'(d_0)^2)^{1/3}$  and  $\mathcal{Z}$  follows the standard Chernoff distribution.

*Remark 4.2.2.* The standard Chernoff distribution is the distribution of

$$\mathcal{Z} = \underset{t \in \mathbb{R}}{\operatorname{argmin}} (W(t) + t^2),$$

where  $\{W(t), t \in \mathbb{R}\}$  is a two-sided Brownian Motion with  $W(0) = 0$ . Accurate numerical quantiles are available in [Groeneboom and Wellner \(2001\)](#).

The proof of [Theorem 4.2.1](#) follows by a minor adaptation of the arguments from [Theorem 1](#) in [Banerjee and Wellner \(2005\)](#). A sketch of this proof is provided in [Appendix 4.8.1](#). From [Theorem 4.2.1](#) and the symmetry of the Chernoff distribution, the theoretical Wald-type  $1 - \alpha$  asymptotic confidence interval of  $d_0$  is given by

$$[d_I \pm n^{-1/3} C_{d_I} g(d_0)^{-1/3} q(\mathcal{Z}, 1 - \alpha/2)], \quad (4.6)$$

where  $q(\xi, p)$  is the  $p$ th quantile of a random variable  $\xi$ .

Next, consider log-likelihood ratio type (LR-type) confidence intervals for  $d_0$ . The pair of hypotheses of interest are

$$H_0 : m^{-1}(\theta_0) = x_0 \leftrightarrow H_a : m^{-1}(\theta_0) \neq x_0. \quad (4.7)$$

Then, the LR-type test statistic is given by

$$2 \log \lambda_I = 2 \log \lambda_I(x_0) = 2 [l_n(m_I) - l_n(m_{Ic})], \quad (4.8)$$

where

$$l_n(m) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - m(X_i))^2, \quad (4.9)$$

and  $m_{Ic}$  is the constrained isotonic regression of  $m$  under the above null hypothesis. It is known that  $m_{Ic}$  uniquely exists (For example, see [Banerjee \(2000\)](#)). The asymptotic distribution of  $2 \log \lambda_I$  is given in the following Theorem.

*Theorem 4.2.3.* Under assumption **(A)** and the null hypothesis  $H_0 : m^{-1}(\theta_0) = x_0$ , we have

$$2 \log \lambda_I \xrightarrow{d} \mathbb{D},$$

where  $\mathbb{D}$  is a pivot random variable.

*Remark 4.2.4.* The theoretical characterization of  $\mathbb{D}$  is derived in [Banerjee and Wellner \(2001\)](#).

From [Theorem 4.2.3](#), the LR-type  $(1 - \alpha)$  asymptotic confidence region of  $d_0$  is given by

$$\{x \in [a, b] : 2 \log \lambda_I(x) \leq q(\mathbb{D}, 1 - \alpha)\}. \quad (4.10)$$

This LR-type confidence region is an interval. It is usually asymmetric around  $d_I$ , unlike the Wald-type confidence interval [\(4.6\)](#). Note that in the Wald-type confidence interval [\(4.6\)](#), both  $m'(d_0)$  and  $\sigma$  are unknown. On the other hand, in the LR-type confidence interval [\(4.10\)](#), only  $\sigma$  is unknown. Since it is usually difficult to estimate  $m'(d_0)$  well without a large sample, the LR-type confidence interval offers practical advantages.

Thus, there are two kinds confidence intervals for  $d_0$ , the Wald-type and LR-type ones. The corresponding plans are denoted by OSIRP-Wald and OSIRP-LR, which are briefly summarized as follows.

- Procedure for OSIRP-Wald:
  1. Generate the fixed design points  $\{X_i\}_{i=1}^n$  and obtain the corresponding responses  $\{Y_i\}_{i=1}^n$ .
  2. Compute the one-stage isotonic regression  $m_I$  of  $m$  with the one-stage data  $\{(X_i, Y_i)\}$ .
  3. Calculate the one-stage isotonic regression estimator  $d_I$  of  $d_0$  from  $m_I$  and  $\theta_0$ .
  4. Construct Wald-type confidence intervals by (4.6).
- Procedure for OSIRP-LR:
  1. Do the first two steps in OSIRP-Wald.
  2. Compute the constrained isotonic regression  $m_{Ic}$  of  $m$  with a given  $x_0 \in (a, b)$ .
  3. Construct LR-type confidence intervals by (4.10).

#### 4.2.2 One-Stage Smoothed Isotonic Regression Procedure (OSSIRP)

From Theorem 4.2.1, we see that the convergence rate of the isotonic regression estimator  $d_I$  of  $d_0$  in OSIRP is only  $n^{1/3}$ . In order to increase the convergence rate, we next consider a one-stage smoothed isotonic regression procedure (OSSIRP).

Suppose we have the same one-step data  $\{(X_i, Y_i)\}$  as in OSIRP. After obtaining the one-stage isotonic regression  $m_I$  of  $m$ , we compute the smoothed isotonic regression  $m_{Is}$  with a kernel  $K$  and a bandwidth  $h_n$ . More specifically, the smoothed isotonic regression is given by

$$m_{Is}(x) = \frac{\sum_{i=1}^n W_i(x) m_I(X_i)}{\sum_{i=1}^n W_i(x)}, \quad \text{for } x \in [a, b], \quad (4.11)$$

where  $W_i(x) = K((x - X_i)/h_n)$ . When  $K$  is log concave,  $m_{Is}$  is increasing. (See, for example, Remark 2.1 in Mukerjee (1988).) Then, the one-stage smoothed isotonic regression estimator of  $d_0$  is  $d_{Is} = m_{Is}^{-1}(\theta_0)$ . For further analysis, we formulate  $h_n$  as  $cn^{-\gamma}$  with  $c > 0$  and  $\gamma > 0$  and make the following assumptions:

**(B1)** The regression function  $m$  is three-times continuously differentiable and  $m'(d_0) > 0$ .

**(B2)** The design density  $g$  is three-times continuously differentiable and  $g(d_0) > 0$ .

**(B3)** The kernel function  $K$  is a symmetric density function around 0 defined on a compact interval and has log-concavity.

We denote these three assumptions together as **(B)**. Then, the following theorem gives the asymptotic distributions of  $d_{I_s}$ .

*Theorem 4.2.5.* Suppose assumption **(B)** holds. For  $\gamma = 1/5$ , we have

$$n^{2/5} (d_{I_s} - d_0) \xrightarrow{d} B_{d_{I_s}} + C_{d_{I_s}} N(0, 1); \quad (4.12)$$

whereas for  $\gamma \in (1/5, 1/3)$ , we have

$$n^{(1-\gamma)/2} (d_{I_s} - d_0) \xrightarrow{d} C_{d_{I_s}} N(0, 1), \quad (4.13)$$

where

$$B_{d_{I_s}} = -m'(d_0)^{-1} c^2 B(d_0), \quad C_{d_{I_s}} = m'(d_0)^{-1} \sqrt{\kappa / (c g(d_0))} \sigma, \quad \kappa = \int K^2(t) dt,$$

and

$$B(x) = (2g'(x)m'(x) + g(x)m''(x)) / (2g(x)) \int t^2 K(t) dt.$$

A sketch of the proof of Theorem 4.2.5 is provided in Appendix 4.8.1.

*Remark 4.2.6.* The restriction  $\gamma < 1/3$  is due to the fact that the smoothing bandwidth  $h_n$  should be asymptotically larger than the automatic bandwidth of the isotonic regression, which is of order  $n^{-1/3}$ , to make the smoothing step effective. For  $\gamma > 1/3$ , the smoothing step is ignorable and  $m_{I_s} = m_I$  holds asymptotically.

By (4.12) of Theorem 4.2.5, with the bandwidth  $h_n = cn^{-1/5}$ , the theoretical  $1 - \alpha$  Wald-type confidence interval of  $d_0$  is given by

$$\left[ d_{I_s} - n^{-2/5} B_{d_{I_s}} \pm n^{-2/5} C_{d_{I_s}} z_{\alpha/2} \right], \quad (4.14)$$

where  $z_{\alpha/2}$  is the upper  $\alpha/2$  quantile of  $N(0, 1)$ . Similarly, by (4.13) of Theorem 4.2.5, with  $h_n = cn^{-\gamma}$  and  $\gamma \in (1/5, 1/3)$ , the theoretical  $1 - \alpha$  Wald-type confidence

interval of  $d_0$  is given by

$$[d_{I_s} \pm n^{-(1-\gamma)/2} C_{d_{I_s}} z_{\alpha/2}]. \quad (4.15)$$

Since the convergence rate of  $d_{I_s}$  for the case where  $h_n = cn^{-1/5}$  is faster, the former confidence interval (4.14) is asymptotically shorter.

A brief summary of OSSIRP is as follows:

- Procedure of OSSIRP
  1. Follow the first two steps in OSIRP-Wald.
  2. Compute the smoothed one-stage isotonic regression  $m_{I_s}$  of  $m$  with the one-stage data.
  3. Calculate the one-stage isotonic regression estimator  $d_{I_s}$  of  $d_0$ .
  4. Construct Wald-type confidence intervals by (4.14) or (4.15).

### 4.3 Two-Stage Procedures

Although the convergence rate of  $d_{I_s}$  can be  $n^{2/5}$ , faster than that of  $d_I$ , it is still slower than the parametric rate  $\sqrt{n}$ . Since we are in a design setting, we can further improve the convergence rate by collecting data in two stages.

Specifically, at stage one, an initial isotonic regression (IR) or smoothed isotonic regression (SIR) estimator of  $d_0$  is obtained by using a limited number of sample points. At stage two, all the remaining design points are allocated in a neighborhood of  $d_0$  to increase the sample resolution around  $d_0$  and the corresponding second-stage responses are obtained. Then, the convergence rates of the second-stage IR and SIR estimators of  $d_0$  are expected to increase since they both depend on the sample resolution around  $d_0$ .

There are four combinations of the two estimation methods IR and SIR in two stages. In this work, we mainly consider two of them, where IR is used in the first stage for its practical simplicity and without loss of theoretical generality. The other two combinations with SIR at the first stage can be similarly considered. We call the procedure where IR is also used at the second stage as the two-stage isotonic regression procedure (TSIRP) and the procedure where SIR is used at the second



stage as the two-stage smoothed isotonic regression procedure (TSSIRP). Next, we consider TSIRP and TSSIRP in detail.

### 4.3.1 Two-Stage Isotonic Regression Procedure $I + I$ (TSIRP)

We first consider the two-stage isotonic regression procedure (TSIRP). Suppose the first-stage sample proportion is  $p \in (0, 1)$  and denote the first and second-stage sample sizes by  $n_1 = \lfloor np \rfloor$  and  $n_2 = n - n_1$ , respectively. In the first stage, the sample  $\{(X_{1,i}, Y_{1,i})\}_{i=1}^{n_1}$  with a design density  $g_1$  on  $[a, b]$  is used with OSIRP to obtain the first-stage IR estimator  $d_{1,I}$  of  $d_0$ . Then, a neighborhood of  $d_0$  for the second-stage sampling is specified as  $[L_1, U_1] = [d_{1,I} \pm C_1 n_1^{-\gamma_1}] \cap [a, b]$  with  $C_1 > 0$  and  $0 < \gamma_1 < 1/3$ . Note that  $\gamma_1$  being less than  $1/3$  ensures that  $[L_1, U_1]$  contains  $d_0$  with probability going to one. Next, the remaining design points  $\{X_{2,i}\}_{i=1}^{n_2}$  are allocated in the second-stage sampling interval  $[L_1, U_1]$  with a Lebesgue design density  $g_2$  and the corresponding second-stage responses  $\{Y_{2,i}\}_{i=1}^{n_2}$  are obtained. The density  $g_2$  on  $[L_1, U_1]$  can be defined as  $g_2(x) = (C_1 n_1^{-\gamma_1})^{-1} \psi((x - d_{1,I}) / (C_1 n_1^{-\gamma_1}))$  with  $\psi$  being a positive and continuous Lebesgue density on  $[-1, 1]$ . This assumption on  $\psi$  is denoted by **(A3)** and included in the assumption set **(A)** for TSIRP. From the second-stage data, an upgraded estimator  $m_{2,I}$  of  $m$  is computed over  $[L_1, U_1]$ . The second-stage IR estimator of  $d_0$  is  $d_{2,I} = m_{2,I}^{-1}(\theta_0)$  and its asymptotic distribution is given in the following proposition.

*Proposition 4.3.1.* Under assumption **(A)**, we have

$$n^{(1+\gamma_1)/3} (d_{2,I} - d_0) \xrightarrow{d} C_{d_{2,I}} \mathcal{Z},$$

where

$$C_{d_{2,I}} = C_{d_I} \left( \frac{C_1}{(1-p)p^{\gamma_1}\psi(0)} \right)^{1/3}.$$

A sketch of the proof of Proposition 4.3.1 is given in Appendix 4.8.3.

From Proposition 4.3.1, the theoretical Wald-type  $1 - \alpha$  asymptotic confidence interval of  $d_0$  is given by

$$[d_{2,I} \pm n^{-(1+\gamma_1)/3} C_{d_{2,I}} q(\mathcal{Z}, 1 - \alpha/2)]. \quad (4.16)$$

From Proposition 4.3.1 and Theorem 4.2.1, both  $d_{2,I}$  and  $d_I$  are consistent estimators of  $d_0$ . The rate of convergence of  $d_{2,I}$  is  $n^{(1+\gamma_1)/3}$ , faster than that of  $d_I$ , which

is  $n^{1/3}$ . In fact the asymptotic relative efficiency (ARE) of  $d_{2,I}$  with respect to  $d_I$  is

$$ARE(d_{2,I}, d_I) = \frac{s.d.(d_I)}{s.d.(d_{2,I})} = \left( \frac{(1-p)p^\gamma \psi(0)}{C_1 g(0)} \right)^{1/3} n^{\gamma/3} \rightarrow \infty \text{ as } n \rightarrow \infty.$$

Thus,  $d_{2,I}$  is a superior estimator for  $d_0$  than  $d_I$ , in terms of asymptotic mean squared error.

*Remark 4.3.2.* One intuition behind the result in Proposition 4.3.1 is as follows. From Theorem 4.2.1, we have

$$(g(d_0)n)^{1/3}(d_{1,I} - d_0) \xrightarrow{d} C_{d_I} \mathcal{Z}, \quad (4.17)$$

where  $g(d_0)n$  can be viewed as the sample resolution at  $d_0$ . On the other hand, since the second-stage sampling interval  $(L_1, U_1)$  contains  $d_0$  with probability increasing to one as the sample size goes to infinity, we can roughly think that  $d_0 \in (L_1, U_1)$ . Thus, the second-stage sample resolution around  $d_0$  becomes  $g_2(d_0)n_2$  in TSIRP. Since isotonic regression is essentially a kind of local average of the responses, according to Theorem 4.2.1, we have

$$(g_2(d_0)n_2)^{1/3}(d_{2,I} - d_0) \xrightarrow{d} C_{d_I} \mathcal{Z}, \quad (4.18)$$

which leads to Proposition 4.3.1 by noting that  $\psi((d_0 - d_{1,I})/(C_1 n_1^{-\gamma_1})) \xrightarrow{P} \psi(0)$  for  $\gamma \in (0, 1/3)$ .

Similarly to OSIRP, we also consider the second-stage LR-type confidence interval of  $d_0$  with the hypotheses (4.7). After obtaining the second-stage data  $\{(X_i, Y_i)\}$  and the second-stage isotonic regression  $m_{2,I}$ , we compute the constrained second-stage isotonic regression,  $m_{2,Ic}$ , under the null hypothesis  $H_0$ . Then, the LR-type test statistic is given by

$$2 \log \lambda_{2,I} = 2 \log \lambda_{2,I}(x_0) = 2 [l_n(m_{2,I}) - l_n(m_{2,Ic})], \quad (4.19)$$

where

$$l_n(m) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n_2} (Y_{2,i} - m(X_{2,i}))^2. \quad (4.20)$$

and its asymptotic distribution is given in the following Proposition.

*Proposition 4.3.3.* Under assumption **(A)** and the null hypothesis  $H_0: m^{-1}(\theta_0) = x_0$ ,

$$2 \log \lambda_{2,I} \xrightarrow{d} \mathbb{D}.$$

From Proposition 4.3.3, the LR-type  $(1 - \alpha)$  asymptotic confidence interval of  $d_0$  is given by

$$\{x \in [a, b] : 2 \log \lambda_{2,I}(x) \leq q(\mathbb{D}, 1 - \alpha)\}. \quad (4.21)$$

Similar to OSIRP, there are two kinds of confidence intervals for  $d_0$  in TSIRP, the Wald-type and LR-type ones. The corresponding plans are denoted by TSIRP-Wald and TSIRP-LR, which are briefly described next:

- Procedures of TSIRP-Wald and TSIRP-LR:
  1. Implement OSIRP with the first-stage sample of size  $n_1$  and obtain an initial estimator  $d_{1,I}$  of  $d_0$ .
  2. Specify the second-stage sampling interval  $[L_1, U_1]$  and generate the second-stage data.
  3. Compute the unconstrained second-stage isotonic regression  $m_{2,I}$  (and the constrained one  $m_{2,Ic}$ ) of  $m$  from the second-stage data.
  4. Construct the Wald-type confidence intervals (4.16) (or the LR-type ones (4.21)).

#### 4.3.1.1 Using Two Stage Data

In the previous TSIRP, only the second-stage data are used to estimate  $d_0$  and construct confidence intervals after a small neighborhood around  $d_0$  is identified. In fact, after generating the second-stage data, it is more practical to first pool the two-stage data together and then compute the two-stage or pooled estimator of  $d_0$  and the LR-type test statistic. Denote them as  $d_{I,pooled}$  and  $2 \log \hat{\lambda}_{I,pooled}$ , respectively.

Asymptotically speaking, since both  $d_{I,pooled}$  and  $2 \log \lambda_{I,pooled}$  depend on the sample resolution around  $d_0$ , which is determined by the second-stage data, they are equivalent to  $d_{2,I}$  and  $2 \log \lambda_{2,I}$ . More specifically, we have the following two propositions.

*Proposition 4.3.4.* Under assumption **(A)**, we have

$$n^{(1+\gamma_1)/3}(d_{I,pooled} - d_0) \xrightarrow{d} C_{d_{I,pooled}} \mathcal{Z},$$

where  $C_{d_{I,pooled}} = C_{d_{2,I}}$ .

*Proposition 4.3.5.* Under assumption **(A)** and the null hypothesis  $H_0 : m^{-1}(\theta_0) = d_0$ ,

$$2 \log \lambda_{I,pooled} \xrightarrow{d} \mathbb{D}.$$

From Propositions 4.3.4 and 4.3.5, the theoretical Wald-type and LR-type  $1 - \alpha$  asymptotic confidence intervals for  $d_0$  are given by

$$[d_{I,pooled} \pm n^{-(1+\gamma_1)/3} C_{d_{I,pooled}} q(\mathcal{Z}, 1 - \alpha/2)], \quad (4.22)$$

and

$$\{x \in [a, b] : 2 \log \lambda_{I,pooled}(x) \leq q(\mathbb{D}, 1 - \alpha/2)\}. \quad (4.23)$$

Although  $d_{I,pooled}$  and  $2 \log \lambda_{I,pooled}$  are asymptotically equivalent to  $d_{2,I}$  and  $2 \log \lambda_{2,I}$ , they usually have better finite sample properties since they use more sample points.

*Remark 4.3.6.* Our intuition for the result in Proposition 4.3.4 is similar to that in Remark 4.3.2. When two-stage data are used, the sample resolution at  $d_0$  becomes  $g_1(d_0)n_1 + g_2(d_0)n_2$ . Thus, similar to (4.18), we get

$$(g_1(d_0)n_1 + g_2(d_0)n_2)^{1/3}(d_{I,pooled} - d_0) \xrightarrow{d} C_{d_I} \mathcal{Z},$$

which leads to Proposition 4.3.4 by noting  $g_1(d_0)n_1 = o(g_2(d_0)n_2)$ .

#### 4.3.1.2 Multistage Isotonic Regression Procedure

It is straightforward to extend TSIRP to a multistage isotonic regression procedure. Specifically, the  $M$ -stage procedure with  $M \geq 2$  is given by:

1. Specify the  $i$ th stage sample proportion  $p_i > 0$  such that  $\sum_{i=1}^M p_i = 1$ . Then, define the  $i$ th stage sample size  $n_i$  accordingly.
2. Generate the  $i$ th stage data over  $[L_{i-1}, U_{i-1}]$  for  $1 \leq i \leq M - 1$  with  $[L_0, U_0] = [a, b]$ .

3. Obtain the  $i$ th-stage isotonic regression estimator  $d_{i,I}$  of  $d_0$  with the  $i$ th-stage data.
4. Specify the  $(i + 1)$ th-stage sampling interval  $[L_i, U_i] = [d_{i,I} \pm C_i n_i^{-\gamma_i}]$ , where  $C_i > 0$ ,  $\gamma_i \in (0, r_i)$  and  $n_i^{r_i}$  is the convergence rate of  $d_{i,I}$ .
5. Repeat the above steps until  $i = M$ . Then compute the  $M$ th-stage isotonic regression estimator  $d_{M,I}$  of  $d_0$ .

We assume that assumption **(A1)** holds in the  $M$ -stage procedure. Then, one characterization of the convergence rate of  $d_{M,I}$  is given in the following proposition.

*Proposition 4.3.7.* The convergence rate of  $d_{M,I}$  is always less than  $1/2$ . However, it increasingly converges to  $1/2$  as  $M$  goes to infinity.

*Proof.* The convergence rate  $r_1$  of  $d_{1,I}$  is  $\frac{1}{3}$ , so the order of the length of the second-stage sampling interval  $\gamma_1$  is in  $(0, \frac{1}{3})$ . The convergence rate  $r_2$  of  $d_{2,I}$  is  $\frac{1}{3}(1 + \gamma_1)$ , so the order of the length of the third-stage sampling interval  $\gamma_2$  is in  $(0, \frac{1}{3}(1 + \gamma_1))$ . In this way, we can see that the convergence rate  $r_k$  of  $d_{k,I}$  is  $\frac{1}{3}(1 + \gamma_{k-1})$ , and  $\gamma_k$  is in  $(0, \frac{1}{3}(1 + \gamma_{k-1}))$ , for any  $k \geq 2$ . Now since  $\gamma_1 < \frac{1}{2}$ , we have  $\gamma_2 < r_2 = \frac{1}{3}(1 + \gamma_1) < \frac{1}{2}$ . In this way, we have  $r_k = \frac{1}{3}(1 + \gamma_{k-1}) < \frac{1}{2}$  for any  $k \geq 2$ , which means the convergence rate of  $d_{M,I}$  is less than  $\frac{1}{2}$ .

On the other hand, for a small  $\epsilon > 0$ , in order to find a  $k$  such that  $r_{k+1} = \frac{1}{3}(1 + \gamma_k) > \frac{1}{2} - \epsilon$ , it is sufficient to have a  $\gamma_k$  such that  $\gamma_k > \frac{1}{2} - 3\epsilon$ . In order to find such a  $\gamma_k$ , it is sufficient to have a  $\gamma_{k-1}$  such that  $r_k = \frac{1}{3}(1 + \gamma_{k-1}) > \frac{1}{2} - 3\epsilon$ , which leads to  $\gamma_{k-1} > \frac{1}{2} - 9\epsilon$ . In this way, we can see that it is sufficient to find a  $\gamma_{k-l}$  such that  $\gamma_{k-l} > \frac{1}{2} - 3^{l+1}\epsilon$  for some integers  $k > l$ . Letting  $\frac{1}{2} - 3^{l+1}\epsilon < r_1 = 1/3$  gives  $l > -\log_3 \frac{6\epsilon}{3} - 1$ . This means if  $k > -\log_3 \frac{6\epsilon}{3}$ , we have  $r_{k+1} > \frac{1}{2} - \epsilon$ , which means that the convergence rate of  $d_{M,I}$  can be larger than  $1/2 - \epsilon$  with  $M$  large enough.  $\square$

Recall that  $r_k = (1/3)(1 + \gamma_{k-1})$  is the convergence rate of  $d_{k,I}$  with  $\gamma_0 = 0$  and that  $\gamma_k < r_k$  determines the order of the length of the  $(k + 1)$ th-stage sampling interval. Now suppose  $\gamma_k$  has the form  $\gamma_k = \alpha r_k$  where  $\alpha \in (0, 1)$  is a constant. Then, the asymptotic behavior of  $r_k$  is given in the following proposition.

*Proposition 4.3.8.* If  $\gamma_k = \alpha r_k = \alpha(1 + \gamma_{k-1})/3$  for  $k \in \mathbb{N}$  with  $\gamma_0 = 0$  and  $\alpha \in (0, 1)$ , then the convergence rate  $r_k$  converges to  $1/(3 - \alpha)$ , which is less than  $1/2$ .

*Proof.* Let  $b = \frac{\alpha/3}{1-\alpha/3}$ , then  $\gamma_k = \alpha \cdot \frac{1}{3}(1 + \gamma_{k-1})$  leads to  $(\gamma_k - b) = \frac{\alpha}{3}(\gamma_{k-1} - b)$ . Thus  $\gamma_k = \frac{\alpha}{3-\alpha} - (\frac{\alpha}{3})^k(\frac{\alpha}{3-\alpha}) \rightarrow \frac{\alpha}{3-\alpha}$  as  $k \rightarrow \infty$ . So the convergence rate  $r_k$  converges to  $\frac{1}{3}(1 + \frac{\alpha}{3-\alpha}) = \frac{1}{3-\alpha} < \frac{1}{2}$ .  $\square$

Suppose there exists another one stage regression estimator of  $d_0$  whose convergence rate is  $\eta > 0$ . If this regression method can be extended to a multistage plan like the isotonic regression, keeping the same pattern on the sequence of convergence rates, that is,  $r_k = \eta(1 + \gamma_{k-1})$  with  $k \geq 1$  and  $\gamma_0 = 0$ , then, one limit behavior is provided in the following Proposition.

*Proposition 4.3.9.* If  $\eta > 1/3$ , then there exists an interger  $M$  such that the convergence rate  $r_M$  of the estimator of the  $M$ -stage plan is greater than  $1/2$ .

*Proof.* It is sufficient to show that there exists a  $k \in \mathbb{N}$  such that  $\eta(1 + \gamma_k) > 1/2$ , which means  $\gamma_k > \frac{1}{2\eta} - 1$ . This is equivalent to  $\eta(1 + \gamma_{k-1}) > \frac{1}{2\eta} - 1$ , which means  $\gamma_{k-1} > \frac{1}{2\eta^2} - \frac{1}{\eta} - 1$ . In this way, we can see that a sufficient condition is that there exist  $1 \leq l < k$  such that

$$\gamma_{k-l} > \frac{1}{2\eta^{l+1}} - \frac{1}{\eta^l} - \dots - \frac{1}{\eta} - 1.$$

Letting the right side of the above inequality be less than  $\eta$  gives

$$\eta^{l+2} + \eta^{l+1} + \dots + \eta > 1/2.$$

Now since  $\eta > 1/3$ , we have

$$\lim_{l \rightarrow \infty} \sum_{i=1}^l \eta^i = \frac{\eta}{1-\eta} > \frac{1}{2},$$

which means there always exists a large  $M$  such that  $r_M > 1/2$ .  $\square$

*Remark 4.3.10.* Proposition 4.3.9 perhaps can be used to show the nonexistence of such a nonparametric regression method, whose one stage convergence rate is larger than  $1/3$  and which can be extended to a multistage plan without change of the pattern of convergence rates, if, on the other hand, we can prove that all nonparametric regressions methods can at most achieve the parametric rate  $\sqrt{n}$ .

### 4.3.2 Two-Stage Smoothed Isotonic Regression Procedure $I+Is$ (TSSIRP)

From Theorem 4.3.1, we know that the convergence rate of the isotonic regression estimator  $d_{2,I}$  of  $d_0$  in TSIRP is  $n^{(1+\gamma_1)/3}$ . In order to further increase the convergence rate, we next consider the two-stage smoothed isotonic regression procedure (TSSIRP). For this part, we assume the second-stage sampling density is uniform without loss of generality.

In TSIRP, after generating the second-stage data with the uniform density, we estimate  $m$  again by the second-stage isotonic regression  $m_{2,I}$  over the second-stage sampling interval  $[L_1, U_1]$  and then obtain the second-stage estimator  $d_{2,I}$  of  $d_0$  by taking the inverse of  $m_{2,I}$  at  $\theta_0$ . Since smoothed isotonic regression has a faster convergence rate than isotonic regression, at the second stage of TSSIRP, we smooth  $m_{2,I}$  with a kernel  $K$  and a bandwidth  $h_n$ , obtain the second-stage smoothed isotonic regression  $m_{2,Is}$  of  $m$  over  $[L_1, U_1]$ , and then estimate  $d_0$  by the second-stage smoothed isotonic regression estimator  $d_{2,Is} = m_{2,Is}^{-1}(\theta_0)$ .

Since the second-stage sampling density is uniform over  $[L_1, U_1]$ , the second-stage sample resolution around  $d_0$  is given by

$$n_r = \frac{n_2}{U_1 - L_1} = C_{n_r} n^{1+\gamma_1},$$

where  $\gamma_1 \in (0, 1/3)$  and  $C_{n_r} = p^{\gamma_1} q / (2C_1)$ . Note that  $C_{n_r}$  just means the coefficient of  $n_r$  and does not depend on  $n$ . To do further analysis, suppose the bandwidth for the kernel smoothing is of the form  $h_n = cn_r^{-\gamma_2}$  with constants  $c > 0$  and  $\gamma_2 > 0$ .

Asymptotically, there are two internal constraints for  $m_{2,Is}$  to be an effective estimator. Firstly, since kernel smoothing is applied over a shrinking interval, the effective bandwidth should be less than half of the length of the second-stage sampling interval, i.e.  $h_n < C_1 n_1^{-\gamma_1}$ . This inequality holds when  $\gamma_2 > \gamma_1 / (1 + \gamma_1)$ , or, when  $\gamma_2 = \gamma_1 / (1 + \gamma_1)$  and  $c < c^* = c^*(\gamma_2) = 2^{-\gamma_2} C_1^{1-\gamma_2} [(1-p)/p]^{\gamma_2}$ . Secondly, in order to ensure that  $m_{2,Is} = m_{2,s}$  with probability going to one, the bandwidth needs to be larger than the automatic bandwidth of isotonic regression, i.e.  $n^{-(1+\gamma_1)\gamma_2} > n^{-(1+\gamma_1)/3}$ , which is equivalent to  $\gamma_2 < 1/3$ .

Besides the above two internal requirements for  $\gamma_1$  and  $\gamma_2$ , there is another technical one  $\gamma_2 > 1/7$ , which is used to avoid complicated bias terms in the asymptotic results of  $d_{2,Is}$ . (See more details in Theorem 4.8.10 in the Appendix 4.8.4.) Thus, if  $(\gamma_1, \gamma_2)$  with a proper  $c$  lies in the region enclosed by the black dashed borders in

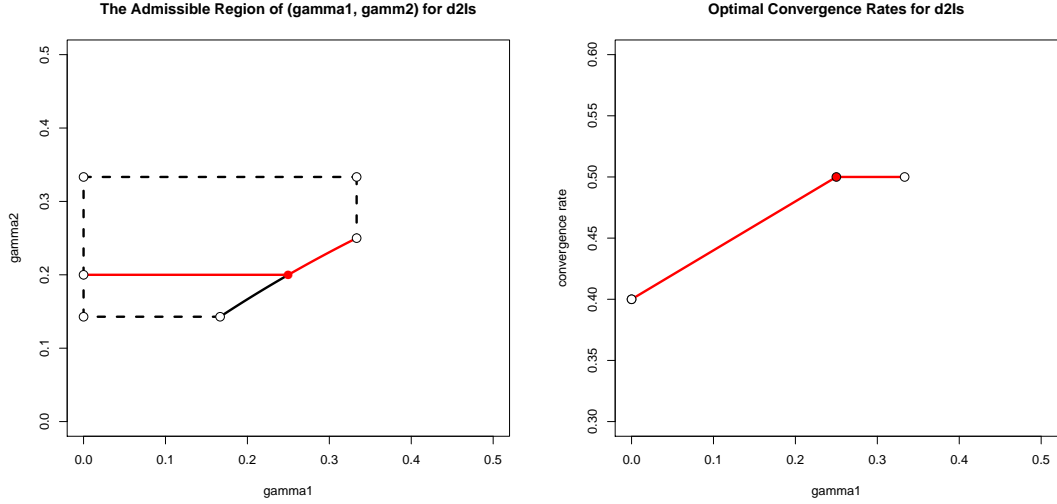


Figure 4.1: The left panel shows the admissible region of  $(\gamma_1, \gamma_2)$  for  $d_{2,Is}$  enclosed by the solid and dashed borders. The pair  $(\gamma_1, \gamma_2)$  can be on the solid borders but not on the dashed ones. When  $(\gamma_1, \gamma_2)$  lies on the red solid segment and curve in the left panel, the convergence rate of  $d_{2,Is}$  for each  $\gamma_1 \in (0, 1/3)$  is optimized. The corresponding optimal rate as a function of  $\gamma_1$  is shown in the right panel.

the left panel of Figure 4.1, all the previous constraints are satisfied simultaneously. With the intuition from Theorem 4.8.10, for each  $\gamma_1 \in (0, 1/3)$ , the optimal  $\gamma_2$  (and a proper  $c$ ) to maximize the convergence rate of  $d_{2,Is}$  is as follows. For  $\gamma_1 \in (0, 1/4]$ , let  $\gamma_2 = 1/5$ ; for  $\gamma_1 \in (1/4, 1/3)$ , let  $\gamma_2 = \gamma_1/(1 + \gamma_1)$  with  $c < c^*$ . The optimal pair  $(\gamma_1, \gamma_2)$  lies on the red solid segments in the left panel of Figure 4.1 and the optimal convergence rate of  $d_{2,Is}(d_0)$  as a function of  $\gamma_1$  is shown in the right panel of Figure 4.1.

Further, the asymptotic distributions on  $d_{2,Is}$  with optimal convergence rates are given in the following Proposition.

*Proposition 4.3.11.* Suppose assumption **(B)** holds. If  $1/4 < \gamma_1 < 1/3$ ,  $\gamma_2 = \gamma_1/(1 + \gamma_1)$  and  $c < c^*(\gamma_2)$ , we have

$$n^{1/2}(d_{2,Is} - d_0) \xrightarrow{d} C_{d_{2,Is}}(\gamma_2)N(0, 1); \quad (4.24)$$

If  $\gamma_1 = 1/4$ ,  $\gamma_2 = 1/5$  and  $c < c^*(1/5)$ , we have

$$n^{1/2}(d_{2,Is} - d_0) \xrightarrow{d} B_{d_{2,Is}} + C_{d_{2,Is}}(1/5)N(0, 1); \quad (4.25)$$



If  $0 < \gamma_1 < 1/4$  and  $\gamma_2 = 1/5$ , we have

$$n^{2(1+\gamma_1)/5}(d_{2,Is} - d_0) \xrightarrow{d} B_{d_{2,Is}} + C_{d_{2,Is}}(1/5)N(0, 1); \quad (4.26)$$

where  $C_{d_{2,Is}}(\gamma) = m'(d_0)^{-1}C_{n_r}^{-(1-\gamma)/2}\sigma\sqrt{\kappa/c}$ ,  $B_{d_{2,Is}} = -m'(d_0)^{-1}C_{n_r}^{-2/5}c^2B_1(d_0)$  and  $B_1(d_0) = m''(d_0) \int t^2 K(t)dt/2$ .

A sketch of the proof of Proposition 4.3.11 is provided in Appendix 4.8.4.

From the asymptotic results (4.24), (4.25) and (4.26), the theoretical Wald-type  $1 - \alpha$  asymptotic confidence intervals for  $d_0$  are given by

$$[d_{2,Is} \pm n^{-1/2}C_{d_{2,Is}}(\gamma_2)z_{\alpha/2}], \quad (4.27)$$

$$[d_{2,Is} - n^{-1/2}B_{d_{2,Is}} \pm n^{-1/2}C_{d_{2,Is}}(1/5)z_{\alpha/2}] \quad (4.28)$$

and

$$[d_{2,Is} - n^{-2(1+\gamma_1)/5}B_{d_{2,Is}} \pm n^{-2(1+\gamma_1)/5}C_{d_{2,Is}}(1/5)z_{\alpha/2}]. \quad (4.29)$$

## 4.4 Practical Procedures

In practice, it is of interest to consider procedures with all unknown parameters specified or estimated, which provide both point estimators of  $d_0$  with small mean square errors and confidence intervals for  $d_0$  with good coverage rates and small average lengths. Next, we consider the practical variants of OSIRP, OSSIRP, TSIRP and TSSIRP.

### 4.4.1 Practical OSIRP (POSIRP)

Denote the practical versions of OSIRP-Wald and OSIRP-LR by POSIRP-Wald and POSIRP-LR. In POSIRP-Wald, we need to estimate both  $\sigma^2$  and  $m'(d_0)$ ; in POSIRP-LR, we only need to estimate  $\sigma^2$ . We first consider the estimation of  $\sigma^2$  and then  $m'(d_0)$ .

For the estimation of  $\sigma^2$ , we employ the nonparametric estimator proposed by [Gasser et al. \(1986\)](#), which is an average of local estimators from local linear fitting. Suppose the data  $\{(X_i, Y_i)\}_{i=1}^n$  are already sorted in ascending order of the  $X_i$ 's.

Then, we calculate

$$S^2 = (n-2)^{-1} \sum_{i=2}^{n-1} c_i^2 \tilde{\epsilon}_i^2,$$

where  $\tilde{\epsilon}_i = a_i Y_{i-1} + b_i Y_{i+1} - Y_i$ ,  $c_i^2 = (a_i^2 + b_i^2 + 1)^{-1}$ ,  $a_i = (X_{i+1} - X_i)/(X_{i+1} - X_{i-1})$  and  $b_i = (X_i - X_{i-1})/(X_{i+1} - X_{i-1})$ , for  $i = 2, 3, \dots, n-1$ .

An estimate of  $m'(d_0)$  is obtained through the local quadratic regression estimator proposed by [Fan and Gijbels \(1996\)](#), at the estimate  $d_I$ . Specifically, let  $K(\cdot)$  denote the Epanechnikov kernel function and  $h > 0$  the bandwidth, so that  $K_h(\cdot) = (1/h)K(\cdot/h)$ . Further, let  $\hat{\eta} = (\hat{\eta}_0, \hat{\eta}_1, \hat{\eta}_2)$  be given by

$$\hat{\eta} = \underset{\eta \in \mathbb{R}^3}{\operatorname{argmin}} \sum_{i=1}^n \left[ Y_i - \sum_{j=0}^2 \eta_j (X_i - d_I)^j \right]^2 K_h(X_i - d_I).$$

Then, the local quadratic regression estimator of  $m'(d_I)$  is given by  $\hat{\eta}_1$ . The bandwidth  $h$  is chosen by first fitting a fifth order polynomial function to the data to obtain  $\hat{m}_{pol}(x) = \sum_{j=0}^5 \hat{\alpha}_j x^j$ . Next, the estimate of the third order derivative of  $m$  at  $d_I$  is obtained by  $\hat{m}_{pol}^{(3)}(d_I) = 6\hat{\alpha}_3 + 24\hat{\alpha}_4 d_I + 60\hat{\alpha}_5 d_I^2$ . Finally the practically optimal value for the bandwidth  $h$  is given by

$$\hat{h}_{opt} = C_{1,2}(K) \left[ S^2 / (\hat{m}_{pol}^{(3)}(d_I))^2 \right]^{1/7} n^{-1/7},$$

where  $C_{1,2}(K) = 2.275$ .

#### 4.4.2 Practical OSSIRP (POSSIRP)

Denote the Practical version of OSSIRP as POSSIRP. For the smoothing step for  $m_I$  in POSSIRP, we need to specify the kernel  $K$  and the bandwidth  $h_n$ . For example, we can use the triangular kernel  $K(x) = 1 - |x|$  for  $x \in [-1, 1]$ , which is log-concave and thus guarantees the isotonicity of  $m_{I_s}$  (see Remark 2.1 of [Mukerjee \(1988\)](#)). With this triangular kernel, we have  $\int x^2 K(x) dx = 1/6$  and  $\int K^2(x) dx = 2/3$ .

As to the bandwidth, we are interested in  $h_n = cn^{-1/5}$ , since the convergence rate of the corresponding  $m_{I_s}$  reaches  $\sqrt{n}$ , the parametric rate. A rule of thumb suggests choosing  $c$  to be the sample standard deviation of the design points  $X_i$ 's. Usually the design density  $g$  is uniform if there is no prior information on  $d_0$ . For this case, we can also let  $c$  be the standard deviation of the uniform distribution on  $[a, b]$ .

For constructing the confidence interval (4.14), we need to estimate unknown

parameters  $\sigma^2$ ,  $m'(d_0)$  and  $m''(d_0)$ . The first two can be estimated with the same methods described for POSIRP. The last one can be estimated in the same way of estimating  $m'(d_0)$  except using local cubic regression and fitting a pilot sixth degree polynomial function to the data.

#### 4.4.3 Practical TSIRP (PTSIRP)

Since the point estimator of  $d_0$  from TSIRP has a faster convergence rate than the OSIRP one, it is certainly of interest to consider the practical version of TSIRP, which is denoted by PTSIRP.

In PTSIRP, one important step is to specify the second-stage sampling interval  $[L_1, U_1]$ . We can use a high probability Wald-type confidence interval as  $[L_1, U_1]$ . Specifically, in TSIRP,  $[L_1, U_1]$  is defined as  $[d_{1,I} \pm C_1 n_1^{-\gamma_1}]$  with  $C_1 > 0$  and  $\gamma_1 \in (0, 1/3)$ . Then, we need to specify or estimate  $C_1$  and  $\gamma_1$  properly. One practical idea is to employ a high probability confidence interval as  $[L_1, U_1]$ . From Theorem 4.2.1,  $[L_1, U_1]$  can be taken as the following  $1 - \beta$  Wald-type confidence interval

$$[d_{1,I} \pm n_1^{-1/3} \hat{C}_{d_I} g_1(d_{1,I})^{-1/3} q(\mathcal{Z}, 1 - \alpha/2)] \cap [a, b], \quad (4.30)$$

where  $\hat{C}_{d_I}$  is a consistent estimator of  $C_{d_I}$ , which can be obtained by first estimating  $\sigma^2$  and  $m'(d_0)$  with the methods described for POSIRP, and  $\beta$  is a small positive number such as 0.01. With this confidence interval, we essentially choose  $C_1$  and  $\gamma_1$  such that

$$C_1 n_1^{-\gamma_1} = n_1^{-1/3} \hat{C}_{d_I} g_1(d_{1,I})^{-1/3} q(\mathcal{Z}, 1 - \alpha/2). \quad (4.31)$$

That is,

$$\gamma_1 = 1/3 \text{ and } C_1 = \hat{C}_{d_I} g_1(d_{1,I})^{-1/3} q(\mathcal{Z}, 1 - \alpha/2). \quad (4.32)$$

Although  $\gamma_1 = 1/3$  is not in  $(0, 1/3)$ , as required in TSIRP, we expect that Propositions 4.3.1 and 4.3.3 provide good approximations in practice, since  $1/3$  is at the boundary of  $(0, 1/3)$ .

This Wald-type second-stage sampling interval  $[L_1, U_1]$  requires the estimation of  $\sigma^2$  and  $m'(d_0)$ . Although they can be estimated by the methods introduced for POSIRP, usually it is difficult to estimate  $m'(d_0)$  well, especially by using only the first-stage data. In order to avoid this difficult estimation of  $m'(d_0)$ , an alternative

approach is to consider a LR-type confidence interval based on the first-stage data:

$$[L_1, U_1] = \{x \in [a, b] : 2 \log \lambda_{1,I}(x) \leq d_\beta\} \cap [a, b]. \quad (4.33)$$

Although this LR-type sampling interval is not centered at  $d_{1,I}$ , and thus Propositions 4.3.1 and 4.3.3 for TSIRP might not be applicable, it is practically attractive since it follows the essential two-stage idea and avoids the estimation of difficult unknown parameters like  $m'(d_0)$ .

Next, we consider the specification of the first-stage sample proportion  $p$  and the first stage and second-stage design densities  $g_1$  and  $g_2$  or equivalently  $\psi$ ).

By Proposition 4.3.1, the asymptotic standard deviation of  $d_{2,I}$  is given by

$$a.s.d.(d_{2,I}) = n^{-(1+\gamma_1)/3} C_{d_{2,I}} \sigma_{\mathcal{Z}},$$

where

$$C_{d_{2,I}} = C_{d_I} \left( \frac{C_1}{(1-p)p^{\gamma_1}\psi(0)} \right)^{1/3} = \left( \frac{4\sigma^2}{m'(d_0)^2} \right)^{1/3} \left( \frac{C_1}{(1-p)p^{\gamma_1}\psi(0)} \right)^{1/3}$$

and  $\sigma_{\mathcal{Z}}$  is the standard deviation of  $\mathcal{Z}$ . As a function of  $p$ ,  $a.s.d.(d_{2,I})$  is minimized when  $p = \gamma_1/(1 + \gamma_1)$ . With this  $p$ ,  $a.s.d.(d_{2,I})$ , as a function of  $\gamma_1 \in (0, 1/3]$ , is minimized when  $\gamma_1 = 1/3$ . Although theoretically  $\gamma$  should not be  $1/3$ , practically it can be used since it is at the boundary. Thus, in practice we let  $p = 1/4$ .

When there is no information about  $d_0$  at the first stage,  $g_1$  is usually taken to be the uniform density on  $[a, b]$ . If there is some prior information about  $d_0$ , We should incorporate the prior information by using a non-uniform density. From Proposition 4.3.1, large  $h(0)$  makes the asymptotic variance of  $d_{2,I}$  small. So theoretically it is proper to let  $h$  peak at 0. However, in practice  $[L_1, U_1]$  is taken to be a high probability confidence interval. Thus, it is safer to let  $h$  be uniform on  $[-1, 1]$ , otherwise big bias might occur. That is,  $g_2$  is uniform on  $[L_1, U_1]$ .

After having the second-stage data, we can compute the second-stage isotonic regression  $d_{2,I}$  of  $d_0$  and construct a practical version of the Wald-type  $1 - \alpha$  confidence interval (4.16) or the LRT-type one (4.21) with the second-stage data, where  $\alpha$  is a small positive real, for example 0.05. Again, in order to obtain practical Wald-type

confidence intervals, we need estimators of  $\sigma^2$  and (or)  $m'(d_0)$ , which can be obtained by the same methods used in POSIRP with the second-stage data.

Since both stage data become available at the second stage, practically it is better to combine them together to obtain the two-stage isotonic regression  $d_{I,pooled}$  of  $d_0$  and the practical Wald-type  $1 - \alpha$  confidence interval (4.22), or LRT-type one (4.23), where  $\sigma^2$  and (or)  $m'(d_0)$  can be estimated as before with the two-stage data.

Thus, a typical procedure of PTSIRP is as follows:

1. Let the first-stage sample proportion be  $p = 1/4$  and the first stage and second stage sample sizes be  $n_1 = \lfloor np \rfloor$  and  $n_2 = n - n_1$ , respectively, where  $n$  is the total sample size.
2. Use  $n_1$  design points on  $[a, b]$  with a continuous density  $g_1$  (usually the uniform density) and obtain the first-stage data  $\{(X_{1,i}, Y_{1,i})\}_{i=1}^{n_1}$ .
3. Compute the first stage isotonic regression  $d_{1,I}$  of  $d_0$  with the first-stage data. Then, let the second-stage sampling interval  $[L_1, U_1]$  be the  $1 - \beta$  Wald-type confidence interval (4.30) or LR-type one (4.33), where  $\beta$  is a small positive real.
4. Use  $n_2$  design points on  $[L_1, U_1]$  with a continuous density  $g_2$  (usually the uniform density) and obtain the second-stage data  $\{(X_{2,i}, Y_{2,i})\}_{i=1}^{n_2}$ .
5. Compute the two-stage isotonic regression  $d_{I,pooled}$  of  $d_0$  with the two-stage data and construct a practical version of the Wald-type  $1 - \alpha$  confidence interval (4.22) or the LRT-type one (4.23) with the combined two-stage data, where  $\alpha$  is a small positive real such as 0.05.

Note that if we employ LR-type confidence intervals for both the second-stage sampling interval  $[L_1, U_1]$  and the final  $1 - \alpha$  confidence interval, we only need to estimate  $\sigma^2$ . It is easy to implement, usually more robust and often preferred in practice.

#### 4.4.4 Practical TSSIRP (PTSSIRP)

Since the convergence rate of the estimator of  $d_0$  from TSSIRP reaches the parametric rate  $\sqrt{n}$ , it is of interest to consider the practical version of TSSIRP, which

is denoted by PTSSIRP. The average lengths of confidence intervals from PTSSIRP are expected to be relatively short.

In PTSSIRP, the second-stage sampling interval  $[L_1, U_1]$  is taken to be the  $1 - \beta$  Wald-type confidence interval (4.30) as in PTSIRP. Thus, we essentially take  $\gamma_1 = 1/3$ . Since  $1/3$  is at the boundary of  $(1/4, 1/3)$ , the asymptotic result (4.24) in Proposition 4.3.11 may still provide good approximation and we take the practical version of the confidence interval (4.27) as the final  $1 - \alpha$  confidence interval of  $d_0$ .

The first-stage sample proportion  $p$  is taken to be  $\gamma_1/(1 + \gamma_1)$  to minimize the asymptotic standard deviation of  $d_{2,Is}$ , which is inversely proportional to  $(1 - p)p^{\gamma_1}$  by the asymptotic result (4.24) in Proposition 4.3.11. Thus, we let  $p = 1/4$  since  $\gamma_1 = 1/3$ . For the same reasons in PTSIRP, both  $g_1$  and  $g_2$  are set to be uniform on  $[a, b]$  and  $[L_1, U_1]$ , respectively.

In the smoothing step of computing  $d_{2,Is}$ , we need to specify the kernel  $K$  and smoothing bandwidth  $h_n$ . Similar to POSSIRP, we can use the triangular kernel  $K(x) = 1 - |x|$  for  $x \in [-1, 1]$ .

Recall the bandwidth  $h_n$  is  $cn_r^{-\gamma_2}$ . The parameter  $\gamma_2$  needs to be  $\gamma_1/(1 + \gamma_1)$  to accelerate the convergence rate of  $d_{2,Is}$  to  $\sqrt{n}$ . Thus, we take  $\gamma_2 = 1/4$  because of  $\gamma_1 = 1/3$ . The second stage sample resolution  $n_r$  depends on  $C_1$ . Note that  $c < c^*(\gamma_2 = 1/4)$  and that  $c^*$  also depends on  $C_1$ . Since  $[L_1, U_1]$  is the Wald-type confidence interval (4.30),  $C_1$  depends on  $\hat{C}_{d_I}$ , which can be obtained by estimating  $\sigma^2$  and  $m'(d_0)$  as in POSIRP but now with the two-stage data since they are already available. Usually, we can set  $c = c^*(1/4)/\rho$  for some  $\rho > 1$ . The parameter  $\rho$  controls the ratio of the half length of the second-stage sampling interval  $[L_1, U_1]$  and the bandwidth  $h_n$ . If this ratio is too large, the smoothing step will not be practically effective. On the other hand, if this ratio is too close to 1, there will be a serious boundary effect and the estimator will usually have large bias. Our simulation studies indicate that  $\rho \in [1.5, 2.0]$  usually works fine. So our empirical suggestion for  $\rho$  is 1.75.

Thus, all the unknown parameters in PTSSIRP have been estimated or specified and we can obtain the final  $1 - \alpha$  confidence interval for  $d_0$ , which is the practical version of the confidence interval (4.27).

Thus, a typical procedure of PTSSIRP using the asymptotic result (4.24) in Propo-

sition 4.3.11 is as follows.

1. Let the first-stage sample proportion be  $p = 1/4$  and the first stage and second stage sample sizes be  $n_1 = \lfloor np \rfloor$  and  $n_2 = n - n_1$ , respectively, where  $n$  is the sample size.
2. Use  $n_1$  design points on  $[a, b]$  with a continuous density  $g_1$  (usually the uniform density) and obtain the first-stage data  $\{(X_{1,i}, Y_{1,i})\}_{i=1}^{n_1}$ .
3. Compute the first stage isotonic regression  $d_{1,I}$  of  $d_0$  with the first-stage data. Then, let the second-stage sampling interval  $[L_1, U_1]$  be the Wald-type  $1 - \beta$  confidence interval (4.30) (or the LR-type  $1 - \beta$  confidence interval (4.33)) with  $\beta$  being a small positive real.
4. Use  $n_2$  design points on  $[L_1, U_1]$  with a continuous density  $g_2$  (usually the uniform density) and obtain the second-stage data  $\{(X_{2,i}, Y_{2,i})\}_{i=1}^{n_2}$ .
5. Compute the second stage smoothed isotonic regression  $d_{2,I_s}$  of  $d_0$  with the second-stage data, estimate  $\sigma^2$  and  $m'(d_0)$  again with two-stage data, and construct a practical version of the Wald-type  $1 - \alpha$  confidence interval (4.27) by the asymptotic result (4.24) in Proposition 4.3.11, where  $\alpha$  is a small positive real such as 0.05.

## 4.5 Simulation Study

In the simulation study, we compare the practical procedures: POSIRP-Wald, POSIRP-LR, POSSIRP with Wald-type confidence intervals, PTSIRP with LR-type second-stage sampling intervals and LR-type second stage confidence intervals, and PTSSIRP with Wald-type second-stage sampling intervals and Wald-type second-stage confidence intervals.

The simulation settings are as follows: The design points lie in the interval  $[a, b] = [0, 1]$ . The regression function  $m$  is the sigmoid function  $m(x) = \exp(4(x - 0.5)) / [1 + \exp(4(x - 0.5))]$ , the quadratic function  $m(x) = x^2$  or the isotonic sine function  $m(x) = (1/40) \sin(6\pi x) + 1/4 + (1/2)x + (1/4)x^2$ . The target point  $d_0$  is 0.4, 0.5 or 0.6. The random error follows  $N(0, \sigma^2)$ , where the standard deviation of the random error  $\sigma$  is 0.1, 0.3 or 0.5. The total sample size  $n$  ranges from 100 to 700 in increments

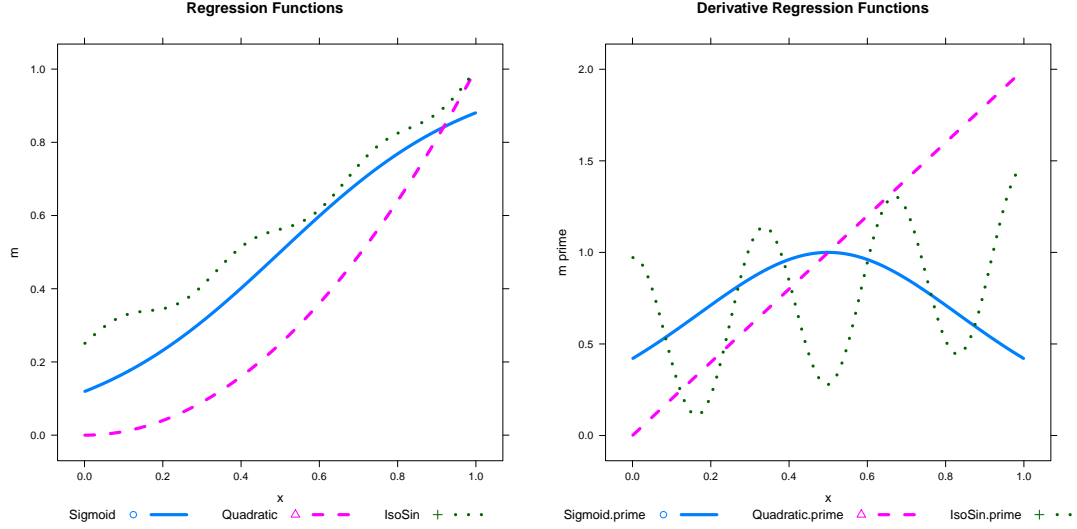


Figure 4.2: The left plot shows the regression functions: sigmoid, quadratic and isotonic sine functions. The right plot shows their derivatives.

of 100. All the design densities  $g$ ,  $g_1$  and  $g_2$  are uniform. The confidence level is  $1 - \alpha$  with  $\alpha$  being 0.05. The default number of iterations is  $n_{iter}$  is 5000.

Figure 4.2 shows the regression functions and their derivatives. Figure 4.3 shows the scatter plots of data with different regression functions, standard deviations and sample sizes, where the design points are equally spaced.

In PTSIRP and PTSSIRP, the quantiles of  $\mathbb{D}$  and  $\mathcal{Z}$  for constructing the second-stage sampling intervals are set to be 4 and 2, respectively. The corresponding values of  $\beta$  are less than 0.01.

In the simulations for POSSIRP and PTSSIRP, we found that fitting pilot fifth or sixth polynomial function usually gave bad initial estimators of  $m^{(3)}(d_0)$  or  $m^{(4)}(d_0)$ , though the resulting bandwidth seemed still reasonable. So, instead, we use local polynomial smoothing with bandwidth  $b - a$  to obtain the pilot estimators.

The coverage rates and average lengths of the 95% confidence intervals for  $d_0$  are obtained. Figure 4.4 shows the coverage rates and the average lengths of the five practical procedures with the sigmoid and quadratic functions and different values of  $\sigma$  and  $d_0$ . We can see that usually the coverage rates are close to the nominal level 95%. A larger noise increases the variation of the coverage rates and a larger sample size is usually needed to achieve better coverage rates, especially for POSSIRP-Wald, POSSIRP and PTSSIRP. Usually POSSIRP has some over coverage and PTSSIRP under coverage. POSSIRP-LR and PTSIRP consistently have best coverage rates. This



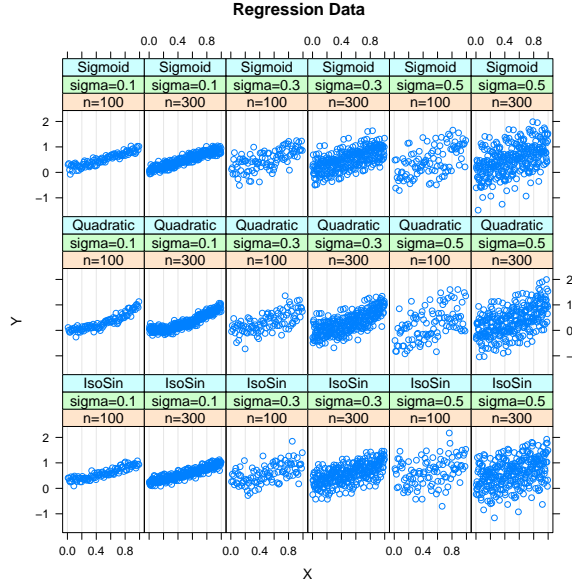


Figure 4.3: These scatter plots show data with different regression functions, standard deviations and sample sizes, where the design points follow the fixed uniform design.

tells that usually LR-type confidence intervals are stable with respect to noise. The average lengths decreases with the increase of the sample size and a larger noise results in longer confidence intervals. The confidence intervals from POSIRP are longest and those from PTSIRP are significantly shorter, but the shortest confidence intervals are from PTSSIRP. So, PTSSIRP is the most aggressive procedure for constructing confidence intervals and usually works well for the sigmoid and quadratic functions.

The left panel of Figure 4.5 shows the coverage rates of the five practical procedures for the isotonic sine function. We can see that, for the case with  $d_0 = 0.5$ , the coverage rates of the confidence intervals from POSIRP-Wald and POSSIRP even decrease as the sample size increases. This is because that the estimation of  $m'(d_0)$  is not accurate enough. More specifically, the true value of  $m'(d_0)$  is around 0.279 and the corresponding kernel estimators of  $m'(d_0)$  are usually around 0.8, significantly larger than the true value. This makes the confidence interval too short to cover  $d_0$  and the coverage rates decrease. When  $d_0 = 0.4$  or  $0.6$  and  $\sigma = 0.3$ , the coverage rates from PTSSIRP also decreases with the sample size. This is perhaps because that the asymptotically ignorable bias term plays a practical role for these cases. Fortunately, for this wiggly isotonic sine function, POSIRP-LR and PTSIRP usually have good coverage rates for all simulation cases. This shows that LR-type confidence intervals

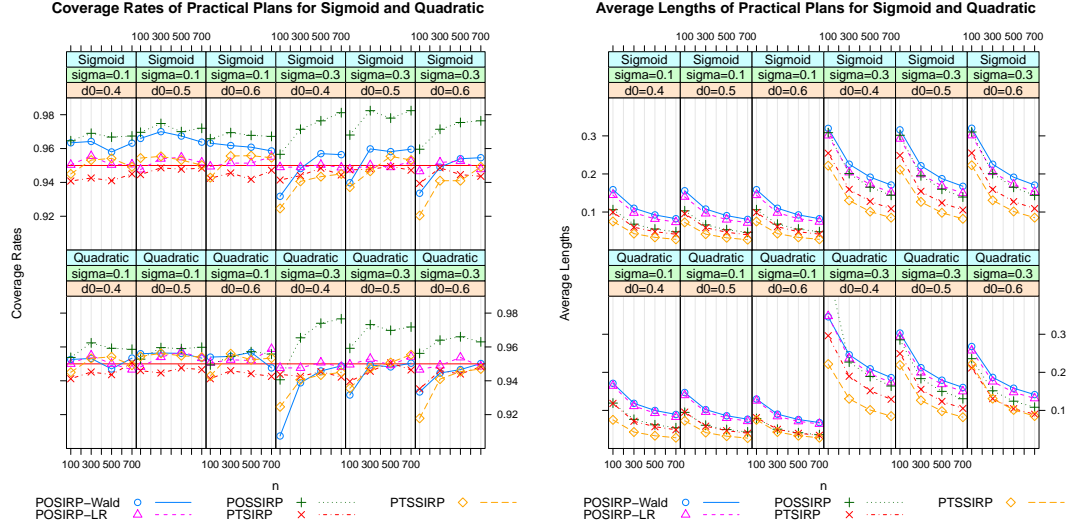


Figure 4.4: The left and right panels show the coverage rates and average lengths of the 95% confidence intervals for  $d_0$  from the five practical procedures with the sigmoid and quadratic functions and different values of  $\sigma$ ,  $d_0$  and  $n$ .

are usually robust with different regression functions. The average lengths of the confidence intervals are shown in the right panel of Figure 4.5. Without surprise, PTSIRP achieves shorter average lengths since it is a two-stage procedure.

Therefore, when the underlying regression function is well-behaved, we can use the most aggressive PTSSIRP. Otherwise, we use the conservative but stable PTSIRP. One way to decide if the underlying function is regular or not is to checking the first-stage smoothed isotonic regression estimator of  $m$ . Sometimes, we can even spend a larger number of sample points at the first stage to obtain a better first-stage estimator of  $m$  and then make the judgement more reliable.

## 4.6 An Application to Real Data

In recent years, one of the most important features for a vehicle is its fuel efficiency or economy (FE) in the unit of miles per gallon (MPG). In the USA, the National Highway Traffic Safety Administration (NHTSA) regulates the Corporate Average Fuel Economy (CAFE) standards to encourage automobile manufacturers to improve the average fuel economy of their fleets of vehicles. The CAFE standard for 2011 model year is 30.2 MPG. A manufacturer must pay a penalty if the average fuel economy of its annual fleet of production can not achieve or exceed the defined

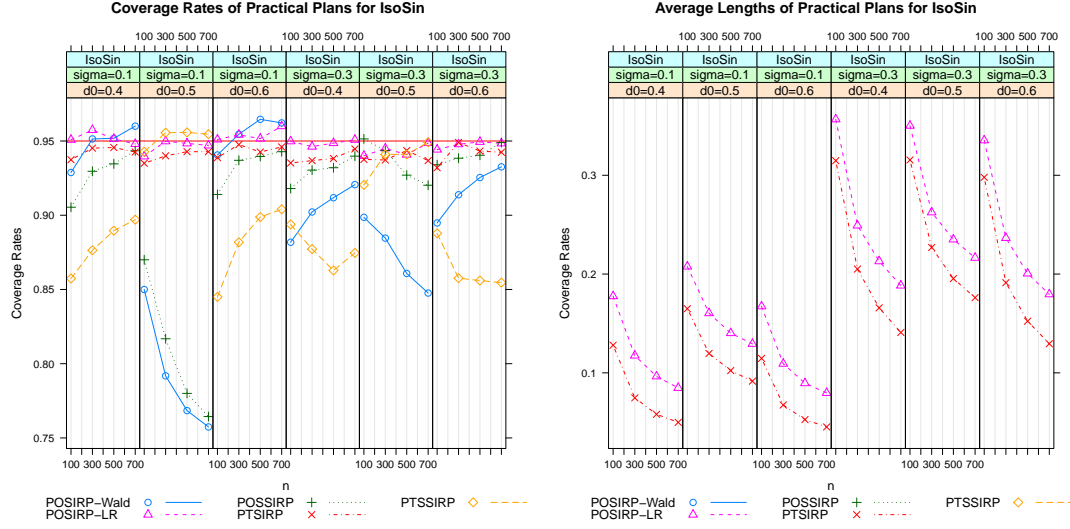


Figure 4.5: The left panel shows the coverage rates of the 95% confidence intervals for  $d_0$  from the five practical procedures with the isotonic sine functions and different values of  $\sigma$ ,  $d_0$  and  $n$ . The right panel shows the average lengths of the 95% confidence intervals for  $d_0$  from POSSIRP-LR and PTSIRP.

standard. On the other hand, the US Energy Tax Act of 1978 imposes a gas guzzler tax on the sale of new model year cars with low fuel efficiency. Although both the penalty and tax only apply to manufactures and importers of vehicles, presumably they are at least partially passed along to consumers in the form of higher prices. So these regulations intend to encourage manufacturers and consumers to produce and purchase vehicles with high fuel efficiency.

In the USA, while the CAFE standards are regulated by NHTSA, the vehicle fuel efficiency is evaluated by the Environmental Protection Agency (EPA). From 2008, EPA measures the fuel efficiency of a vehicle in two testing modes, the city and highway modes, with the consideration of faster speeds and acceleration, air conditioner usage and colder outside temperatures to better approximate the real-world fuel efficiency. From the unadjusted city and highway fuel efficiency, the unadjusted combined fuel efficiency is calculated as follows:

$$\text{Combined FE} = \frac{1}{.495/\text{City FE} + .351/\text{Highway FE}} + .15.$$

Since usually there is a decreasing relationship between a vehicle's horse power and its fuel efficiency, it is of interest to study this relationship and identify the horse

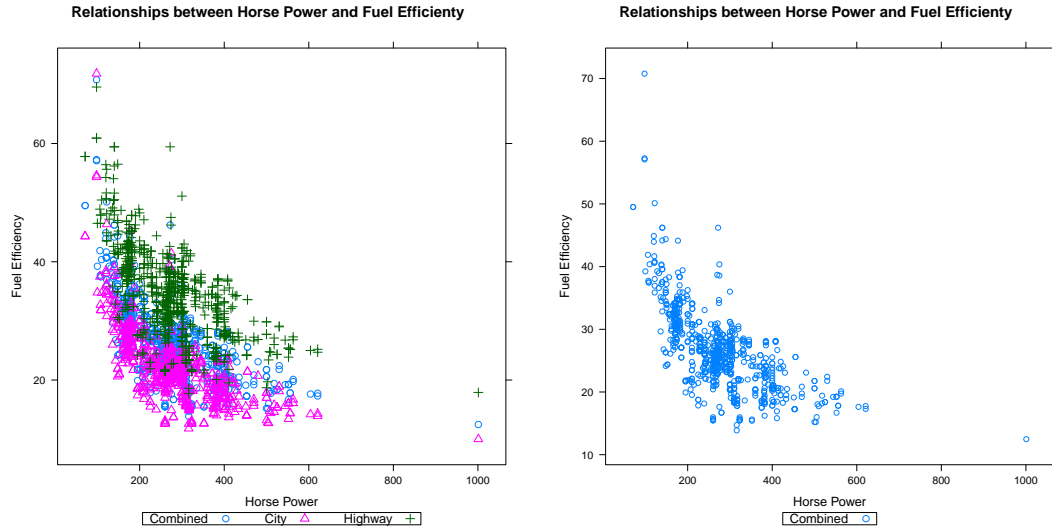


Figure 4.6: The left panel shows the scatter plots on the relationships between the horse power and the combined, city and highway fuel efficiency and the right panel shows the scatter plot on the relationships between the horse power and the combined fuel efficiency.

power at which the fuel efficiency is equal to a given fuel efficiency.

The web site [www.fueleconomy.gov](http://www.fueleconomy.gov) provides the fuel economy data of the 2011 model year vehicles. This data set contains the unadjusted city, highway and combined fuel efficiency for 1052 vehicles. We collect the horse power data for 726 non-hybrid vehicles with auto transmission gearboxes. The goal of this real data analysis is to estimate the horse power at which the combined fuel efficiency is equal to 30 MPG, around the 2011 model year CAFE standard.

#### 4.6.1 Preliminary Analysis

The scatter plots in the left panel of Figure 4.6 confirms the decreasing relationships between horse power and the unadjusted city, highway and combined fuel efficiency. Since the combined fuel efficiency is a modified harmonic average of the city and highway fuel efficiency, its plot lies in the middle. Regulations set requirements on the combined fuel efficiency. So we next focus on the relationship between the horse power and the combined fuel efficiency, whose scatter plot is shown in the right panel of Figure 4.6.

The left panel of Figure 4.7 is the same to the right panel of Figure 4.6 except that the method of air aspiration is used as a group variable. It shows that the turbo

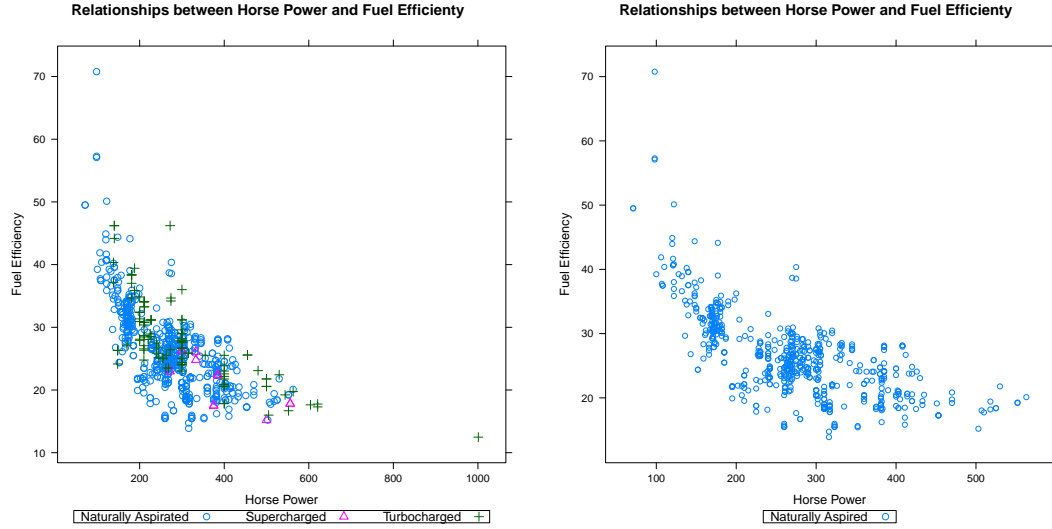


Figure 4.7: The left panel shows the scatter plot on the relationship between horse power and combined fuel efficiency grouped with three methods of air aspiration. The right panel shows the scatter plot on the relationship between horse power and combined fuel efficiency with the natural air aspiration.

charged air aspiration method might significantly improve fuel efficiency. So, for the homogeneity of the data, we next pay attention to the vehicles with the natural air aspiration method since this method is currently most popular. The corresponding scatter plot with the sample size 601 is shown in the right panel of Figure 4.7.

The frequency plot in the right panel of Figure 4.8 indicates that there are multiple values of the fuel efficiency for many of the 153 distinguished values of the horse power. We simply take the arithmetic average of the multiple values of the fuel efficiency at a value of the horse power and the scatter plot with a sample size 153 in the left panel of Figure 4.8 shows the decreasing relationship between the horse power and the average combined fuel efficiency.

For the simplicity of analysis, we ignore the weights of the averages and view the the 153 data points in the left panel of Figure 4.8 as i.i.d. sample points from the regression model (4.1), with  $m$  now being a decreasing function. We are interested in the estimation of  $d_0 = m^{-1}(30)$ , the value of the horse power at which the combined fuel efficiency is equal to 30 MPG.

The practical procedures in Section 4.4 are proposed in the design setting, but the 153 sample points from the real data set are actually already observed. However,

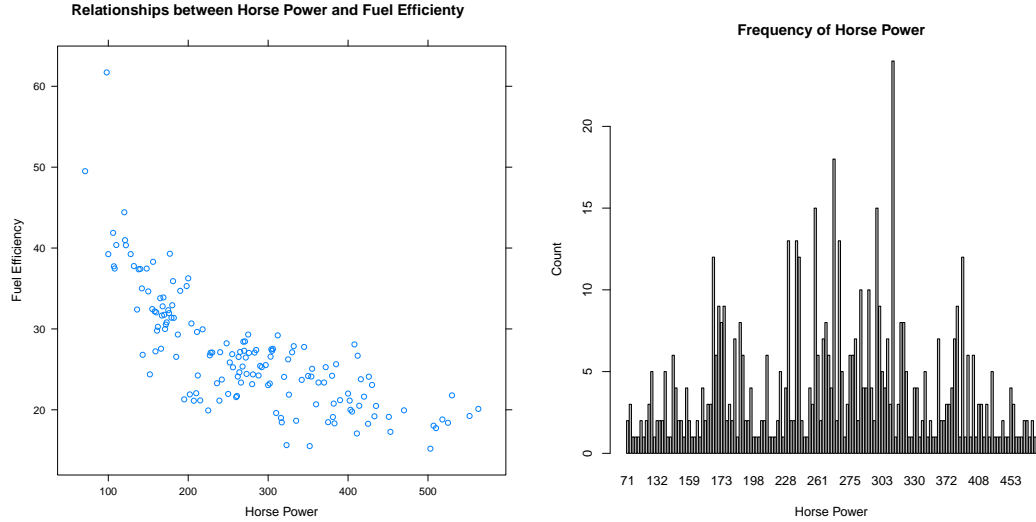


Figure 4.8: The left panel shows the scatter plot on the relationship between the horse power and average combined fuel efficiency. The right panel shows the frequencies of the values for the horse power.

we can simulate the design setting as follows. Suppose we have no knowledge of the existence of the 153 observed sample points and the total budget for estimating  $d_0$  is 100 sample points. Then, for a one-stage procedure, we choose 100 sample points from the 153 ones with the corresponding 100 values of the horse power more or less equally spaced. On the other hand, for a two-stage procedure, we first choose 50 sample points and construct a second-stage sampling interval. Then, from the remaining 103 sample points, choose another 50 sample points with the values of the horse power located in the second-stage sampling interval if possible. If there are less than 50 values of the horse power within that interval, we choose all of them. Thus, we can compare the point estimators and confidence intervals for  $d_0$  from different procedures. As for the practical “true” value of  $d_0$ , we take the isotonic regression estimator of  $d_0$  with all the 153 observed sample points, which is around 200.265.

#### 4.6.2 Practical Procedures

We apply the five practical procedures POSIRP-Wald, POSIRP-LR, POSSIRP, PTSIRP and PTSSIRP evaluated in the simulation study of Section 4.5 to obtain the point estimators and confidence intervals for  $d_0$ .

For the one-stage procedures, the sample size is 100. From POSIRP, the one-stage isotonic regression estimator of  $d_0$  is about 192.168 and the 95% Wald-type and

Table 4.1: Data Analysis Results for the Five Practical Procedures

Procedure	Estimator	Distance	95% CI	Coverage	Length	$n$
POSIRP-Wald	192.168	8.097	[189.185, 195.150]	No	5.965	100
POSIRP-LR	192.168	8.097	[150, 198]	No	48	100
POSSIRP	191.755	8.51	[177.191, 201.416]	No	24.225	100
PTSIRP	200.265	0	[179, 204]	Yes	25	80
PTSSIRP	191.422	8.843	[172.054, 210.790]	Yes	38.736	60

LR-type confidence intervals are [189.185, 195.150] and [150, 198], respectively. The one-stage smoothed regression estimator of  $d_0$  from POSSIRP is around 191.755 and the corresponding 95% Wald-type confidence interval is [177.191, 201.416].

For the two-stage procedures, the first-stage sample size is 50. For PTSIRP, the LR-type second-stage sampling interval is [136, 204], within which there are only another 30 values for the horse power. So, the second-stage sample size is 30 and the total or two-stage sample size is 80. The two-stage estimator of  $d_0$  is 200.265 and the two-stage LR-type 95% confidence interval is [179, 204]. The Wald-type second-stage sampling interval in PTSSIRP is [172.830, 191.248], centered at 182.039, the first-stage isotonic regression estimator of  $d_0$ . Since within this interval only 10 more values for the horse power are available, the second-stage sample size is 10 and the two-stage sample size is 60. The two-stage estimator of  $d_0$  from PTSSIRP is 191.422 and the corresponding two-stage Wald-type 95% confidence interval is [172.054, 210.790].

All the results are summarized in Table 4.1 and plotted in Figure 4.9. In Table 4.1, “Procedure” indicates the name of a procedure, “Estimator” shows the value of a point estimator for  $d_0$ , “Distance” measures the distance between a point estimator for  $d_0$  and the practical “true” value of  $d_0$ , “CI” stands for “confidence interval”, “Coverage” indicates whether a confidence interval covers the practical “true” value of  $d_0$ , “Length” shows the length of a confidence interval and “ $n$ ” is the total sample size.

Figure 4.9 includes four plots for the five procedures. POSIRP-Wald and POSIRP-LR share the top-left plot. In these plots, a number “1” stands for a one-stage or first-stage sample point and a number “2” for a second-stage one. Since the estimators of  $d_0$  vary from about 190 to 200, the x-axes of the plots are all set to be from 60 to 300 to improve visualization. In each plot, the red curve is a final estimator of the whole function  $m$ , the red round solid point and two red star points at the

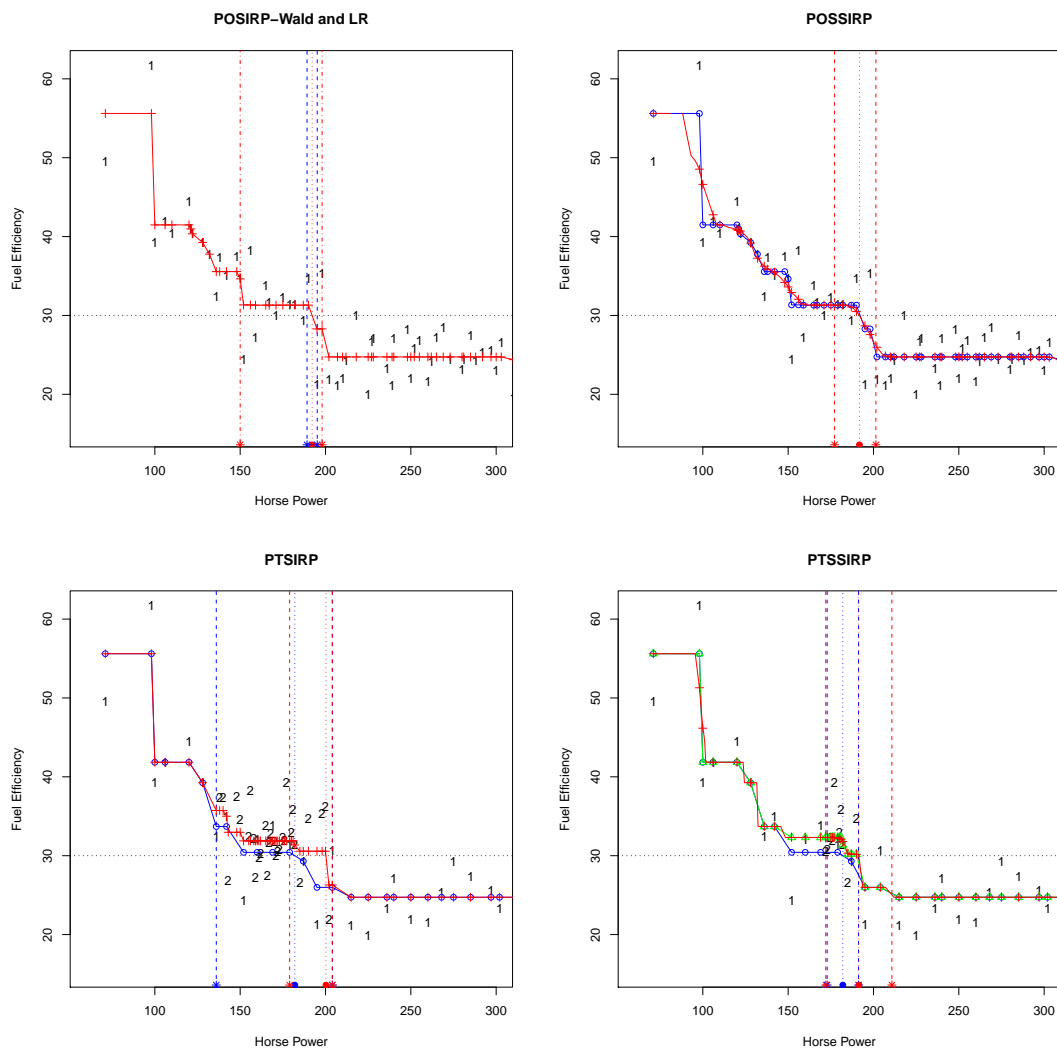


Figure 4.9: The top-left panel is for POSIRP-Wald and POSIRP-LR. The top-right panel is for POSSIRP. The bottom-left panel is for PTSIRP. The bottom-right panel is for PTSSIRP.



x-axis indicate the location of an estimator of  $d_0$  and the two ends of a final 95% confidence interval, while in the plot with the title “POSIRP-Wald and LR” the two blue star points at the x-axis are for the Wald-type confidence intervals. The blue curves in the plots for POSSIRP and the two-stage procedures are the piecewise-linear isotonic regression estimators of  $m$  and in the plot for PTSSIRP the green curve is the two-stage piecewise-linear isotonic regression estimator of  $m$ . In the plots for the two-stage procedures, the blue round solid points stand for the first-stage estimators of  $d_0$  and the blue star points for the ends of second-stage sampling intervals. Vertical dashed lines with the same colors to the end points are draw for all the confidence intervals and the second-stage sampling intervals to make comparison easier.

From Table 4.1 or Figure 4.9, we can see that, for OSIRP, the Wald-type confidence interval is centered at the one-stage estimator of  $d_0$  while the LR-type confidence interval is asymmetric. Further, the LR-type confidence interval is much larger than the Wald-type one, especially for the low limit. This is because that the LR-type confidence interval is more adaptive to the data. In fact, the top-left plot in Figure 4.9 shows that there is, to the left side of the one-stage estimator, a long flat part in the isotonic regression of  $m$ , which is captured in the LR-type confidence interval. The aggressiveness of the Wald-type confidence interval comparing to the LR-type one is also revealed by comparing the second-stage stage sampling intervals in the plots for the two-stage procedures.

Most of the estimators for  $d_0$  are around 190 and only the estimator from PTSIRP is quite different. In this procedure, the second-stage sampling interval covers a proper region so that the second-stage data have good quality with respect to the estimation of  $d_0$ . Together with the fact that isotonic regression is a local estimator, the second-stage isotonic regression estimator of  $d_0$  in this procedure is just equal to the practical “true” value of  $d_0$ .

The one-stage confidence intervals do not cover the practical “true” value of  $d_0$ , but the two-stage confidence intervals do. The confidence interval from PTSSIRP is significantly larger than that from PTSIRP. It is partially because the total sample size for the former procedure is smaller. Although the sample size for PTSIRP is only 4/5 of that of POSSIRP, the lengths of their confidence intervals are similar, about 25. Therefore, from all the above analysis, PTSIRP performs best for this real data set, in terms of providing the point estimator and 95% confidence interval.

## 4.7 Conclusion

In this chapter, we consider the estimation of  $d_0$ , the inverse of an increasing regression function at a given point in a design setting. When the regression function is locally linear around  $d_0$ , the two-stage hybrid procedure in the previous chapter not only has good theoretical properties but also usually performs well in practice. However, it usually can not handle the the case well where the regression function is very locally nonlinear around  $d_0$  and the sample size is not large enough. To address this problem, we propose alternative two-stage procedures: the two-stage isotonic regression procedure (TSIRP) and the two-stage smoothed isotonic regression procedure (TSSIRP). The convergence rate of the second-stage estimator of  $d_0$  in TSSIRP can even achieve the parametric rate  $\sqrt{n}$ . We also provide practical variants of TSIRP and TSSIRP with all the unknown parameters being estimated and specified. A simulation study shows that the practical TSSIRP usually does somewhat better for well-behaved regression functions while the practical TSIRP can still provide confidence intervals of  $d_0$  with good coverage rates for wiggly regression functions. This stability of TSIRP is of practical interest.

## 4.8 Appendix

In this appendix, we provide proofs for some theorems in the main text.

### 4.8.1 Appendix for OSIRP

Here we provide a sketch of the proof for Theorem 4.2.1 based on [Banerjee and Wellner \(2005\)](#).

*The sketch of the proof of Theorem 4.2.1.* For every  $x \in \mathbb{R}$ ,

$$\begin{aligned}
 P(n^{1/3}(d_I - d_0) \leq x) &= P(d_I \leq d_0 + xn^{-1/3}) \\
 &= P(\theta_0 \leq m_I(d_0 + xn^{-1/3})) \\
 &= P(n^{1/3}(m_I(d_0 + xn^{-1/3}) - \theta_0) \geq 0) \\
 &= P(n^{1/3}(m_I(d_0 + xn^{-1/3}) - m(d_0)) \geq 0).
 \end{aligned}$$

Then, the problem on  $d_I$  is translated into the problem on the isotonic regression  $m_I$ . Let  $\{W(t), t \in \mathbb{R}\}$  be a two-sided Brownian Motion with  $W(0) = 0$ ;  $X_{a,b} = \{X_{a,b}(t) =$

$aW(t) + bt^2, t \in \mathbb{R}$ , where  $a > 0, b > 0$ ;  $G_{a,b} = \{G_{a,b}(t), t \in \mathbb{R}\}$  the greatest convex minorant function of  $X_{a,b}$ ;  $g_{a,b} = \{g_{a,b}(t), t \in \mathbb{R}\}$  the left derivative function of  $G_{a,b}$ . Note that there is a little abuse of  $a$  and  $b$ . They are generic positive constants here, but not the two end points of the original sampling interval  $[a, b]$ . With these notations, we have that, under the assumption **(A)**,

$$n^{1/3} (m_I(d_0 + xn^{-1/3}) - m(d_0)) \xrightarrow{d} g_{a,b}(x), \quad (4.34)$$

where  $a = \sigma/\sqrt{g(d_0)}$  and  $b = m'(d_0)/2$ . Then, from the switching relationship

$$g_{a,b}(x) \geq \lambda \iff \underset{t \in \mathbb{R}}{\operatorname{argmin}}(X_{a,b}(t) - \lambda t) \leq x, \text{ for } \lambda \in \mathbb{R}, \quad (4.35)$$

we have

$$P(n^{1/3}(d_I - d_0) \leq x) = P\left(\underset{t \in \mathbb{R}}{\operatorname{argmin}} X_{a,b}(t) \leq x\right). \quad (4.36)$$

Further, from Problem 5 on Page 308 of [van der Vaart and Wellner \(1996\)](#), we have

$$\underset{t \in \mathbb{R}}{\operatorname{argmin}}(X_{a,b}(t) - \lambda t) \stackrel{d}{=} (a/b)^{2/3} \underset{t \in \mathbb{R}}{\operatorname{argmin}} X_{1,1}(t) + \lambda/(2b). \quad (4.37)$$

Thus, we have

$$P(n^{1/3}(d_I - d_0) \leq x) = P\left((a/b)^{2/3} \underset{t \in \mathbb{R}}{\operatorname{argmin}} X_{1,1}(t) \leq x\right). \quad (4.38)$$

Since  $\underset{t \in \mathbb{R}}{\operatorname{argmin}} X_{1,1}(t)$  is  $\mathcal{Z}$  and  $(a/b)^{2/3} = C_{d_I} g(d_0)^{-1/3}$ , Theorem 4.2.1 follows. □

From (4.34), (4.35) and (4.37), we can obtain the asymptotic distribution on  $m_I$ . *Theorem 4.8.1.* Under the assumption **(A)**, we have, for  $x \in (a, b)$ ,

$$n^{1/3}(m_I(x) - m(x)) \xrightarrow{d} C_{m_I}(x)g(x)^{-1/3}\mathcal{Z}, \quad (4.39)$$

where  $C_{m_I}(x) = (4\sigma^2 m'(x))^{1/3}$ .

## 4.8.2 Appendix for OSSIRP

Before providing the sketch of the proof of Theorem 4.2.5, we state a related Theorem on the smoothed isotonic regression  $m_{I_S}$ .

*Theorem 4.8.2.* Suppose the assumption **(B)** holds,  $nh_n^3 \rightarrow \infty$  and  $nh_n^7 \rightarrow 0$ . Then, for  $x \in (a, b)$ ,

$$\sqrt{nh_n} (m_{I_s}(x) - m(x) - h_n^2 B(x)) \xrightarrow{d} N(0, \kappa\sigma^2/g(x)), \quad (4.40)$$

where

$$B(x) = (2g'(x)m'(x) + g(x)m''(x)) / (2g(x)) \int t^2 K(t) dt, \quad \kappa = \int K^2(t) dt.$$

*Proof.* According to [Mukerjee \(1988\)](#) and [Mammen \(1991\)](#), for a fixed  $x \in (a, b)$ ,  $m_{I_s}(x) = m_s(x)$  with probability tending to 1. Then, by the classical asymptotic results on  $m_s$  (For example, see Theorem 2.2 in [Li and Racine \(2007\)](#)), where  $m_s$  is the kernel smoothing estimator of  $m$ . under the assumption **(B)** and  $\gamma \in (1/7, 1/3)$ , we have the asymptotic result (4.40).  $\square$

From [Theorem 4.8.2](#), we have the following Corollary under the formulation  $h_n = cn^{-\gamma}$  with  $c > 0$  and  $\gamma > 0$ .

*Corollary 4.8.3.* Suppose the assumption **(B)** holds. For  $x \in (a, b)$  and  $\gamma = 1/5$ , we have

$$n^{2/5} (m_{I_s}(x) - m(x)) \xrightarrow{d} c^2 B(x) + N(0, \kappa\sigma^2/(cg(x))); \quad (4.41)$$

for  $\gamma \in (1/5, 1/3)$ , we have

$$n^{(1-\gamma)/2} (m_{I_s}(x) - m(x)) \xrightarrow{d} N(0, \kappa\sigma^2/(cg(x))). \quad (4.42)$$

Next, we provide the sketch of the proof of [Theorem 4.2.5](#).

*The sketch of the proof of [Theorem 4.2.5](#).* We only show the sketch for the first result (4.12) and the second result (4.13) can be shown in the same way.

By [Corollary 4.8.3](#), for  $h_n = cn^{-1/5}$  with  $c > 0$ , we have (4.41). On the other hand, for each  $t \in \mathbb{R}$ , we have

$$\begin{aligned} P(n^{2/5}(d_{I_s} - d_0) \leq t) &= P(m_{I_s}(d_0 + tn^{-2/5}) - m(d_0) \geq 0) \\ &= P(T_1 + T_2 \geq 0), \end{aligned}$$

where

$$\begin{aligned} T_1 &= n^{2/5} (m_{Is}(d_0 + tn^{-2/5}) - m_{Is}(d_0) - m'(d_0)tn^{-2/5}), \\ T_2 &= n^{2/5} (m_{Is}(d_0) - m(d_0) + m'(d_0)tn^{-2/5}). \end{aligned}$$

Since  $T_1 \xrightarrow{P} 0$ , the asymptotic distribution of  $T_1 + T_2$  is the same to that of  $T_2$ . Then, from

$$P(T_2 \geq 0) = P(-m'(d_0)^{-1}n^{2/5}(m_{Is}(d_0) - m(d_0)) \leq t)$$

and the asymptotic result (4.41), the result (4.12) in Theorem 4.2.5 follows.  $\square$

### 4.8.3 Appendix of TSIRP

To study the asymptotic properties of  $d_{2,I}$  and  $2 \log \lambda_{2,I}$ , the basic idea is to approximate them by the corresponding statistics in the following ideal two-stage isotonic regression procedure (ITSIRP). More specifically, the procedure of ITSIRP is as follows:

1. Set the first-stage sample proportion  $p \in (0, 1)$  and let the first and second-stage sample sizes be  $n_1 = \lfloor np \rfloor$  and  $n_2 = n - n_1$ , respectively, where  $n$  is the total sample size.
2. Let the ideal second-stage sampling interval be  $[L_1, U_1] = [d_0 \pm C_1 n_1^{-\gamma_1}]$  with  $C_1 > 0$  and  $\gamma_1 > 0$ .
3. Allocate the second-stage design points  $\{X_{2,i}\}_{i=1}^{n_2}$  with a Lebesgue density  $g_2$  on  $[L_1, U_1]$  and then the corresponding i.i.d. second-stage responses  $\{Y_{2,i}\}_{i=1}^{n_2}$ .
4. Compute the unconstrained isotonic regression  $m_{oI}$  (and the constrained one  $m_{oIc}$  under the null hypothesis  $m^{-1}(\theta_0) = d_0$ ) of  $m$  over  $[L_1, U_1]$  from the second-stage data.
5. Obtain  $d_{oI} = m_{oI}^{-1}(\theta_0)$  and  $2 \log \lambda_{oI} = 2 \log \lambda_{oI}(d_0) = 2 [l_n(m_{oI}) - l_n(m_{oIc})]$ , the ideal second-stage isotonic regression of  $d_0$  and log likelihood ratio test statistics under  $H_0 : m^{-1}(\theta_0) = d_0$ .

Note that ‘ $o$ ’ means ‘ideal’ and the Lebesgue design density  $g_2$  on  $[L_1, U_1]$  is defined as  $g_2(x) = (C_1 n_1^{-\gamma_1})^{-1} \psi((x - d_0)/(C_1 n_1^{-\gamma_1}))$  with  $\psi$  being a Lebesgue density on  $[-1, 1]$ .

Comparing ITSIRP and TSIRP, we can see that ITSIRP is the same to TSIRP except that in ITSIRP the second-stage sampling interval is centered at  $d_0$  but not at  $d_{1,I}$ . Since  $d_{1,I}$  goes to  $d_0$  at the convergence rate of  $n^{1/3}$ , the second-stage sample resolution at  $d_0$  would be the same for those two procedures given  $\gamma_1 < 1/3$ . On the other hand, all the statistics  $d_{2,I}$ ,  $d_{oI}$ ,  $2 \log \lambda_{2,I}$  and  $2 \log \lambda_{oI}$  are asymptotically determined by the sample resolution around  $d_0$ . Then, heuristically speaking, both  $n^{(1+\gamma_1)/3}(d_{2,I} - d_{oI})$  and  $2 \log \lambda_{2,I} - 2 \log \lambda_{oI}$  would converge to 0 in probability. This intuitive statement needs rigorous justification. Now, suppose the above heuristic holds, the asymptotic distributions on  $d_{2,I}$  and  $2 \log \lambda_{2,I}$  are completely determined by those on  $d_{oI}$  and  $2 \log \lambda_{oI}$ . The following theorems contains the limiting distributions on  $m_{oI}$ ,  $d_{oI}$  and  $2 \log \lambda_{oI}$ , from which the asymptotic results in Propositions 4.3.1 and 4.3.3 follow directly.

*Theorem 4.8.4.* Under the assumption **(A)**, we have

$$n^{(1+\gamma_1)/3}(m_{oI}(d_0) - m(d_0)) \xrightarrow{d} C_{m_{oI}} \mathcal{Z},$$

where

$$C_{m_{oI}} = C_{m_I}(d_0) \left( \frac{C_1}{(1-p)p^{\gamma_1}\psi(0)} \right)^{1/3}$$

and  $C_{m_I}$  is defined in Theorem 4.8.1.

*Theorem 4.8.5.* Under the assumption **(A)**, we have

$$n^{(1+\gamma_1)/3}(d_{oI} - d_0) \xrightarrow{d} C_{d_{oI}} \mathcal{Z},$$

where

$$C_{d_{oI}} = C_{d_I} \left( \frac{C_1}{(1-p)p^{\gamma_1}\psi(0)} \right)^{1/3}.$$

*Theorem 4.8.6.* Under the assumption **(A)** and the null hypothesis  $H_0 : m^{-1}(\theta_0) = d_0$ ,

$$2 \log \lambda_{oI} \xrightarrow{d} \mathbb{D}.$$

Next we provide a sketch of the proof of Theorem 4.8.5, which is similar to that for Theorem 4.2.1 in Appendix 4.8.1. The convergence rate of  $d_{oI}$  is determined by the sample resolution at  $d_0$ , which is  $n^{1+\gamma_1}$ . By the property of isotonic regression,

the convergence rate of  $d_{oI}$  is  $n^{(1+\gamma_1)/3}$ . Then, we have, for every  $x \in \mathbb{R}$ ,

$$\begin{aligned}
P\left(n_2^{(1+\gamma_1)/3}(d_{oI} - d_0) \leq x\right) &= P\left(d_{oI} \leq d_0 + xn_2^{(1+\gamma_1)/3}\right) \\
&= P\left(\theta_0 \leq m_{oI}(d_0 + xn_2^{(1+\gamma_1)/3})\right) \\
&= P\left(n_2^{(1+\gamma_1)/3}(m_{oI}(d_0 + xn_2^{(1+\gamma_1)/3}) - \theta_0) \geq 0\right) \\
&= P\left(n_2^{(1+\gamma_1)/3}(m_{oI}(d_0 + xn_2^{(1+\gamma_1)/3}) - m(d_0)) \geq 0\right).
\end{aligned}$$

Thus, it is sufficient to derive the limiting distribution of  $n_2^{(1+\gamma_1)/3}(m_{oI}(d_0 + xn_2^{(1+\gamma_1)/3}) - m(d_0))$ . In fact, we have the following theorem.

*Theorem 4.8.7.* Under the assumption **(A)**, we have, for every  $x \in \mathbb{R}$ ,

$$n_2^{(1+\gamma_1)/3}(m_{oI}(d_0 + xn_2^{-(1+\gamma_1)/3}) - m(d_0)) \xrightarrow{d} g_{a,b}(x),$$

where  $a = (\sigma^2[(1-p)/p]^{\gamma_1} C_1 / \psi(0))^{1/2}$ ,  $b = m'(d_0)/2$ .

Then, by the switching relationship (4.35), we have

$$P\left(n_2^{(1+\gamma_1)/3}(d_{oI} - d_0) \leq x\right) \rightarrow P\left(\underset{t \in \mathbb{R}}{\operatorname{argmin}} X_{a,b}(t) \leq x\right). \quad (4.43)$$

Further, by the equality in distribution (4.37), we have

$$P\left(n_2^{(1+\gamma_1)/3}(d_{oI} - d_0) \leq x\right) \rightarrow P\left((a/b)^{2/3} \underset{t \in \mathbb{R}}{\operatorname{argmin}} X_{1,1}(t) \leq x\right). \quad (4.44)$$

Since

$$(a/b)^{2/3} = \left(\frac{4\sigma^2}{m'(d_0)^2} \cdot \frac{(1-p)^{\gamma_1} C_1}{p^{\gamma_1} \psi(0)}\right)^{1/3},$$

we have

$$n_2^{(1+\gamma_1)/3}(d_{oI} - d_0) \xrightarrow{d} C_{d_I} \left(\frac{(1-p)^{\gamma_1} C_1}{p^{\gamma_1} \psi(0)}\right)^{1/3} \mathcal{Z},$$

which leads to  $n^{(1+\gamma_1)/3}(d_{oI} - d_0) \xrightarrow{d} C_{d_{oI}} \mathcal{Z}$ , the result in Theorem 4.8.5.

A sketch of the proof for Theorem 4.8.7 is as follows. Without loss of generality, we let  $[a, b] = [0, 1]$ . It basically consists of three steps: the first one is to use a switching relationship to change the original problem into an M-Estimation problem; the second one is to solve the M-Estimation problem in the framework of the empirical process theory; the third one is to simplify the final limit distribution.

More specifically, in the first step, we show

*Lemma 4.8.8.* For  $t \in [0, 1]$  and  $s \in \mathbb{R}$ ,

$$m_{oI}(t) \leq s \Leftrightarrow \underset{x \in [x_0 \pm C_1 n_1^{-\gamma_1}]}{\operatorname{argmin}} \{V_{n_2}(x) - sG_{n_2}(x)\} \geq T(t), \quad (4.45)$$

where, for any  $x \in [0, 1]$ ,

$$V_{n_2}(x) = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i 1(X_i \leq x), \quad G_{n_2}(x) = \frac{1}{n_2} \sum_{i=1}^{n_2} 1(X_i \leq x); \quad (4.46)$$

$T(t)$  is the least  $X_i$  greater than or equal to  $t$  given  $t \leq X_{n_2}$ ; otherwise,  $T(t) = X_{n_2}$ .

By Lemma 4.8.8, for every  $x_0 \in (0, 1)$  and  $z \in \mathbb{R}$ ,

$$\begin{aligned} & P \left( n_2^{(1+\gamma_1)/3} (m_{oI}(x_0 + x n_2^{-(1+\gamma_1)/3}) - m(x_0)) \leq z \right) \\ &= P \left( m_{oI}(x_0 + x n_2^{-(1+\gamma_1)/3}) \leq m(x_0) + z n_2^{-(1+\gamma_1)/3} \right) \\ &= P \left( \underset{x \in [x_0 \pm K n_1^{-\gamma_1}]}{\operatorname{argmin}} \left\{ V_{n_2}(x) - (m(x_0) + z n_2^{-(1+\gamma_1)/3}) G_{n_2}(x) \right\} \geq T(x_0 + x n_2^{-(1+\gamma_1)/3}) \right). \end{aligned}$$

Then in the second step, we will show

*Theorem 4.8.9.* Under the assumption **(A)**, we have, as  $n \rightarrow \infty$ ,

$$\begin{aligned} & P \left( \underset{x \in [x_0 \pm K n_1^{-\gamma_1}]}{\operatorname{argmin}} \left\{ V_{n_2}(x) - (m(x_0) + z n_2^{-(1+\gamma_1)/3}) G_{n_2}(x) \right\} \geq T(x_0 + x n_2^{-(1+\gamma_1)/3}) \right) \\ & \rightarrow P \left( \underset{h \in \mathbb{R}}{\operatorname{argmin}} \{ aW(h) + bh^2 - zh \} \geq x \right), \end{aligned}$$

where  $a = ((\sigma^2[(1-p)/p]^{\gamma_1} K/h(0)))^{1/2}$  and  $b = m'(x_0)/2$ .

In the third step, by the switching relationship (4.35), we have

$$P \left( \underset{h \in \mathbb{R}}{\operatorname{argmin}} \{ aW(h) + bh^2 - zh \} \geq x \right) = P(g_{a,b}(x) \leq z).$$

Thus, Theorem 4.8.7 follows.

Finally, we provide a sketch of the proof for Theorem 4.8.6. Without the loss of generality, we assume the second-stage sampling density is uniform on  $[L_1, U_1]$ . That



is,  $g_2(x) = (2C_1 n_1^{-\gamma_1})^{-1}$  for  $x \in [L_1, U_1]$ . Then, similar to Theorem 4.8.7, we have

$$\begin{pmatrix} n_2^{(1+\gamma_1)/3}(m_{oI}(d_0 + xn_2^{-(1+\gamma_1)/3}) - m(d_0)) \\ n_2^{(1+\gamma_1)/3}(m_{oIc}(d_0 + xn_2^{-(1+\gamma_1)/3}) - m(d_0)) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} g_{a,b}(x) \\ g_{a,b}^o(x) \end{pmatrix}, \quad (4.47)$$

where  $a = (2C_1 \sigma^2 [(1-p)/p]^{\gamma_1})^{1/2}$  and  $b = m'(d_0)/2$ . In fact, the weak convergence (4.47) holds not only finite dimensionally, but also in the  $L_2$  sense (more rigorously, with the topology of  $L_2$  convergence on compacta times itself) because of the monotonicity of both  $m_{oI}$  and  $m_{oIc}$ .

Denote  $D_{n_2}$  as the set on which  $m_{oI}$  and  $m_{oIc}$  are not equal. Then, given any  $\epsilon > 0$ , there exists a real number  $M > 0$  such that  $n_2^{(1+\gamma_1)/3}(D_{n_2} - d_0) \subset [-M, M]$  holds with probability larger than  $1 - \epsilon$ . This fact can be derived by using the same arguments in [Banerjee et al. \(2007\)](#).

We now consider the ideal second-stage likelihood ratio test statistics  $2 \log \lambda_{oI}$ . By the definition of  $2 \log \lambda_{oI}$ , we have

$$\begin{aligned} 2 \log \lambda_{oI} &= 2 \left[ \frac{1}{2\sigma^2} \sum_{i=1}^{n_2} (Y_{2,i} - m_{oIc}(X_{2,i}))^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n_2} (Y_{2,i} - m_{oI}(X_{2,i}))^2 \right] \\ &= \frac{1}{\sigma^2} \left\{ \sum_{i=1}^{n_2} [(Y_{2,i} - \theta_0) - (m_{oIc}(X_{2,i}) - \theta_0)]^2 - \sum_{i=1}^{n_2} [(Y_{2,i} - \theta_0) - (m_{oI}(X_{2,i}) - \theta_0)]^2 \right\} \\ &= -\frac{2}{\sigma^2} \left[ \sum_{i=1}^{n_2} (Y_{2,i} - \theta_0)(m_{oIc}(X_{2,i}) - \theta_0) - \sum_{i=1}^{n_2} (Y_{2,i} - \theta_0)(m_{oI}(X_{2,i}) - \theta_0) \right] \\ &\quad + \frac{1}{\sigma^2} \sum_{i=1}^{n_2} [(m_{oIc}(X_{2,i}) - \theta_0)^2 - (m_{oI}(X_{2,i}) - \theta_0)^2] \\ &= -\frac{2}{\sigma^2} \sum_{i=1}^{n_2} (Y_{2,i} - m_{oIc}(X_{2,i}))(m_{oIc}(X_{2,i}) - \theta_0) \\ &\quad + \frac{2}{\sigma^2} \sum_{i=1}^{n_2} (Y_{2,i} - m_{oI}(X_{2,i}))(m_{oI}(X_{2,i}) - \theta_0) \\ &\quad - \frac{1}{\sigma^2} \sum_{i=1}^{n_2} [(m_{oIc}(X_{2,i}) - \theta_0)^2 - (m_{oI}(X_{2,i}) - \theta_0)^2] \\ &= \frac{1}{\sigma^2} \sum_{i=1}^{n_2} [(m_{oI}(X_{2,i}) - \theta_0)^2 - (m_{oIc}(X_{2,i}) - \theta_0)^2], \end{aligned}$$

where the last equation is from the fact that both the unconstrained and constrained isotonic regression estimators of  $m$ ,  $m_{oI}$  and  $m_{oIc}$ , are block averages, which ensures

that

$$\sum_{i=1}^{n_2} (Y_{2,i} - m_{oI}(X_{2,i}))(m_{oI}(X_{2,i}) - \theta_0) \text{ and } \sum_{i=1}^{n_2} (Y_{2,i} - m_{oIc}(X_{2,i}))(m_{oIc}(X_{2,i}) - \theta_0)$$

are both equal to 0.

Now denote  $\mathbb{P}_{n_2}$  as the empirical measure of the second-stage covariates  $\{X_{2,i}\}_{i=1}^{n_2}$  and  $P_{n_2}$  as the corresponding uniform probability measure of  $X_{2,i}$ . Then, we have

$$2 \log \lambda_{oI} = \frac{n_2}{\sigma^2} \mathbb{P}_{n_2} [(m_{oI}(x) - \theta_0)^2 - (m_{oIc}(x) - \theta_0)^2] \{x \in D_{n_2}\} = T_1 + T_2,$$

where

$$\begin{aligned} T_1 &= \frac{n_2}{\sigma^2} (\mathbb{P}_{n_2} - P_{n_2}) [(m_{oI}(x) - \theta_0)^2 - (m_{oIc}(x) - \theta_0)^2] \{x \in D_{n_2}\}, \\ T_2 &= \frac{n_2}{\sigma^2} P_{n_2} [(m_{oI}(x) - \theta_0)^2 - (m_{oIc}(x) - \theta_0)^2] \{x \in D_{n_2}\}. \end{aligned}$$

Since  $(1 - 2\gamma_1)/3 < 1/2$ , both  $(m_{oI}(x) - \theta_0)$  and  $(m_{oIc}(x) - \theta_0)$  are  $O_P(n_2^{-\frac{1+\gamma_1}{3}})$  and

$$T_1 = \frac{n_2}{\sigma^2} n_2^{\frac{1-2\gamma_1}{3}} (\mathbb{P}_{n_2} - P_{n_2}) \left\{ \left[ n_2^{\frac{1+\gamma_1}{3}} (m_{oI}(x) - \theta_0) \right]^2 - \left[ n_2^{\frac{1+\gamma_1}{3}} (m_{oIc}(x) - \theta_0) \right]^2 \right\} \{x \in D_{n_2}\},$$

we can show that  $T_1$  converges to 0 in probability by empirical process theory arguments.

From the definition of  $T_2$ , it is equal to

$$\begin{aligned} & \frac{n_2}{\sigma^2} \int_{D_{n_2}} [(m_{oI}(x) - \theta_0)^2 - (m_{oIc}(x) - \theta_0)^2] \frac{n_1^{\gamma_1}}{2C_1} dx \\ &= \frac{1}{2C_1 \sigma^2} \left( \frac{p}{1-p} \right)^{\gamma_1} n_2^{1+\gamma_1} \int_{D_{n_2}} [(m_{oI}(x) - \theta_0)^2 - (m_{oIc}(x) - \theta_0)^2] dx \\ &= \frac{1}{a^2} n_2^{1+\gamma_1} \int_{D_{n_2}} [(m_{oI}(x) - \theta_0)^2 - (m_{oIc}(x) - \theta_0)^2] dx \\ &= \frac{1}{a^2} \int \left\{ \left[ n_2^{\frac{1+\gamma_1}{3}} (m_{oI}(d_0 + tn_2^{-\frac{1+\gamma_1}{3}}) - \theta_0) \right]^2 - \left[ n_2^{\frac{1+\gamma_1}{3}} (m_{oIc}(d_0 + tn_2^{-\frac{1+\gamma_1}{3}}) - \theta_0) \right]^2 \right\} dt \\ &\xrightarrow{d} \frac{1}{a^2} \int [g_{a,b}(t)^2 - g_{a,b}^o(t)^2] dt = \mathbb{D}. \end{aligned}$$

The last equality above is from the change of variable  $x = d_0 + tn_2^{-(1+\gamma_1)/3}$  and the weak convergence is ensured by the weak convergence result (4.47) in the  $L_2$  sense. Therefore, the likelihood ratio test statistic  $2 \log_{\lambda_n}$  converges to  $\mathbb{D}$  and Theorem 4.8.6 holds.

#### 4.8.4 Appendix for TSSIRP

To study the asymptotic properties of  $d_{2,Is}$ , the basic idea is to approximate it by the corresponding statistics in the following ideal two-stage smoothed isotonic regression procedure (ITSSIRP). More specifically, the procedure of ITSSIRP with an ideal second-stage sampling interval is as follows:

1. Set the first-stage sample proportion  $p \in (0, 1)$  and let the first and second-stage sample sizes be  $n_1 = \lfloor np \rfloor$  and  $n_2 = n - n_1$ , respectively, where  $n$  is the total sample size.
2. Let the ideal second-stage sampling interval be  $[L_1, U_1] = [d_0 \pm C_1 n_1^{-\gamma_1}]$  with constants  $C_1 > 0$  and  $\gamma_1 > 0$ .
3. Allocate the equally-spaced or uniformly distributed second-stage design points  $\{X_{2,i}\}_{i=1}^{n_2}$  in  $[L_1, U_1]$  and obtain the corresponding independent second-stage responses  $\{Y_{2,i}\}_{i=1}^{n_2}$ .
4. Compute the ideal second-stage isotonic regression  $m_{oI}$  of  $m$  over  $[L_1, U_1]$ .
5. Smooth  $m_{oI}$  to obtain the ideal second-stage smoothed isotonic regression  $m_{oIs}$  of  $m$  over  $[L_1, U_1]$  with a kernel  $K$  and a bandwidth  $h_n$ .
6. Obtain  $d_{oIs} = m_{oIs}^{-1}(\theta_0)$ , the ideal second-stage smoothed isotonic regression estimator of  $d_0$ .

Comparing ITSSIRP and TSSIRP, we can see that ITSSIRP is the same to TSSIRP except that the second-stage sampling interval of ITSSIRP is centered at  $d_0$  but not  $d_{1,I}$ . Since  $d_{1,I}$  converges to  $d_0$  in probability at the rate of  $n^{1/3}$ , the second-stage sample resolution at  $d_0$  would be the same for both two procedures given  $\gamma_1 < 1/3$ . On the other hand, the estimators  $d_{2,Is}$  and  $d_{oIs}$  are asymptotically determined by the sample resolution around  $d_0$ . Thus, heuristically speaking, the difference  $d_{2,I} - d_{oI}$  would converge to 0 in probability with a fast enough rate. This intuition needs rigorous justification. Now, suppose the above heuristic holds. Then,

the asymptotic distribution of  $d_{2,Is}$  is completely determined by that of  $d_{oIs}$ . Next, we focus on the properties of  $d_{oIs}$ .

In ITSSIRP, the second-stage sample resolution at  $d_0$  is  $n_r$ , which is the same to that in TSSIRP. Suppose the bandwidth for the kernel smoothing has the form  $h_n = cn_r^{-\gamma_2}$  with constants  $c > 0$  and  $\gamma_2 > 0$ . Similarly, the two internal constraints for  $m_{2,Is}$  described in Section 4.3.2 also apply for  $m_{oIs}$ . That is, to make  $m_{oIs}$  effective, we need to pose the requirements that  $\gamma_2 > \gamma_1/(1 + \gamma_1)$  or  $\gamma_2 = \gamma_1/(1 + \gamma_1)$  with  $c < c^*$  and that  $\gamma_2 < 1/3$ , with which we obtain the asymptotic distribution of  $m_{oIs}$  in the following result.

*Theorem 4.8.10.* Suppose the assumption **(B)** holds. For  $\gamma_1 \in (0, 1/2)$  and  $\gamma_2 \in (1/7, 1/3)$ , if  $\gamma_2 > \gamma_1/(1 + \gamma_1)$  or if  $\gamma_2 = \gamma_1/(1 + \gamma_1)$  and  $c < c^*$ , we have

$$n^{(1+\gamma_1)(1-\gamma_2)/2} (m_{oIs}(d_0) - m(d_0)) - n^{(1+\gamma_1)(1-5\gamma_2)/2} C_{n_r}^{-2\gamma_2} c^2 B_1(d_0) \xrightarrow{d} C_{n_r}^{-(1-\gamma_2)/2} N(0, \sigma^2 \kappa/c),$$

where  $B_1(d_0) = m''(d_0) \int t^2 K(t) dt/2$ .

From Theorem 4.8.10, we can see there is an additional technical requirement  $\gamma_2 > 1/7$ , which is essentially used to avoid complicated bias terms. On the other hand, note that now  $\gamma_1$  is allowed to take values larger than or equal to  $1/3$  since the second-stage sampling interval centers at  $d_0$  but not  $d_{1,I}$ . It is natural to pose the restriction that  $\gamma_1 < 1/2$  since no essentially nonparametric one-stage estimation method could achieve the parametric rate. Then, all the above requirements for  $\gamma_1$  and  $\gamma_2$  forms an admissible region for  $(\gamma_1, \gamma_2)$ . It is clear that if  $(\gamma_1, \gamma_2)$  (with a proper  $c$ ) lies in the admissible region, Theorem 4.8.10 holds. The admissible region is enclosed by the black dashed borders in the left panel of Figure 4.10.

Theorem 4.8.10 indicates that, for each  $\gamma_1 \in (0, 1/2)$ , the optimal  $\gamma_2$  (and a proper  $c$ ) to maximize the convergence rate is as follows. For  $\gamma_1 \in (0, 1/4]$ , let  $\gamma_2 = 1/5$ ; for  $\gamma_1 \in (1/4, 1/2)$ , let  $\gamma_2 = \gamma_1/(1 + \gamma_1)$  with  $c < c^*$ . The optimal pair  $(\gamma_1, \gamma_2)$  lies on the red solid segments in the left panel of Figure 4.10 and the optimal convergence rate of  $m_{oIs}(d_0)$  as a function of  $\gamma_1$  is shown in the right panel of Figure 4.10.

The asymptotic distributions on  $m_{oIs}(d_0)$  with the optimal convergence rates follow directly from Theorem 4.8.10.

*Corollary 4.8.11.* Suppose the assumption **(B)** holds. If  $1/4 < \gamma_1 < 1/2$ ,  $\gamma_2 = \gamma_1/(1 + \gamma_1)$  and  $c < c^*(\gamma_2)$ , we have

$$n^{1/2} (m_{oIs}(d_0) - m(d_0)) \xrightarrow{d} C_{n_r}^{-(1-\gamma_2)/2} N(0, \sigma^2 \kappa/c). \quad (4.48)$$

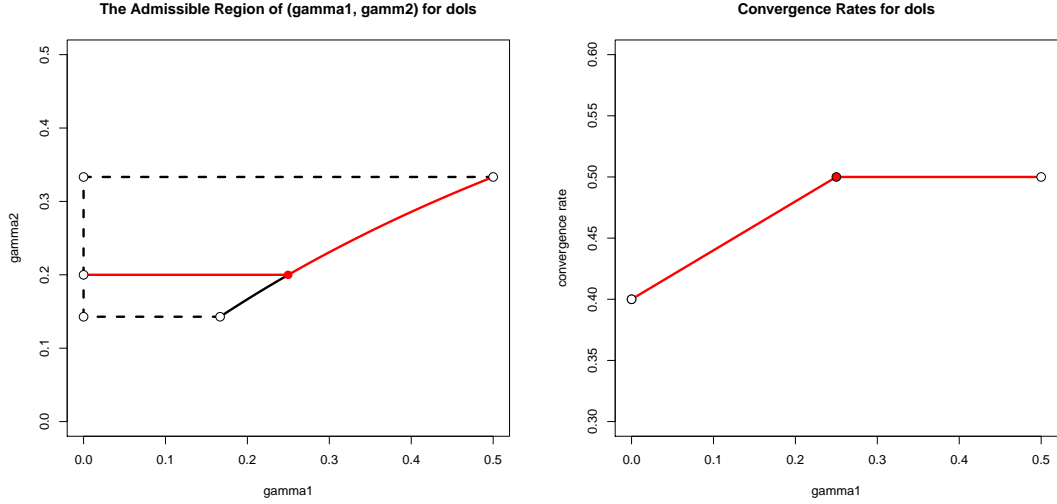


Figure 4.10: The left panel shows the admissible region of  $(\gamma_1, \gamma_2)$  for  $m_{oIs}$  enclosed by the solid and dashed borders. The pair  $(\gamma_1, \gamma_2)$  can be on the solid borders but not on the dashed ones. When  $(\gamma_1, \gamma_2)$  is on the red solid segment and curve in the left panel, the convergence rate of  $m_{oIs}$  for each  $\gamma_1 \in (0, 1/2)$  is optimized. The corresponding optimal rate as a function of  $\gamma_1$  is shown in the right panel.

If  $\gamma_1 = 1/4$ ,  $\gamma_2 = 1/5$  and  $c < c^*(1/5)$ , we have

$$n^{1/2}(m_{oIs}(x_0) - m(d_0)) \xrightarrow{d} C_{n_r}^{-2/5} c^2 B_1(d_0) + C_{n_r}^{-2/5} N(0, \sigma^2 \kappa/c). \quad (4.49)$$

If  $0 < \gamma_1 < 1/4$  and  $\gamma_2 = 1/5$ , we have

$$n^{2(1+\gamma_1)/5}(m_{oIs}(d_0) - m(d_0)) \xrightarrow{d} C_{n_r}^{-2/5} c^2 B_1(d_0) + C_{n_r}^{-2/5} N(0, \sigma^2 \kappa/c). \quad (4.50)$$

From Corollary 4.8.11, we can further derive the asymptotic distributions on  $d_{oIs}$  with the optimal convergence rates.

*Theorem 4.8.12.* Suppose the assumption **(B)** holds. If  $1/4 < \gamma_1 < 1/2$ ,  $\gamma_2 = \gamma_1/(1 + \gamma_1)$  and  $c < c^*(\gamma_2)$ , we have

$$n^{1/2}(d_{oIs} - d_0) \xrightarrow{d} C_{d_{oIs}}(\gamma_2) N(0, 1); \quad (4.51)$$

If  $\gamma_1 = 1/4$ ,  $\gamma_2 = 1/5$  and  $c < c^*(1/5)$ , we have

$$n^{1/2}(d_{oIs} - d_0) \xrightarrow{d} B_{d_{oIs}} + C_{d_{oIs}}(1/5) N(0, 1); \quad (4.52)$$

If  $0 < \gamma_1 < 1/4$  and  $\gamma_2 = 1/5$ , we have

$$n^{2(1+\gamma_1)/5}(d_{oIs} - d_0) \xrightarrow{d} B_{d_{oIs}} + C_{d_{oIs}}(1/5)N(0, 1), \quad (4.53)$$

where  $C_{d_{oIs}}(\gamma) = m'(d_0)^{-1}C_{n_r}^{-(1-\gamma)/2}\sigma\sqrt{\kappa/c}$  and  $B_{d_{oIs}} = -m'(d_0)^{-1}C_{n_r}^{-2/5}c^2B_1(d_0)$ .

*A sketch of the proofs for Theorems 4.8.10 and 4.8.12.* First, we consider the proof for Theorems 4.8.10. Since the second-stage design density is uniform on  $[L_1, U_1]$ , the sample resolution at  $x$  is  $n_r$  and the bandwidth is  $h_n = cn_r^{-\gamma_2}$  with some  $c > 0$  and  $\gamma_2 > 0$ . Then, it is sufficient to show that, for  $\gamma_2 \in (1/7, 1/3)$ ,

$$n_r^{(1-\gamma_2)/2}(m_{oIs}(x) - m(x)) - n_r^{(1-5\gamma_2)/2}c^2B_1(x) \xrightarrow{d} N(0, \kappa\sigma^2/c), \quad (4.54)$$

which is equivalent to

$$(n_r h)^{1/2}(m_{oIs}(x) - m(x) - B_1(d_0)h^2) \xrightarrow{d} N(0, \kappa\sigma^2).$$

Since  $\gamma_2 < 1/3$  ensures that the ideal smoothed isotonic regression estimator  $m_{oIs}(d_0)$  and the ideal kernel smoothing estimator  $m_{os}(d_0)$  are the same with probability going to 1, it is sufficient to show that

$$(n_r h)^{1/2}(m_{os}(x) - m(x) - B_1(d_0)h^2) \xrightarrow{d} N(0, \kappa\sigma^2). \quad (4.55)$$

From the definition of  $m_{os}(d_0)$ , we have

$$\begin{aligned} & m_{os}(d_0) - m(d_0) \\ &= \frac{\sum_{i=1}^{n_2} K((X_{2,i} - d_0)/h)Y_{2,i}}{\sum_{i=1}^{n_2} K((X_{2,i} - d_0)/h)} - m(d_0) \\ &= \frac{\sum_{i=1}^{n_2} K((X_{2,i} - d_0)/h)(Y_{2,i} - m(d_0))}{\sum_{i=1}^{n_2} K((X_{2,i} - d_0)/h)} \\ &= \frac{\sum_{i=1}^{n_2} K((X_{2,i} - d_0)/h)(m(X_{2,i}) - m(d_0)) + \sum_{i=1}^{n_2} K((X_{2,i} - d_0)/h)\epsilon_{2,i}}{\sum_{i=1}^{n_2} K((X_{2,i} - d_0)/h)} \\ &= (T_1 + T_2)/T_3, \end{aligned}$$

where

$$\begin{aligned} T_1 &= (U_1 - L_1) \frac{1}{n_2 h} \sum_{i=1}^{n_2} K((X_{2,i} - d_0)/h)(m(X_{2,i}) - m(d_0)), \\ T_2 &= (U_1 - L_1) \frac{1}{n_2 h} \sum_{i=1}^{n_2} K((X_{2,i} - d_0)/h)\epsilon_{2,i}, \\ T_3 &= (U_1 - L_1) \frac{1}{n_2 h} \sum_{i=1}^{n_2} K((X_{2,i} - d_0)/h). \end{aligned}$$

Then, we have

$$\begin{aligned} m_{os}(d_0) - m(d_0) - B_1(d_0)h^2 &= (T_1/T_3 - B_1(d_0)h^2) + T_2/T_3 \\ &= (T_1 - B_1(d_0)h^2 T_3)/T_3 + T_2/T_3 = [T_1 - B_1(d_0)h^2 + B_1(d_0)h^2(1 - T_3)]/T_3 + T_2/T_3. \end{aligned}$$

It can be shown that

$$\begin{aligned} T_3 - 1 &= O_P(h^3 + (n_2 h)^{-1/2}) \\ T_1 - B_1(d_0)h^2 &= O_P(h^3 + (n_2 h)^{-1/2}h) \\ T_2 &= O_P((n_2 h)^{-1/2}). \end{aligned}$$

Thus, for  $\gamma_2 > 1/7$ , we have

$$(n_2 h)^{-1/2} [T_1 - B_1(d_0)h^2 + B_1(d_0)h^2(1 - T_3)] = o_P(1).$$

By noticing that  $T_3$  converges to 1 in probability, we have

$$(n_2 h)^{-1/2}(m_{os}(d_0) - m(d_0) - B_1(d_0)h^2) = (n_2 h)^{-1/2}T_2 + o_P(1).$$

Then, from  $U_1 - L_1 = o_P(1)$ , we have

$$(U_1 - L_1)^{1/2}(n_2 h)^{-1/2}(m_{os}(d_0) - m(d_0) - B_1(d_0)h^2) = (U_1 - L_1)^{1/2}(n_2 h)^{-1/2}T_2 + o_P(1).$$

It can be shown that  $(U_1 - L_1)^{1/2}(n_2 h)^{-1/2}T_2$  converges weakly to  $N(0, \kappa\sigma^2)$  simply by involving the classical triangular central limit theorem. Thus, from the definition  $n_r = n_2/(U_1 - L_1)$ , we have the desired result (4.55).

We next consider the first asymptotic result (4.51) in Theorem 4.8.12 and the other two cases can be handled in the same way. If  $1/4 < \gamma_1 < 1/2$ ,  $\gamma_2 = \gamma_1/(1 + \gamma_1)$  and  $c < 2^{-\gamma_2} C_1^{1-\gamma_2} [(1-p)/p]^{\gamma_2}$ ,  $\gamma_2$  varies within  $(1/5, 1/3)$ . Then, the bias term in the above result (4.54) disappears. Thus, we have

$$n_r^{(1-\gamma_2)/2} (m_{oIs}(x) - m(x)) \xrightarrow{d} N(0, \kappa\sigma^2/c).$$

Similar to the sketch of the proof of Theorem 4.2.5 in Appendix 4.8.2, we can obtain (4.51). □



## BIBLIOGRAPHY

## BIBLIOGRAPHY

- Abrevaya, J., and J. Huang (2005), On the bootstrap of the maximum score estimator, *Econometrica*, *73*, 1175–1204.
- Anbar, D. (1977), A modified Robbins-Monro procedure approximating the zero of a regression function from below, *Ann. Statist.*, *5*(1), 229–234.
- Ayer, M., H. D. Brunk, G. M. Ewing, W. T. Reid, and E. Silverman (1955), An empirical distribution function for sampling with incomplete information, *The Annals of Mathematical Statistics*, *24*, 641–647.
- Bambos, N., and G. Michailidis (2004), Queueing and scheduling in random environments, *Adv. in Appl. Probab.*, *36*(1), 293–317.
- Banerjee (2000), Likelihood ratio inference in regular and non-regular problems, Ph.D. thesis, University of Washington.
- Banerjee, M. (2007), Likelihood based inference for monotone response models, *Annals of Statistics*, *35*, 931–956.
- Banerjee, M. (2009), Inference in exponential family regression models under certain shape constraints, in *Advances in Multivariate Statistical Methods, Statistical Science and Interdisciplinary Research*, vol. 4, pp. 249–272, World Scientific.
- Banerjee, M., and J. Wellner (2001), Likelihood ratio tests for monotone functions, *Annals of Statistics*, *29*, 1699 – 1731.
- Banerjee, M., and J. A. Wellner (2005), Confidence intervals for current status data, *Scand. J. Statist.*, *32*(3), 405–424.
- Banerjee, M., S. Mishra, and D. Mukherjee (2007), Estimating the regression function in a semiparametric logistic model using likelihood ratio tests.
- Barlow, R. E., D. J. Bartholomew, J. M. Bremner, and H. D. Brunk (1972), *Statistical inference under order restrictions: The theory and application of isotonic regression*, xii+388 pp., John Wiley & Sons, London-New York-Sydney, Wiley Series in Probability and Mathematical Statistics.
- Billingsley, P. (1995), *Probability and measure*, Wiley Series in Probability and Mathematical Statistics, third ed., xiv+593 pp., John Wiley & Sons Inc., New York, a Wiley-Interscience Publication.

- Brunk, H. D. (1970), Estimation of isotonic regression, in *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)*, pp. 177–197, Cambridge Univ. Press, London.
- Durot, C. (2002), Sharp asymptotics for isotonic regression, *Probab. Theory Related Fields*, *122*(2), 222–240.
- Durot, C. (2007), On the  $\mathbb{L}_p$ -error of monotonicity constrained estimators, *Ann. Statist.*, *35*, 10801104.
- Durot, C. (2008), Monotone nonparametric regression with random design, *Math. Methods Statist.*, *17*(4), 327–341.
- Fan, J., and I. Gijbels (1996), *Local polynomial modelling and its applications*, *Monographs on Statistics and Applied Probability*, vol. 66, xvi+341 pp., Chapman & Hall, London.
- Gasser, T., L. Sroka, and C. Jennen-Steinmetz (1986), Residual variance and residual pattern in nonlinear regression, *Biometrika*, *73*(3), 625–633.
- Gezmu, M., and N. Flournoy (2006), Group up-and-down designs for dose-finding, *J. Statist. Plann. Inference*, *136*(6), 1749–1764.
- Groeneboom, P., and J. A. Wellner (1992), *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhauser-Verlag.
- Groeneboom, P., and J. A. Wellner (2001), Computing Chernoff’s distribution, *J. Comput. Graph. Statist.*, *10*(2), 388–400.
- Groeneboom, P., G. Hooghiemstra, and H. P. Lopuhaä (1999), Asymptotic normality of the L1 error of the grenander estimator, *The Annals of Statistics*, *27*, 1316–1347.
- Groeneboom, P., M. H. Maathuis, and J. A. Wellner (2008a), Current status data with competing risks: Consistency and rates of convergence of the mle, *The Annals of Statistics*, *36*, 1031–1063.
- Groeneboom, P., M. H. Maathuis, and J. A. Wellner (2008b), Current status data with competing risks: Limiting distribution of the mle, *The Annals of Statistics*, *36*, 1064–1089.
- Gruet, M.-A. (1996), A nonparametric calibration analysis, *Ann. Statist.*, *24*(4), 1474–1492.
- Hu, T. C., F. Móricz, and R. L. Taylor (1989), Strong laws of large numbers for arrays of rowwise independent random variables, *Acta Math. Hungar.*, *54*(1-2), 153–162.
- Huang, J. (1996), Efficient estimation for the proportional hazards model with interval censoring, *The Annals of Statistics*, *24*, 540–568.

- Ivanova, A., A. Montazer-Haghighi, S. G. Mohanty, and S. D. Durham (2003), Improved up-and-down designs for phase I trials, *Stat. Med.*, *22*, 69–82.
- Jewell, N. P., M. van der Laan, and T. Henneman (2003), Nonparametric estimation from current status data with competing risks, *Biometrika*, *90*, 183–197.
- Kim, J., and D. Pollard (1990), Cube root asymptotics, *The Annals of Statistics*, *18*, 191–219.
- Kosorok, M. R. (2008), Bootstrapping the grenander estimator, in *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, pp. 282–292, Institute of Mathematical Statistics: Hayward, CA.
- Lai, T. L. (2003), Stochastic approximation, *Ann. Statist.*, *31*(2), 391–406, dedicated to the memory of Herbert E. Robbins.
- Leurgans, S. (1982), Asymptotic distributions of slope-of-greatest-convex-minorant estimators, *Ann. Statist.*, *10*(1), 287–296.
- Li, Q., and J. S. Racine (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.
- Lin, D. Y., D. Oakes, and Z. Ying (1998), Additive hazards regression with current status data, *Biometrika*, *85*, 289–298.
- Ma, S., and M. R. Kosorok (2005), Penalized log-likelihood estimation for partly linear transformation models with current status data, *The Annals of Statistics*, *33*, 2256–2290.
- Maathuis, M., and M. Hudgens (2010), Nonparametric inference for competing risks current status data with continuous, discrete or grouped observation times, *Biometrika (to appear) arXiv:0909.4856v2*.
- Makowski, G. (1975), A rate of convergence for a nondecreasing regression estimator, *Bull. Inst. Math. Acad. Sinica*, *3*(1), 61–64.
- Makowski, G. G. (1973), Laws of the iterated logarithm for permuted random variables and regression applications, *Ann. Statist.*, *1*, 872–887.
- Mammen, E. (1991), Estimating a smooth monotone regression function, *The Annals of Statistics*, *19*, 724–740.
- Morgan, B. J. T. (1985), The cubic logistic model for quantal assay data, *Applied Statistics*, *34*(2), 105–113.
- Morgan, B. J. T. (1992), *Analysis of Quantal Response Data*, Chapman & Hall, London Glasgow New York Tokyo Melbourne Madras.
- Mukerjee, H. (1988), Monotone nonparametric regression, *The Annals of Statistics*, *16*, 741–750.

- Müller, H.-G., and T. Schmitt (1990), Choice of number of doses for maximum likelihood estimation of the ED50 for quantal dose-response data, *Biometrics*, *46*, 117–129.
- Müller, H.-G., and U. Stadtmüller (1987), Estimation of heteroscedasticity in regression analysis, *The Annals of Statistics*, *15*, 610–625.
- Osborne, C. (1991), Statistical calibration: A review, *International Statistical Review*, *59*, 309–336.
- Pollard, D. (2002), *A User's Guide to Measure Theoretic Probability*, Cambridge University Press.
- Prakasa Rao, B. L. S. (1969), Estimation of a unimodal density, *Sankhya Ser. A*, *31*, 23–36.
- Rai, K., and J. V. Ryzin (1981), A generalized multihit dose-response model for low-dose extrapolation, *Biometrics*, *37*, 341–352.
- Robbins, H., and S. Monro (1951), A stochastic approximation method, *Ann. Math. Statistics*, *22*, 400–407.
- Robertson, T., F. T. Wright, and R. L. Dykstra (1988), *Order restricted statistical inference*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, xx+521 pp., John Wiley & Sons Ltd., Chichester.
- Rosenberger, W. F. (1996), New directions in adaptive designs, *Statist. Sci.*, *11*, 137–149.
- Rosenberger, W. F., and L. M. Haines (2002), Competing designs for phase I clinical trials: a review, *Stat. Med.*, *21*, 2757–2770.
- Schick, A., and Q. Yu (2000), Consistency of the gmle with mixed case interval-censored data, *Scandinavian Journal of Statistics*, *27*, p. 45 Length: 11 pages.
- Sen, B., and M. Banerjee (2006), A pseudo-likelihood method for analyzing interval censored data, *Biometrika*.
- Sen, B., M. Banerjee, and M. B. Woodroffe (2010), Inconsistency of bootstrap: the grenander estimator, *Annals of Statistics*, *38*, 1953–1977.
- Shiboski, S. C. (1998), Generalized additive models for current status data, *Lifetime Data Analysis*, *4*, 29–50.
- Shiryayev, A. N. (1995), *Probability*, second ed., Springer-Verlag.
- Staniswalis, J. G., and V. Cooper (1988), Kernel estimates of dose response, *Biometrics*, *44*(4), 1103–1119.

- Stylianou, M., and N. Flournoy (2002), Dose finding using the biased coin up-and-down design and isotonic regression, *Biometrics*, *58*(1), 171–177.
- Sun, J. (1999), A nonparametric test for current status data with unequal censoring, *Journal of the Royal Statistical Society: Series B*, *61*, 243–250.
- Turnbull, B. W. (1976), The empirical distribution function with arbitrarily grouped, censored and truncated data, *Journal of the Royal Statistical Society. Series B (Methodological)*, *38*(3), 290–295.
- van der Vaart, A. (1991), On differentiable functionals, *The Annals of Statistics*, *19*, 178–204.
- van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press.
- van der Vaart, A. W., and J. A. Wellner (1996), *Weak Convergence and Empirical Processes*, Springer.
- Wang, X., and J. Shen (2010), A class of grouped brunk estimators and penalized spline estimators for monotone regression, *Biometrika*, *97*, 585–601.
- Wright, F. T. (1981), The asymptotic behavior of monotone regression estimates, *Ann. Statist.*, *9*(2), 443–448.
- Yu, Q., A. Schick, L. Li, and G. Y. C. Wong (1998), Asymptotic properties of the gmle in the case 1 interval-censorship model with discrete inspection times, *The Canadian Journal of Statistics*, *26*, 619–627.
- Zhang, R., J. Kim, and M. Woodroffe (2001), Asymptotic analysis of isotonic estimation for grouped data, *JSPI*, *98*, 107–117.