Submitted to the Annals of Statistics
arXiv: math.PR/0000000

A TWO-STAGE HYBRID PROCEDURE FOR ESTIMATING AN INVERSE REGRESSION FUNCTION

BY RUNLONG TANG, MOULINATH BANERJEE* AND GEORGE MICHAILIDIS†

University of Michigan

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

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 ϵ an error term with zero mean and finite variance σ^2. Further, without loss of generality it is assumed that f is isotonic and $X \in [0, 1]$. We are interested in estimating $d_0 = f^{-1}(\theta_0)$ for some θ_0 in 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 [16]. 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, briefly described below and more extensively in Section 5, we introduce a two-stage design for estimating d_0. Specifically, we consider a complex queueing system operating in discrete (slotted) time

*Supported in part by NSF grant DMS-07-05288
†Supported in part by NIH grant P41-18627
AMS 2000 subject classifications: Primary 62G09, 62G20; secondary 62G07
Keywords and phrases: two-stage estimator, bootstrap, adaptive design, asymptotic properties

comprising multiple customer classes and a service resource that can be allocated to them. Further, the efficiency of service is modulated by an external to the system stochastic mechanism. The main issue is to efficiently allocate the service resource so as to maximize the system's throughput (average number of customers processed per unit of time). Such a maximizing allocation scheme is described in Bambos and Michailidis [2]. Unfortunately the customers' average delay which is an important quality of service metric of the performance of the system is not analytically tractable and hence one needs to resort to expensive simulations to obtain it. The average delay (over all customer classes) as a function of the system's loading (number of customers arriving per unit of time) is depicted in Figure 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. 3, 5, 10 and 15 time units), so as to be able to decide whether to upgrade the available resources, or give higher priority to specific customer classes.

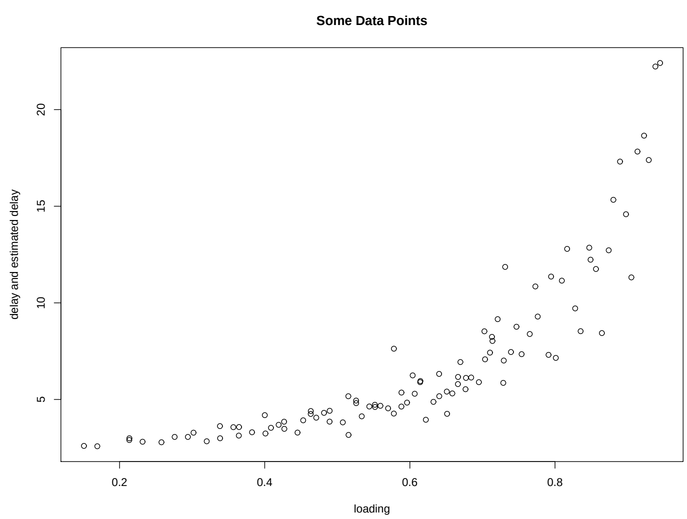

FIG 1. *The average delay as a function of system's loading.*

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 θ_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.

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 [18] gives a comprehensive review of this topic and Gruet [10] 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 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 [16], multihit [19], cubic logistic model [15]), or by nonparametric models, for which kernel estimates [25] and isotonic regression [26] 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 [22] and Rosenberger and Haines [23] for comprehensive reviews of the subject). Stylianou and Flournoy [26] 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. [11] 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. This partially motivates the usage 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 Robins-Monro procedure [20] for finding d_0. It corresponds to a stochastic version of Newton's scheme for root finding problems. Anbar [1] considered a modified Robbins-Monro type procedure approximating the root from one side. A good review of this area is provided in Lai [12], 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 [17].

The remainder of the paper is organized as follows: Section 2 describes the two-stage procedures. The asymptotic properties of the two-stage estimators are obtained in Section 3. Simulation studies and data analysis are presented in Sections 4 and 5, respectively. We close with a discussion in Section 6, which is followed by an appendix containing technical details.

2. Two-Stage Procedures.

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

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. Sample n design points $X_i \in [0,1]$ from a Lebesgue density g positive at $d_0 \in (0,1)$ and subsequently obtain the corresponding responses that are generated according to $Y_i = f(X_i) + \epsilon_i,\ i = 1, 2, \cdots, n$, where f is in $\mathscr{F}_0$, a class of increasing real functions on $[0,1]$ with positive first derivatives at d_0 and ϵ_i are independently and identically distributed (iid) random errors independent of the X_i's, with mean zero and constant variance σ^2.
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. [21]).
3. Estimate d_0 as $\hat{d}_n^{(1)} = \hat{f}^{-1}(\theta_0) = \inf\{x \in [0,1] : \hat{f}(x) \geq \theta_0\}$, where $\theta_0 = f(d_0)$.

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

THEOREM 2.1. *For $f \in \mathscr{F}_0$, we have*

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

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

REMARK 2.1. 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 significantly less dispersion than the standard normal (Its standard deviation is about 0.51.), and its quantiles have been tabled in Groeneboom and Wellner [9].

The proof of this result follows by minor adaptation of the arguments from Theorem 1 in Banerjee and Wellner [3]. Hence, an approximate confidence interval for d_0 with significance level $1 - 2\alpha$ can be constructed as follows

$$[\hat{d}_n^{(1)} - n^{-1/3}\hat{C}q_\alpha, \hat{d}_n^{(1)} + n^{-1/3}\hat{C}q_\alpha] \cap (0, 1), \tag{2.1}$$

where q_α denotes the upper α quantile of Chernoff's distribution and $\hat{C}$ 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.

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. We focus on the simple model $Y = f(X) + \epsilon$, with ϵ independent of X, in order to simplify the presentation and at the same time highlight the main technical challenges that need to be resolved. Nevertheless, some extensions of the model are briefly discussed in the conclusions section.

Assume 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 2.1. (Note that np represents its integer part whenever it is not an integer.)
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 $i = 1, 2, \cdots, n_2$, with $\{\epsilon_i'\}$ and $\{\epsilon_i''\}$ being iid random errors with mean zero and constant variance σ^2, mutually independent and also independent of $\{X_i\}$ and $\{\epsilon_i\}$.
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 $\tilde{d}_{n_2}^{(2)} = (\theta_0 - \hat{\beta}_0)/\hat{\beta}_1$.

Asymptotic properties of $\tilde{d}_{n_2}^{(2)}$ will be established in Subsection 3.1. For example, when f is locally linear at d_0 (i.e. $f''(d_0) = 0$) and $\gamma \in (1/8, 1/3)$, we have

$$n^{1/2}(\tilde{d}_n^{(2)} - d_0) \xrightarrow{d} \frac{\sigma}{f'(d_0)(1-p)^{1/2}} N(0,1), \tag{2.2}$$

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 rate of convergence of the one-stage isotonic regression estimator.

However, when constructing confidence intervals from the asymptotic results like (2.2), 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.

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 $\{Y_i'^{\star}\}_{i=1}^{n_2}$ and $\{Y_i''^{\star}\}_{i=1}^{n_2}$, 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 $\tilde{d}_n^{(2)\star}$, and calculate the corresponding root $R_n^{\star} := n^{1/2}(\tilde{d}_n^{(2)\star} - \tilde{d}_n^{(2)})$.
3. Repeat the previous step B times to obtain $\{R_n^{\star b}\}_{b=1}^{B}$. Subsequently, calculate the lower and upper α quantiles, $q_l^{\star}$ and $q_u^{\star}$, of $\{R_n^{\star b}\}_{b=1}^{B}$. Finally, construct a $1-2\alpha$ bootstrapped Wald-type confidence interval for d_0 as

$$[\tilde{d}_n^{(2)} - n^{-1/2} q_u^{\star},\ \tilde{d}_n^{(2)} - n^{-1/2} q_l^{\star}]. \tag{2.3}$$

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 3.2. For example, when f is locally linear at d_0 and $\gamma \in (0, 1/4)$, we have

$$n^{1/2}(\tilde{d}_n^{(2)\star} - \tilde{d}_n^{(2)}) \xrightarrow{d^{\star}} \frac{\sigma}{f'(d_0)(1-p)^{1/2}} N(0,1),\ (P - a.s.), \tag{2.4}$$

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

From (2.2) and (2.4), the strong consistency of the bootstrapped estimator $\tilde{d}_n^{(2)\star}$ is ensured for $\gamma \in (1/8, 1/4)$. However, weak consistency is achieved for a broader range of γ values and does not require the assumption of local linearity (see Theorems 3.6 and Theorem 3.7). Therefore, the bootstrapped procedure is theoretically validated under certain conditions.

REMARK 2.2. Both the two-stage estimator and its bootstrapped variant rely on the choice of a number of tuning parameters: p, γ and K. Procedures for their selection will be discussed in Section 4.

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 $\tilde{d}_n^{(2)}$.

3.1. *Two-Stage Estimator.* According to the two-stage procedure, we have that

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

It is easy to see that

$$\hat{\beta}_0 = \frac{1}{2n_2}\sum_{i=1}^{n_2}(Y_i'' + Y_i') - \hat{d}_{n_1}^{(1)}\hat{\beta}_1, \quad \hat{\beta}_1 = \frac{1}{2Kn_1^{-\gamma}n_2}\sum_{i=1}^{n_2}(Y_i'' - Y_i'). \tag{3.1}$$

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

$$\tilde{d}_n^{(2)} = \frac{\theta_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{\theta_0 - \frac{1}{2n_2}\sum_{i=1}^{n_2}(Y_i'' + Y_i')}{\hat{\beta}_1} + \hat{d}_{n_1}^{(1)}. \tag{3.2}$$

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 two classes of underlying functions:

$$\mathscr{F}_1 = \{f \in \mathscr{F}_0 : f''(d_0) \neq 0, f'''(x) \text{ is uniformly bounded in a neighborhood of } d_0\}$$

and

$$\mathscr{F}_2 = \{f \in \mathscr{F}_0 : f''(d_0) = 0, f'''(d_0) \neq 0, f^{(4)}(x) \text{ is uniformly bounded in a neighborhood of } d_0\}.$$

REMARK 3.1. A function in $\mathscr{F}_2$ is exactly locally linear at d_0 while that in $\mathscr{F}_1$ is not. Notice that both $\mathscr{F}_2$ and $\mathscr{F}_1$ 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 $\mathscr{F}_2$ if $d_0 = b$ and to $\mathscr{F}_1$ otherwise. Obviously, the size of $\mathscr{F}_2$ is much smaller than that of $\mathscr{F}_1$. However, the asymptotic results for $f \in \mathscr{F}_2$ should also provide good approximations for functions that are approximately linear in the vicinity of d_0. Hence, the class $\mathscr{F}_2$ is also of interest.

We consider next the asymptotic properties of $\tilde{d}_n^{(2)}$ for $f \in \mathscr{F}_1$. We start with the consistency of the two-stage estimator.

LEMMA 3.1. *For $f \in \mathscr{F}_1$ and $\gamma \in (0, 1/2)$, we have:*

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

Based on Lemma 3.1, we derive the asymptotic distribution of $\tilde{d}_n^{(2)}$ given in the next theorem.

THEOREM 3.2. *For $f \in \mathscr{F}_1$, we have*

$$\begin{aligned}
n^{2\gamma}(\tilde{d}_n^{(2)} - d_0) &\xrightarrow{d} C_1, && \textit{for } \gamma \in (0, 1/4) \\
n^{1/2}(\tilde{d}_n^{(2)} - d_0) &\xrightarrow{d} C_1 + C_2 Z_1, && \textit{for } \gamma = 1/4 \\
n^{1/2}(\tilde{d}_n^{(2)} - d_0) &\xrightarrow{d} C_2 Z_1, && \textit{for } \gamma \in (1/4, 1/3) \\
n^{1/2}(\tilde{d}_n^{(2)} - d_0) &\xrightarrow{d} C_2 Z_1 + C_3 \mathbb{Z} Z_2, && \textit{for } \gamma = 1/3 \\
n^{(5/6-\gamma)}(\tilde{d}_n^{(2)} - d_0) &\xrightarrow{d} C_3 \mathbb{Z} Z_2, && \textit{for } \gamma \in (1/3, 1/2)
\end{aligned}$$

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 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 3.2. Theorem 3.2 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. On the other hand, for the boundary values of $\gamma = 1/4$ and $1/3$, there exists asymptotic bias in the former case, while in the latter case the asymptotic variance increases. For $\gamma > 1/3$, the local interval $[L, U]$ is too short to contain d_0 with probability going to 1, whereas for $\gamma < 1/4$ it proves too long for a satisfactory linear approximation of f in the vicinity of d_0. The results theoretically suggest selecting γ in the $(1/4, 1/3)$ range.

Next, we examine the case where $f \in \mathscr{F}_2$. Analogous to Lemma 3.1 and Theorem 3.2, we establish consistency and derive the asymptotic distribution of the two-stage estimator.

LEMMA 3.3. *For $f \in \mathscr{F}_2$ and $\gamma \in (0, 1/2)$, we have*

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

THEOREM 3.4. *For $f \in \mathscr{F}_2$, we have*

$$\begin{aligned} n^{1/2}(\tilde{d}_n^{(2)} - d_0) &\xrightarrow{d} C_2 Z_1, && \textit{for } \gamma \in (1/8, 1/3) \\ n^{1/2}(\tilde{d}_n^{(2)} - d_0) &\xrightarrow{d} C_2 Z_1 + C_3 \mathbb{Z} Z_2, && \textit{for } \gamma = 1/3 \\ n^{(5/6-\gamma)}(\tilde{d}_n^{(2)} - d_0) &\xrightarrow{d} C_3 \mathbb{Z} Z_2, && \textit{for } \gamma \in (1/3, 1/2) \end{aligned}$$

where C_2, C_3, Z_1, Z_2, $\mathbb{Z}$ and C are as in Theorem 3.2.

REMARK 3.3. Comparing Theorem 3.2 and Theorem 3.4, we see that the asymptotic results 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.

REMARK 3.4. Notice also that for the $\mathscr{F}_2$ function class, the $n^{1/2}$ rate of convergence is achieved for a slightly larger range of values for γ. 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 3.5. The case of $\gamma < 1/8$ has been omitted, since it involves a Taylor expansion of f up to its fifth derivative. Nevertheless, in principle no other technical challenges are in play.

3.2. *Bootstrapped Two-Stage Estimator.* We consider next the asymptotic properties of the bootstrapped two-stage estimator, which is:

$$\tilde{d}_n^{(2)\star} = \frac{\theta_0 - \hat{\beta}_0^\star}{\hat{\beta}_1^\star} = \frac{f(d_0) - \frac{1}{2n_2}\sum_{i=1}^{n_2}(Y_i''^\star + Y_i'^\star)}{\hat{\beta}_1^\star} + \hat{d}_{n_1}^{(1)}, \tag{3.3}$$

where

$$\hat{\beta}_0^\star = \frac{1}{2n_2}\sum_{i=1}^{n_2}(Y_i''^\star + Y_i'^\star) - \hat{d}_{n_1}^{(1)}\hat{\beta}_1^\star, \quad \hat{\beta}_1^\star = \frac{1}{2Kn_1^{-\gamma}n_2}\sum_{i=1}^{n_2}(Y_i''^\star - Y_i'^\star). \tag{3.4}$$

The data generation mechanism is rather involved and presents a number of technical subtleties that need to be carefully addressed, in order to establish the asymptotic properties of the bootstrapped estimator rigorously. In summary, the design points and random errors involved in the sampling mechanism come from sequences but not from triangular arrays, whose treatment requires special care. We present next the necessary probabilistic framework needed to establish the results.

Let X be a continuous random variable with a positive density function g on the interval $[0,1]$, and ϵ be a continuous random variable in $\mathbb{R}$ with mean 0, positive constant variance σ^2 and finite third absolute moment τ. It is well known that there exists a probability space $(\Omega, \mathscr{A}, P)$, on which $\{X_i\}_{i=1}^{\infty}$ are iid X, $\{\epsilon_i\}_{i=1}^{\infty}$, $\{\epsilon_i^{'}\}_{i=1}^{\infty}$ and $\{\epsilon_i^{''}\}_{i=1}^{\infty}$ iid ϵ, and $\{X_i\}$, $\{\epsilon_j\}$, $\{\epsilon_k^{'}\}$ and $\{\epsilon_l^{''}\}$ are mutually independent. For example, we can first construct product probability spaces $(\Omega_i, \mathscr{A}_i, P_i)$,with $i = 1, 2, 3$ and 4, for $\{\tilde{X}_i\}_{i=1}^{\infty}$ being iid copies of X, and $\{\tilde{\epsilon}_i\}_{i=1}^{\infty}$, $\{\tilde{\epsilon}_i^{'}\}_{i=1}^{\infty}$ and $\{\tilde{\epsilon}_i^{''}\}_{i=1}^{\infty}$ being iid copies of ϵ, and then obtain the product probability space of interest $(\Omega, \mathscr{A}, P)$ with $\Omega = \times_{i=1}^{4}\Omega_i$, $\mathscr{A} = \otimes_{i=1}^{4}\mathscr{A}_i$ and $P = \times_{i=1}^{4}P_i$. Finally, let $X_i(\omega) = \tilde{X}_i(\omega_1)$, $\epsilon_i(\omega) = \tilde{\epsilon}_i(\omega_2)$, $\epsilon_i'(\omega) = \tilde{\epsilon}_i'(\omega_3)$, and $\epsilon_i''(\omega) = \tilde{\epsilon}_i''(\omega_4)$, for $\omega = (\omega_1, \omega_2, \omega_3, \omega_4) \in \Omega$.

Given $\omega \in \Omega$, $\{X_i(\omega)\}_{i=1}^{\infty}$, $\{\epsilon_i(\omega)\}_{i=1}^{\infty}$, $\{\epsilon_i^{'}(\omega)\}_{i=1}^{\infty}$ and $\{\epsilon_i^{''}(\omega)\}_{i=1}^{\infty}$ represent *fixed sequences* of real numbers. Let n_1 and $2n_2$ with $n = n_1 + 2n_2$ denote the first and second stage sample sizes, respectively.

According to the sampling mechanism used in the bootstrapped procedure, the data obtained from the first stage are given by $\{(X_i(\omega), Y_i(\omega))\}_{i=1}^{n_1}$, where $Y_i(\omega) = f(X_i(\omega)) + \epsilon_i(\omega)$ for $i = 1, 2, \cdots, n_1$, 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 $\{U(\omega), Y_i^{''}(\omega)\}$, where $Y_i^{'}(\omega) = f(L(\omega)) + \epsilon_i^{'}(\omega)$ and $Y_i^{''}(\omega) = f(U(\omega)) + \epsilon_i^{''}(\omega)$ for $i = 1, 2, \cdots, n_2$, 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 equivalent to bootstrapping $\{\epsilon_i^{'}(\omega)\}_{i=1}^{n_2}$ and $\{\epsilon_i^{''}(\omega)\}_{i=1}^{n_2}$ to get $\{\epsilon_i^{'\star}\}_{i=1}^{n_2}$ and $\{\epsilon_i^{''\star}\}_{i=1}^{n_2}$, so that $Y_i^{'\star} = f(L(\omega)) + \epsilon_i^{'\star}$ and $Y_i^{''\star} = f(U(\omega)) + \epsilon_i^{''\star}$ for $i = 1, 2, \cdots, n_2$. Note that given ω and n, the bootstrapped second-stage random errors $\{\epsilon_i^{'\star}\}_{i=1}^{n_2}$ or $\{\epsilon_i^{''\star}\}_{i=1}^{n_2}$ are iid uniform random variables on $\{\epsilon_i^{'}(\omega)\}_{i=1}^{n_2}$ or $\{\epsilon_i^{''}(\omega)\}_{i=1}^{n_2}$. Finally, the bootstrapped estimate $\tilde{d}_n^{(2)\star}$ is calculated from $\{(L(\omega), Y_i^{'\star}), (U(\omega), Y_i^{''\star})\}_{i=1}^{n_2}$, the bootstrapped second-stage data.

The upshot of this construction is that given ω and n, the randomness of $\tilde{d}_n^{(2)\star}$ comes from the bootstrapping step; given ω, with n increasing more design points and random errors are sampled from the four fixed sequences. On the other hand, the bootstrapped random errors $\{\epsilon_i^{'\star}\}_{i=1}^{n_2}$ and $\{\epsilon_i^{''\star}\}_{i=1}^{n_2}$ form triangular arrays.

Under the above theoretical setting, it turns out that the asymptotic behavior of $\tilde{d}_n^{(2)\star}$ for both functional classes $\mathscr{F}_1$ and $\mathscr{F}_2$ is the same. Hence, for the remainder of this Section, it is assumed that f belongs to $\mathscr{F} = \mathscr{F}_1 \cup \mathscr{F}_2$.

The following lemma shows the strong consistency of $\hat\beta_1$ and the conditional weak consistency of $\hat\beta_1^\star$.

LEMMA 3.5. *For $f \in \mathscr{F}$ and $\gamma \in (0, 1/2)$, we have*

$$\hat\beta_1 \to f'(d_0),\ (P - a.s.),\ \ \hat\beta_1^\star \xrightarrow{P^\star} f'(d_0),\ (P - a.s.),$$

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

We establish next strong and weak consistency results for the bootstrapped two-stage estimator. The latter result provides a unified framework for using the estimator in practice, since it covers both functional classes ($\mathscr{F}_1$ and $\mathscr{F}_2$) and allows γ to go up to $1/3$.

THEOREM 3.6. *For $f \in \mathscr{F}$ and $\gamma \in (0, 1/4)$, we have*

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

where $\xrightarrow{d^\star}$ denotes convergence in probability conditional on a given ω, and C_2 and Z_1 are as in Theorem 3.2. That is,

$$\sup_{t\in\mathbb{R}} |P^\star\left(n^{1/2}(\tilde d_n^{(2)\star} - \tilde d_n^{(2)}) \le t\right) - P\left(C_2 Z_1 \le t\right)| \xrightarrow{a.s} 0,$$

where $P^\star$ denotes the probability of the bootstrapped data conditional on the original data.

THEOREM 3.7. *For $f \in \mathscr{F}$, $\gamma \in (0, 1/3)$ and $t \in \mathbb{R}$, we have*

$$\sup_{t\in\mathbb{R}} |P^\star\left(n^{1/2}(\tilde d_n^{(2)\star} - \tilde d_n^{(2)}) \le t\right) - P\left(C_2 Z_1 \le t\right)| \xrightarrow{P} 0,$$

where $P^\star$ is as in the previous Theorem and C_2 and Z_1 are as in Theorem 3.2.

REMARK 3.6. Comparing Theorem 3.6 with Theorem 3.2 and Theorem 3.4, we see that the bootstrapped estimator is strongly consistent for $f \in \mathscr{F}_2$ and $\gamma \in (1/8, 1/4)$. However, for $f \in \mathscr{F}_1$, our current result does not ensure the strong consistency of the bootstrapped estimator. The fundamental reason is that $n^{1/4}$ is the currently available upper bound on the almost sure convergence rate of the first-stage estimator of d_0. If the upper bound could be increased to n^a for some $a \in (1/4, 1/3)$, it would have been possible to establish the strong consistency of the bootstrapped estimator for $f \in \mathscr{F}_1$ and $\gamma \in (1/4, \alpha)$. (For technical details see Lemma A.1 and the proof of Lemma A.2.)

On the other hand, comparing Theorem 3.7 with Theorem 3.2 and Theorem 3.4 shows that the bootstrapped estimator is weakly consistent for $f \in \mathscr{F}_2$ and $\gamma \in (1/8, 1/3)$ or for $f \in \mathscr{F}_1$ and $\gamma \in (1/4, 1/3)$.

Finally, comparing Theorem 3.6 and Theorem 3.7 shows that a weakening of the mode of convergence allows γ to go up to 1/3. Hence, weak consistency of the bootstrapped estimator for functions in the $\mathscr{F}_1$ class is established.

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 σ^2 need to be obtained; the resulting procedures are called POSP, PTSP and PBTSP, respectively (Practical OSP, TSP and BTSP). For σ^2, we employ the nonparametric estimator proposed by Gasser et al. [8], 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

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

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 $i = 2, 3, \cdots, n-1$. An estimate of $f'(d_0)$ is obtained through the local quadratic regression estimator proposed by Fan and Gijbels [7], 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 $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ given by

$$\hat{\beta} = argmin_{\beta \in \mathbb{R}^2} \sum_{i=1}^{n} \left[Y_i - \sum_{j=0}^{2} \beta_j (X_i - \hat{d}_n^{(1)})^j \right]^2 K_h(X_i - \hat{d}_n^{(1)}).$$

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

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

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

For the two-stage procedures, a number of tuning parameters (p, γ 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 confidence interval* calculated from the first-stage data; that is, γ and K satisfy

$$Kn_1^{-\gamma} = Cq_\beta n_1^{-1/3}, \tag{4.1}$$

where q_β is the upper β quantile of $\mathbb{Z}$. On the other hand, a good quantitative rule for selecting the first-stage proportion of samples 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 were considered.

Finally, due to presence of small sample sizes the following modification of the second-stage estimator in both two-stage procedures 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.

The basic settings of the simulation study are as follows: two regression functions are considered, $f_1(x) = x^2$ 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 the uniform distribution. Further, the target is set to $d_0 = 0.5$, the standard deviation of the random error σ 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 from 0.2 to 0.8 in increments of 0.1. Finally, the levels of significance α and β are set to 0.025. The following quantities are computed: coverage rates and average lengths of confidence intervals, and mean squared errors of second-stage estimators.

REMARK 4.1. Choosing γ and K via equation (4.1) is theoretically equivalent to having $\gamma = 1/3$ and $K = Cq_\beta$. Notice that strictly speaking, weak consistency for $\gamma = 1/3$ is not expected to hold for the bootstrapped estimator, since the normalized first-stage estimator $n^{1/3}(\hat{d}_{n_1}^{(1)} - d_0)$ has neither an almost sure nor an in probability limit (see Theorem 3.7). However, it is reasonable to expect that for relatively small samples, the performance of the bootstrap would be satisfactory, since $\gamma = 1/3$ is at the boundary of weak consistency. The obtained simulation results certainly vindicate this expectation.

4.1. *Comparison of Two-Stage Procedures.* Recall that from the first-stage data, we obtain the following asymptotic $(1 - 2\beta)$ confidence interval for d_0 with the true parameter (see Theorem 2.1):

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

We consider the above confidence interval, as the sampling interval $[L, U]$ with $\gamma = 1/3$ and $K = Cq_\beta$. Then, for the case $\gamma = 1/3$ in Theorem 3.2 and Theorem 3.4, we get for $f \in \mathscr{F}$,

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

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

$$[\tilde{d}_n^{(2)} - \tilde{q}_\alpha n^{-1/2}, \tilde{d}_n^{(2)} + \tilde{q}_\alpha n^{-1/2}] \bigcap [0, 1], \tag{4.2}$$

where $\tilde{q}_\alpha$ is the upper α 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 2, the coverage rates of the (4.2) confidence intervals for combinations of f, n and σ are shown based on 5000 replications, using the *true* parameters $f'(d_0)$ and σ (i.e. the true C, C_2 and C_3 in constructing the confidence 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. 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 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 σ at stage one. It can be seen that for f_2, for almost all proportions p, noise levels and larger sample size ($n \geq 150$) the coverage rates achieve the nominal level. On the other hand, for f_1, this is the case only for small noise levels ($\sigma \leq 0.3$), larger sample sizes ($n \geq 150$) and first-stage proportion $0.5 \leq p \leq 0.7$. It can be concluded, that the BTSP exhibits superior performance to the TSP for settings with small noise and relatively large sample sizes.

Finally, the third row in Figure 2 depicts the coverage rates of the bootstrapped procedure, when both $f'(d_0)$ and σ are estimated at stage one, as outlined above. 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.

The above findings strongly suggest that $p = 0.4$ is a good choice for functions exhibiting a strong linear trend in the vicinity of d_0, while $p = 0.6$ is preferable otherwise.

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 the confidence intervals, as well as the mean squared errors of the estimates of d_0. The

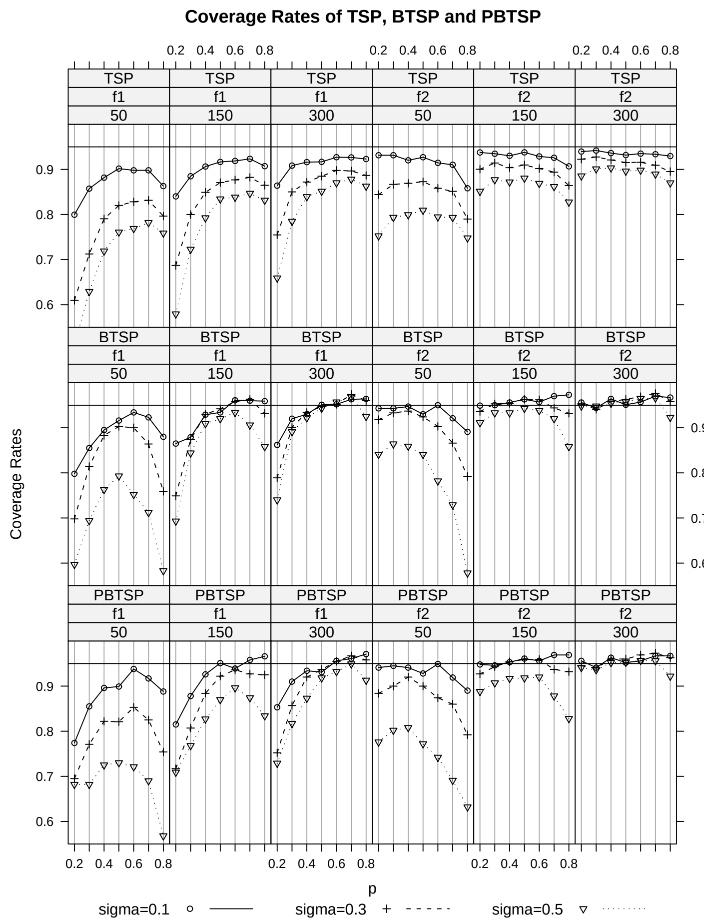

FIG 2. *Coverage Rate plot grouped with different σ's.*

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 the results shown in Table 1 that except for the case where the sample size is small ($n = 50$) and the noise large ($\sigma = 0.5$) both procedures perform well in terms of coverage rates. Further, under the PBTSP, confidence intervals usually have shorter average lengths, while the estimates for d_0 smaller mean squared errors, with the biggest gains accruing in the f_2 case. However, both procedures suffer in the f_1 case with large noise and small to moderate sample sizes.

REMARK 4.2. One of the advantages of the bootstrap procedure, as pointed out

TABLE 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.

f	p	σ	n	CR	CR1	AL	AL1	MSE	MSE1
f_1	0.6	0.1	50	**0.938**	**0.958**	0.15	0.19	1.1e-03	2.1e-03
			100	**0.942**	**0.950**	0.09	0.15	4.5e-04	1.3e-03
			150	**0.939**	**0.940**	0.07	0.13	2.9e-04	1.1e-03
			200	**0.949**	**0.943**	0.06	0.12	1.9e-04	9.2e-04
			300	**0.956**	**0.936**	0.05	0.10	1.3e-04	7.4e-04
		0.3	50	0.853	0.899	0.43	0.37	2.5e-02	1.1e-02
			100	0.917	**0.932**	0.33	0.30	9.6e-03	6.5e-03
			150	**0.940**	**0.932**	0.27	0.27	4.4e-03	5.0e-03
			200	**0.945**	**0.942**	0.22	0.24	2.8e-03	4.2e-03
			300	**0.955**	**0.949**	0.17	0.21	1.8e-03	3.0e-03
		0.5	50	0.721	0.831	0.48	0.48	5.1e-02	2.3e-02
			100	0.820	0.911	0.48	0.41	3.0e-02	1.3e-02
			150	0.896	0.903	0.43	0.37	1.9e-02	1.0e-02
			200	0.919	0.919	0.40	0.34	1.4e-02	8.7e-03
			300	**0.932**	**0.934**	0.33	0.30	8.0e-03	5.9e-03
f_2	0.4	0.1	50	**0.941**	**0.955**	0.09	0.20	6.1e-04	2.4e-03
			100	**0.956**	**0.955**	0.06	0.16	2.4e-04	1.4e-03
			150	**0.953**	**0.954**	0.05	0.14	1.6e-04	1.1e-03
			200	**0.952**	**0.955**	0.04	0.12	1.2e-04	9.4e-04
			300	**0.963**	**0.961**	0.03	0.11	7.1e-05	7.0e-04
		0.3	50	0.920	**0.935**	0.38	0.39	1.1e-02	9.6e-03
			100	**0.941**	**0.938**	0.25	0.32	4.2e-03	6.3e-03
			150	**0.950**	**0.951**	0.19	0.28	2.0e-03	5.1e-03
			200	**0.961**	**0.955**	0.16	0.25	1.5e-03	3.8e-03
			300	**0.960**	**0.950**	0.12	0.22	8.8e-04	3.0e-03
		0.5	50	0.808	0.840	0.51	0.49	4.0e-02	2.2e-02
			100	0.898	0.899	0.44	0.42	1.7e-02	1.3e-02
			150	0.917	0.927	0.36	0.38	1.2e-02	9.0e-03
			200	**0.941**	**0.931**	0.32	0.35	6.7e-03	8.2e-03
			300	**0.952**	**0.939**	0.24	0.31	4.0e-03	5.9e-03

in Subsection 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 4.3. Notice that the sigmoid function f_2 belongs to class $\mathscr{F}_2$ 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 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 4.4. 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 σ^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 the likelihood-ratio-based confidence interval for d_0 obtained from the first-stage data. The likelihood ratio procedure tests the hypothesis $H_{0,d} : f^{-1}(\theta_0) = d$ vs $H_{1,d} : f^{-1}(\theta_0) \neq d$ and assigns d to the confidence set if the likelihood-ratio statistic falls below an appropriate threshold determined by a pre-specified quantile of its limit distribution (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 (see [3] for details). Alternatively, we can use the extremities of the likelihood-ratio-based confidence interval itself 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.

REMARK 4.5. In the case of $f \in \mathscr{F}_1$, 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 4.6. Our simulation results indicate that good choices for p are 0.6 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.

5. Data Application. A brief description of the engineering system under consideration is given next. We consider a complex queueing system comprised of 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 [2]).

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 [2] is shown in Figure 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 20,000 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 $\theta_0 = 3$, 5, 10 and 15 units of delay. For comparison purposes between the one- and two-stage procedures we fix a budget of $n = 100$. The obtained one stage data (black points) are shown in Figure 3. 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 was estimated locally, by choosing an appropriate radius $r > 0$ and using only the points in the interval $[\hat{d}_n^{(1)} \pm r]$. After careful examination of the data plot, we chose r to be 0.15, 0.15, 0.075 and 0.075 for the various θ_0's.

When implementing the two-stage procedure, we selected 50 points from those used in the one-stage procedure ($p = 0.5$). Then, the following values of r were selected for obtaining the second-stage sampling intervals: $r = 0.15, 0.1, 0.1, 0.075$ for the various θ_0's. After obtaining the 50 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 confidence intervals from the POSP and the PBTSP are given in the first two blocks of Table 2 and the first two plots in Figure 3. It can be seen that the point estimates are fairly similar for $\theta_0 = 3, 5$ and 10, but differ somewhat for $\theta_0 = 15$. Notice that the PBTSP should provide a better estimate in the latter case, since in that case d_0 is close to the boundary of the domain and isotonic regression does not work particularly well; on the other hand, there is strong linearity of f around $d_0 = 0.85$, which supports the obtained estimates from the two-stage procedure. More significantly, the length of the confidence intervals from PBTSP are much smaller than those from POSP. 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) in the data.

We also used the PBTSP with only 20 second stage points, by selecting at random 10 out of the original 25 from L and U. The results are shown in the last block of Table 2 and the last plot of Figure 3. It is clear that the point estimates are quite close to those of the two-stage procedure based on a total budget of $n = 100$; specifically, the confidence intervals become wider, but are still much narrower than those of the one-stage procedure.

TABLE 2
Comparing PBTSP and POSP

	θ	3	5	10	15
	$\hat{d}_n^{(1)}$	0.270	0.578	0.806	0.906
POSP	95%	0.196	0.542	0.763	0.885
$n = 100$	CI	0.344	0.614	0.846	0.927
	$\tilde{d}_n^{(2)}$	0.265	0.569	0.779	0.861
PBTSP	95%	0.259	0.565	0.776	0.856
$n = 50 + 2 \times 25$	CI	0.272	0.572	0.782	0.866
	$\tilde{d}_n^{(2)}$	0.272	0.566	0.780	0.864
PBTSP	95%	0.262	0.559	0.777	0.856
$n = 50 + 2 \times 10$	CI	0.281	0.574	0.782	0.872

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 regression function; nevertheless, other non-parametric procedures could be used. Further, the local approximation was primarily based on a linear model, although quadratic and when 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 a derivative of the regression function at the unknown target point, especially with small to moderate sample sizes. Its asymptotic properties are also investigated and its consistency established.

Our simulation results indicate that the practical bootstrapped procedure performs well in a variety of settings. Nevertheless, for relatively small budgets, a sequence of deterministic points (e.g. quantile based) may yield an improved performance.

Finally, we note that the main results generalize readily to heteroskedastic mod-

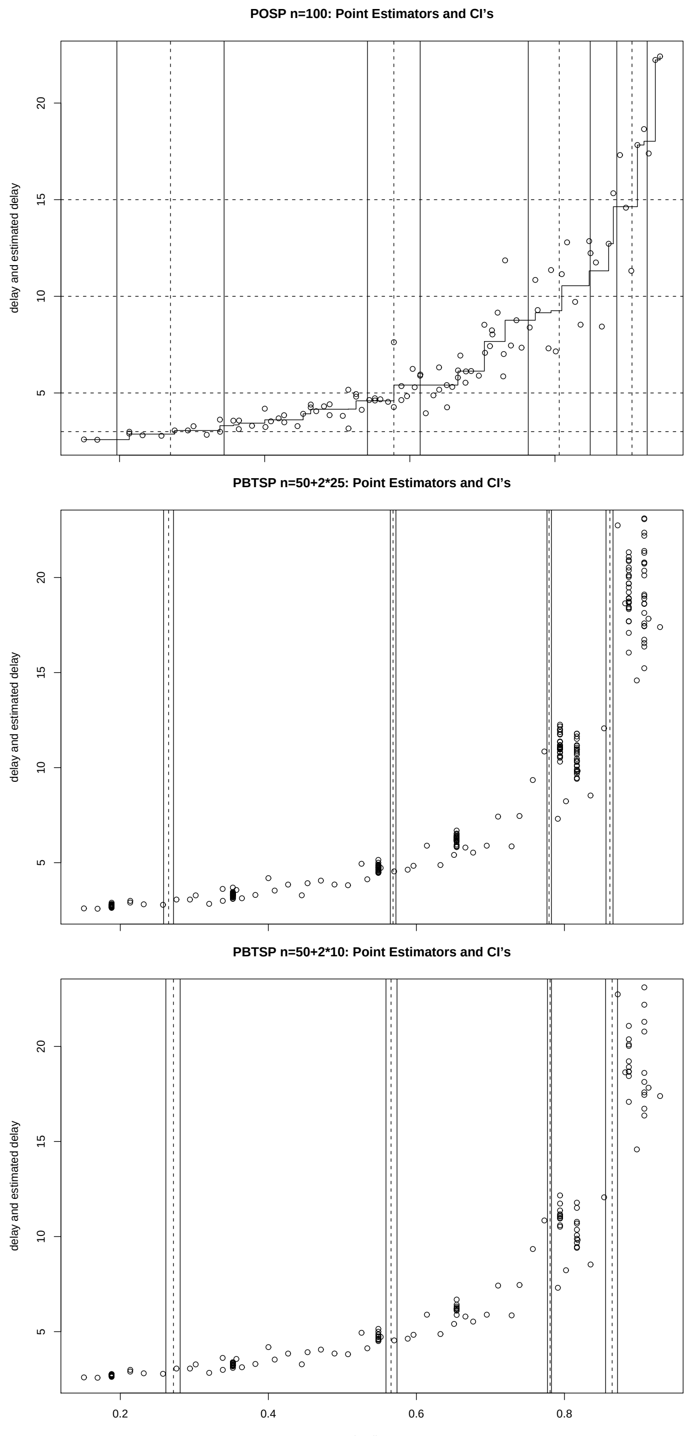

FIG 3. *Comparing PBTSP and POSP*

els of the form $Y = f(X) + \sigma(X)\epsilon$, where the error term ϵ is *independent* of the design variable X and $\sigma(X)$ is a scaling function that determines the conditional 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.

Acknowledgements. The authors thank Professor Michael Woodroofe for his suggestion of employing a local linear approximation in a shrinking neighborhood of the target quantity and for many useful discussions on this topic.

APPENDIX A: APPENDIX

In order to establish the asymptotic distribution of the bootstrapped two-stage estimator, we need a rate of the almost sure convergence for the one-stage estimator of d_0 obtained by isotonic regression. Lemma A.1 below shows that $n^{1/4}$ is a possible boundary rate.

Before proceeding, a theoretical clarification is required. In Lemma A.1 below, we intend to prove that an event involving the first-stage estimator of d_0, say B, happens P-almost surely, where P is the product probability for the product space $(\Omega, \mathscr{A})$ defined in Subsection 3.2. Since B only depends on the first-stage data, there exists $\tilde{B} \in \mathscr{A}_1 \otimes \mathscr{A}_2$, such that $B = \{\omega \in \Omega : (\omega_1, \omega_2) \in \tilde{B}, \omega_3 \in \Omega_3, \omega_4 \in \Omega_4\}$ and that $P(B) = P_{12}(\tilde{B})$, where $P_{12} = P_1 \times P_2$. Thus, it suffices to show that $P_{12}(\tilde{B}) = 1$, which is exactly what the next lemma does. For simplicity of notation, we use $(\Omega, \mathscr{A}, P)$ generically. That is, we still denote the product probability space for the first-stage data as $(\Omega, \mathscr{A}, P)$, where $\Omega = \Omega_1 \times \Omega_2$, $\mathscr{A} = \mathscr{A}_1 \otimes \mathscr{A}_2$ and $P = P_{12}$. This principle of using generic notation is applied to similar situations in the following part of this paper.

LEMMA A.1. *Suppose $f \in \mathscr{F}_0$ is Lipschitz continuous, and f' is positive and continuous in a neighborhood of d_0. Then, for any $a_n \to \infty$ we have*

$$P\left(\lim_{n\to\infty} \frac{1}{\alpha_n}\left(\frac{n}{\log\log n(\log n)^2}\right)^{1/4} |\hat{d}_n^{(1)} - d_0| = 0\right) = 1.$$

Thus, for all $\alpha < 1/4$, we have

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

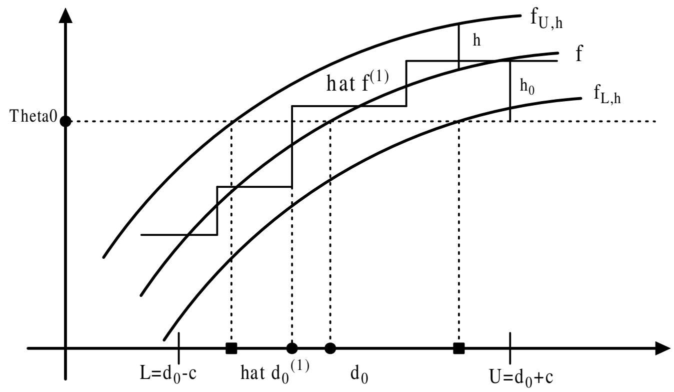

FIG 4. *For the proof of Lemma A.1*

REMARK A.1. We ask the reader to refer to Figure 4 as a pictorial aid for the proof below.

PROOF. Given the assumptions on f, there exists $c > 0$ and $0 < a < b < 1$ such that f is continuous and strictly increasing on $[L, U] = [d_0 \pm c] \subset (a, b)$ and that $\rho = \inf_{x \in [L,U]} f'(x)$ is positive.

Let $f_{U,h}(x) = f(x) + h$ and $f_{L,h}(x) = f(x) - h$ for each $x \in [L, U]$ and every $h > 0$. Let $h_0 = \min\{f(U) - \theta_0, \theta_0 - f(L)\}$. Then, for all $h \in (0, h_0)$, we have $d_0 \in [f^{-1}_{U,h}(\theta_0), f^{-1}_{L,h}(\theta_0)] = [f^{-1}(\theta_0 - h), f^{-1}(\theta_0 + h)] \subset [L, U]$.

By Lemma A.5, we have

$$P\left(\lim_{n\to\infty} \frac{1}{\alpha_n}\left(\frac{n}{\log\log n(\log n)^2}\right)^{1/4} \sup_{x\in[a,b]} |\hat{f}^{(1)}(x) - f(x)| = 0\right) = 1.$$

Denote $M_n = \sup_{x\in[a,b]} |\hat{f}^{(1)}(x) - f(x)|$. Taking α_n to be a sequence slowly converging to ∞ (e.g. $\log n$). Then we have $M_n \to 0$ $(a.s.)$. Thus, M_n converges to 0 in probability.

When $M_n < h_0$, we have $d_0 \in [f^{-1}_{U,M_n}(\theta_0), f^{-1}_{L,M_n}(\theta_0)] \subset [L, U]$. Now on $[a, b]$, $f_{L,M_n} \le \hat{f}^{(1)} \le f_{U,M_n}$, $\hat{f}^{(1)}$ is a right-continuous increasing step function, and $\hat{d}^{(1)}_n$ is the first jump point at which $\hat{f}^{(1)}$ is greater than or equal to θ_0. Thus, $\hat{d}^{(1)}_n \in [f^{-1}_{U,M_n}(\theta_0), f^{-1}_{L,M_n}(\theta_0)]$. In fact, if $\hat{d}^{(1)}_n < f^{-1}_{U,M_n}(\theta_0)$, we have $f_{U,M_n}(\hat{d}^{(1)}_n) < \theta_0$ since f_{U,M_n} is continuous and strictly increasing on $[L, U]$. On the other hand, we have $f_{U,M_n}(\hat{d}^{(1)}_n) \ge \hat{f}^{(1)}(\hat{d}^{(1)}_n) \ge \theta_0$. Thus, we have obtained a contradiction which establishes that $\hat{d}^{(1)}_n \ge f^{-1}_{U,M_n}(\theta_0)$. Similarly, if $\hat{d}^{(1)}_n > f^{-1}_{L,M_n}(\theta_0)$, by the definition of $\hat{d}^{(1)}_n$, we have $\hat{f}^{(1)}(x^\star) < \theta_0$, for each $x^\star \in (f^{-1}_{L,M_n}(\theta_0), \hat{d}^{(1)}_n)$. Hence, we have $f_{L,M_n}(x^\star) \leq \hat{f}^{(1)}(x^\star) < \theta_0$. On the other hand, since $x^\star > f^{-1}_{L,M_n}(\theta_0)$ and f_{L,M_n} is continuous and strictly increasing, we have $f_{L,M_n}(x^\star) > \theta_0$. Again, we obtain a contradiction which establishes that $\hat{d}_n^{(1)} \leq f^{-1}_{L,M_n}(\theta_0)$.

So, when $M_n < h_0$, we have

$$|\hat{d}_n^{(1)} - d_0| \leq f^{-1}_{L,M_n}(\theta_0) - f^{-1}_{U,M_n}(\theta_0) \leq \frac{2}{\rho} M_n. \tag{A.1}$$

The second inequality holds because

$$\begin{aligned}
& f^{-1}_{L,M_n}(\theta_0) - f^{-1}_{U,M_n}(\theta_0) = f^{-1}(\theta_0 + M_n) - f^{-1}(\theta_0 - M_n) \\
\leq\ & 2M_n \sup_{t\in[\theta_0 - M_n, \theta_0 + M_n]} \frac{d}{dt} f^{-1}(t) \leq 2M_n \left(\inf_{t\in[\theta_0 - h_0, \theta_0 + h_0]} f'(f^{-1}(t)) \right)^{-1} \\
\leq\ & 2M_n \left(\inf_{x\in[L,U]} f'(x) \right)^{-1} = \frac{2}{\rho} M_n.
\end{aligned}$$

Putting things together we get

$$\begin{aligned}
& P\left(\lim_{n\to\infty} \frac{1}{\alpha_n} \left(\frac{n}{\log\log n (\log n)^2} \right)^{1/4} |\hat{d}_n^{(1)} - d_0| = 0 \right) \\
\geq\ & P\left(\lim_{n\to\infty} \frac{1}{\alpha_n} \left(\frac{n}{\log\log n (\log n)^2} \right)^{1/4} |\hat{d}_n^{(1)} - d_0| = 0, M_n < h_0 \right) \\
\overset{\text{by (A.1)}}{\geq}\ & P\left(\lim_{n\to\infty} \frac{1}{\alpha_n} \left(\frac{n}{\log\log n (\log n)^2} \right)^{1/4} M_n = 0, M_n < h_0 \right) \\
=\ & P\left(\lim_{n\to\infty} \frac{1}{\alpha_n} \left(\frac{n}{\log\log n (\log n)^2} \right)^{1/4} M_n = 0 \right) \\
-\ & P\left(\lim_{n\to\infty} \frac{1}{\alpha_n} \left(\frac{n}{\log\log n (\log n)^2} \right)^{1/4} M_n = 0, M_n \geq h_0 \right) \\
>\ & 1 - P\left(M_n \geq h_0 \right).
\end{aligned}$$

Since M_n converges to 0 in probability, taking the limit gives

$$P\left(\limsup_{n\to\infty} \frac{1}{\alpha_n} \left(\frac{n}{\log\log n (\log n)^2} \right)^{1/4} |\hat{d}_n^{(1)} - d_0| = 0 \right) = 1.$$

Then, for all $\alpha < 1/4$, we have

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

□

A.1. Proofs for Results in Subsection 3.1.

PROOF OF LEMMA 3.1. We start by considering the following Taylor series expansions:

$$
\begin{aligned}
f(U) &= f(\hat{d}_{n_1}^{(1)} + K n_1^{-\gamma}) \\
&= f(d_0) + f'(d_0)(\hat{d}_{n_1}^{(1)} - d_0 + K n_1^{-\gamma}) \\
&\quad + \frac{f''(d_0)}{2}(\hat{d}_{n_1}^{(1)} - d_0 + K n_1^{-\gamma})^2 + R_U,
\end{aligned} \tag{A.2}
$$

and

$$
\begin{aligned}
f(L) &= f(\hat{d}_{n_1}^{(1)} - K n_1^{-\gamma}) \\
&= f(d_0) + f'(d_0)(\hat{d}_{n_1}^{(1)} - d_0 - K n_1^{-\gamma}) \\
&\quad + \frac{f''(d_0)}{2}(\hat{d}_{n_1}^{(1)} - d_0 - K n_1^{-\gamma})^2 + R_L,
\end{aligned} \tag{A.3}
$$

where $R_U = f'''(\xi_1)(\hat{d}_{n_1}^{(1)} - d_0 + K n_1^{-\gamma})^3/6$, $R_L = f'''(\xi_2)(\hat{d}_{n_1}^{(1)} - d_0 - K n_1^{-\gamma})^3/6$, ξ_1 lies between d_0 and $\hat{d}_{n_1}^{(1)} + K n_1^{-\gamma}$ and ξ_2 lies between d_0 and $\hat{d}_{n_1}^{(1)} - K n_1^{-\gamma}$.

Then, from (3.1), the definitions of Y_i' and Y_i'' and the Taylor expansions (A.2) and (A.3), we get

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{1}{2K n_1^{-\gamma} n_2} \sum_{i=1}^{n_2} (Y_i'' - Y_i') \\
&= \frac{1}{2K n_1^{-\gamma}} \left(f(U) - f(L)\right) + \frac{1}{2K n_1^{-\gamma} n_2} \sum_{i=1}^{n_2} (\epsilon_i'' - \epsilon_i') \\
&= f'(d_0) + f''(d_0)(\hat{d}_{n_1}^{(1)} - d_0) \\
&\quad + \frac{1}{2K n_1^{-\gamma}}(R_U - R_L) + \frac{1}{2K n_1^{-\gamma} n_2} \sum_{i=1}^{n_2} (\epsilon_i'' - \epsilon_i').
\end{aligned}
$$

From Theorem 2.1, $\hat{d}_{n_1}^{(1)} - d_0 \xrightarrow{P} 0$; and by the standard central limit theorem, for $\gamma \in (0, 1/2)$, $(n_1^{\gamma}/n_2) \sum_{i=1}^{n_2} (\epsilon_i'' - \epsilon_i') \xrightarrow{P} 0$. Next We show that $(R_U - R_L)/(2K n_1^{-\gamma}) \xrightarrow{P} 0$ 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 $\gamma \in (0, 1/2)$. We only show the former; the latter follows in an analogous manner.

From the definition of R_U, we have

$$
\begin{aligned}
n_1^{\gamma} R_U &= n_1^{\gamma} \frac{f'''(\xi_1)}{6} (\hat{d}_{n_1}^{(1)} - d_0 + K n_1^{-\gamma})^3 \\
&= \frac{f'''(\xi_1)}{6} \Big[n_1^{\gamma} (\hat{d}_{n_1}^{(1)} - d_0)^3 + 3K (\hat{d}_{n_1}^{(1)} - d_0)^2 \\
&\qquad + 3K^2 n_1^{-\gamma} (\hat{d}_{n_1}^{(1)} - d_0) + K^3 n_1^{-2\gamma} \Big] . \qquad \text{(A.4)}
\end{aligned}
$$

Theorem 2.1 coupled with Slutsky's Lemma, shows that the sum of the four terms within the square bracket in (A.4) is $o_P(1)$ for $\gamma \in (0,1)$. Thus, we have $n_1^{\gamma} R_U = f'''(\xi_1) o_P(1)$. Since ξ_1 lies between d_0 and $\hat{d}_{n_1}^{(1)} + K n_1^{-\gamma}$ and the latter converges to d_0 in probability, we get that ξ_1 converges to d_0 in probability. On the other hand, $f'''(\cdot)$ is uniformly bounded around d_0. Hence $f'''(\xi_1) o_P(1) = o_P(1)$. This shows that $n_1^{\gamma} R_U$ converges to 0 in probability for $\gamma \in (0,1)$.

Next, consider the consistency of $\hat{\beta}_0$. From (3.1), the definitions of Y_i' and Y_i'' and the Taylor expansions (A.2) and (A.3), for $\gamma \in (0, 1/2)$ we have

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

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

PROOF OF THEOREM 3.2. From (3.2), the definitions of Y_i' and Y_i'' and the Taylor expansions (A.2) and (A.3), we get

$$
\begin{aligned}
\tilde{d}_n^{(2)} - d_0 &= \frac{f(d_0) - \frac{1}{2n_2} \sum_{i=1}^{n_2} (Y_i'' + Y_i')}{\hat{\beta}_1} + \left(\hat{d}_{n_1}^{(1)} - d_0 \right) \\
&= \frac{1}{f'(d_0)} \left[f(d_0) - \frac{1}{2n_2} \sum_{i=1}^{n_2} (Y_i'' + Y_i') \right] + \left(\hat{d}_{n_1}^{(1)} - d_0 \right) \\
&+ \frac{1}{f'(d_0) \hat{\beta}_1} (f'(d_0) - \hat{\beta}_1) \left[f(d_0) - \frac{1}{2n_2} \sum_{i=1}^{n_2} (Y_i'' + Y_i') \right] \\
&= S_1 + S_2 \times S_3
\end{aligned}
$$

where

$$\begin{aligned} S_1 &= -\frac{f''(d_0)}{2f'(d_0)}\left[(\hat{d}_{n_1}^{(1)} - d_0)^2 + K^2 n_1^{-2\gamma}\right] \\ &\quad -\frac{1}{2f'(d_0)}(R_U + R_L) - \frac{1}{2f'(d_0)n_2}\sum_{i=1}^{n_2}(\epsilon_i'' + \epsilon_i'), \end{aligned}$$

$$\begin{aligned} S_2 &= \frac{1}{f'(d_0)\hat{\beta}_1}\left[f''(d_0)(\hat{d}_{n_1}^{(1)} - d_0) + \frac{1}{2Kn_1^{-\gamma}}(R_U - R_L)\right. \\ &\quad \left. + \frac{1}{2Kn_1^{-\gamma}n_2}\sum_{i=1}^{n_2}(\epsilon_i'' - \epsilon_i')\right], \end{aligned}$$

and

$$\begin{aligned} S_3 &= \left\{f'(d_0)(\hat{d}_{n_1}^{(1)} - d_0) + \frac{f''(d_0)}{2}\left[(\hat{d}_{n_1}^{(1)} - d_0)^2 + K^2 n_1^{-2\gamma}\right]\right. \\ &\quad \left. + \frac{1}{2}(R_U + R_L) + \frac{1}{2n_2}\sum_{i=1}^{n_2}(\epsilon_i'' + \epsilon_i')\right\}. \end{aligned}$$

We consider next the exact stochastic orders of the terms $S_1,\ S_2$ and S_3. We start with S_1. From Theorem 2.1, we have $(\hat{d}_{n_1}^{(1)} - d_0)^2 = O_P(n^{-2/3})$; for $\gamma > 0$, we have $n_1^{-2\gamma} = O_P(n^{-2\gamma})$, $R_U = O_P(n^{-1}) + O_P(n^{-3\gamma})$, $R_L = O_P(n^{-1}) + O_P(n^{-3\gamma})$, and $n_2^{-1}\sum_{i=1}^{n_2}(\epsilon_i'' + \epsilon_i') = O_P(n^{-1/2})$. Note that these are the exact rates of weak convergence. Then, for $\gamma \in (0, 1/2)$, we have

$$S_1 = T_1 + T_2 + o_P(n^{-2\gamma} \vee n^{-1/2}),$$

where

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

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

$$\begin{aligned} T_3 &= \frac{1}{\hat{\beta}_1}(\hat{d}_{n_1}^{(1)} - d_0)\frac{1}{2Kn_1^{-\gamma}n_2}\sum_{i=1}^{n_2}(\epsilon_i'' - \epsilon_i'), \\ R &= o_P(n^{-2\gamma} \vee n^{-1/2} \vee n^{\gamma - 5/6}). \end{aligned}$$

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 γ 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 2.1, the standard central limit theorem, Slutsky's Lemma and the Continuous Mapping Theorem, and noting that $n_1^{1/3}(\hat{d}_{n_1}^{(1)} - d_0)$ is independent of $n_2^{-1/2}\sum_{i=1}^{n_2}(\epsilon_i'' + \epsilon_i')$ and $n_2^{-1/2}\sum_{i=1}^{n_2}(\epsilon_i'' - \epsilon_i')$ and that $\epsilon_i'' + \epsilon_i'$ is uncorrelated with $\epsilon_i'' - \epsilon_i'$, we obtain the results for the five cases defined by the different ranges of γ 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

$$\left(n_1^{1/3}(\hat{d}_{n_1}^{(1)} - d_0), \frac{1}{\sqrt{n_2}}\sum_{i=1}^{n_2}(\epsilon_i'' + \epsilon_i'), \frac{1}{\sqrt{n_2}}\sum_{i=1}^{n_2}(\epsilon_i'' - \epsilon_i')\right) \xrightarrow{d} (C\mathbb{Z}, cZ_1, cZ_2),$$

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

$$n^{1/2}(T_2 + T_3) \xrightarrow{d} -\frac{\sigma}{f'(d_0)(1-p)^{1/2}}Z_1 + \frac{\sigma}{Kf'(d_0)(1-p)^{1/2}}C\mathbb{Z}Z_2.$$

In similar fashion, we obtain the asymptotic results for the other four cases. □

The proofs of Lemma 3.3 and Theorem 3.4 are essentially the same to those of Lemma 3.1 and Theorem 3.2. In fact, the only difference comes from the Taylor expansions of $f(U)$ and $f(L)$. Therefore, we only provide an outline of the main steps.

PROOF OF LEMMA 3.3. The following Taylor series expansions are considered:

$$\begin{aligned} f(U) &= f(\hat{d}_{n_1}^{(1)} + Kn_1^{-\gamma}) \\ &= f(d_0) + f'(d_0)(\hat{d}_{n_1}^{(1)} - d_0 + Kn_1^{-\gamma}) \\ &\quad + \frac{f'''(d_0)}{6}(\hat{d}_{n_1}^{(1)} - d_0 + Kn_1^{-\gamma})^3 + R_U, \end{aligned} \tag{A.5}$$

and

$$
\begin{aligned}
f(L) &= f(\hat{d}_{n_1}^{(1)} - Kn_1^{-\gamma}) \\
&= f(d_0) + f'(d_0)(\hat{d}_{n_1}^{(1)} - d_0 - Kn_1^{-\gamma}) \\
&\quad + \frac{f'''(d_0)}{6}(\hat{d}_{n_1}^{(1)} - d_0 - Kn_1^{-\gamma})^3 + R_L,
\end{aligned} \tag{A.6}
$$

where $R_U = f^{(4)}(\xi_1)(\hat{d}_{n_1}^{(1)} - d_0 + Kn_1^{-\gamma})^4/24$, $R_L = f^{(4)}(\xi_2)(\hat{d}_{n_1}^{(1)} - d_0 - Kn_1^{-\gamma})^4/24$, ξ_1 lies between d_0 and $\hat{d}_{n_1}^{(1)} + Kn_1^{-\gamma}$ and ξ_2 lies between d_0 and $\hat{d}_{n_1}^{(1)} - Kn_1^{-\gamma}$.

From (3.1), the definitions of Y_i' and Y_i'', and the Taylor expansions (A.5) and (A.6), we have, for $\gamma \in (0, 1/2)$,

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{1}{2Kn_1^{-\gamma}n_2}\sum_{i=1}^{n_2}(Y_i'' - Y_i') \\
&= f'(d_0) + \frac{f'''(d_0)}{2}(\hat{d}_{n_1}^{(1)} - d_0)^2 + \frac{f'''(d_0)}{6}K^2 n_1^{-2\gamma} \\
&\quad + \frac{1}{2Kn_1^{-\gamma}}(R_U - R_L) + \frac{1}{2Kn_1^{-\gamma}n_2}\sum_{i=1}^{n_2}(\epsilon_i'' - \epsilon_i') \xrightarrow{P} f'(d_0),
\end{aligned}
$$

$$
\begin{aligned}
\hat{\beta}_0 &= \frac{1}{2n_2}\sum_{i=1}^{n_2}(Y_i'' + Y_i') - \hat{d}_{n_1}^{(1)}\hat{\beta}_1 \\
&= f(d_0) + f'(d_0)(\hat{d}_{n_1}^{(1)} - d_0) + \frac{f'''(d_0)}{2}(\hat{d}_{n_1}^{(1)} - d_0)K^2 n_1^{-2\gamma} \\
&\quad + \frac{1}{2}(R_U + R_L) + \frac{1}{2n_2}\sum_{i=1}^{n_2}(\epsilon_i'' + \epsilon_i') - \hat{d}_{n_1}^{(1)}\hat{\beta}_1 \xrightarrow{P} f(d_0) - d_0 f'(d_0),
\end{aligned}
$$

and $\tilde{d}_n^{(2)} = (\theta_0 - \hat{\beta}_0)/\hat{\beta}_1 \xrightarrow{P} d_0$. ☐

PROOF OF THEOREM 3.4. From (3.2), the definitions of Y_i' and Y_i'', and the Taylor expansions (A.5) and (A.6), we have, for $\gamma \in (1/8, 1/2)$,

$$
\begin{aligned}
\tilde{d}_n^{(2)} - d_0 &= \frac{f(d_0) - \frac{1}{2n_2}\sum_{i=1}^{n_2}(Y_i'' + Y_i')}{\hat{\beta}_1} + \left(\hat{d}_{n_1}^{(1)} - d_0\right) \\
&= \frac{1}{f'(d_0)}\left[f(d_0) - \frac{1}{2n_2}\sum_{i=1}^{n_2}(Y_i'' + Y_i')\right] + \left(\hat{d}_{n_1}^{(1)} - d_0\right) \\
&\quad + \frac{1}{f'(d_0)\hat{\beta}_1}(f'(d_0) - \hat{\beta}_1)\left[f(d_0) - \frac{1}{2n_2}\sum_{i=1}^{n_2}(Y_i'' + Y_i')\right] \\
&= S_1 + S_2 \times S_3,
\end{aligned}
$$

where

$$\begin{aligned} S_1 &= -\frac{f'''(d_0)}{6f'(d_0)}\left(\hat{d}_{n_1}^{(1)}-d_0\right)^3-\frac{f'''(d_0)}{2f'(d_0)}\left(\hat{d}_{n_1}^{(1)}-d_0\right)K^2n_1^{-2\gamma} \\ &\quad -\frac{1}{2f'(d_0)}(R_U+R_L)-\frac{1}{2f'(d_0)n_2}\sum_{i=1}^{n_2}\left(\epsilon_i''+\epsilon_i'\right), \end{aligned}$$

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

and

$$\begin{aligned} S_3 &= \left\{f'(d_0)\left(\hat{d}_{n_1}^{(1)}-d_0\right)+\frac{f'''(d_0)}{6}\left(\hat{d}_{n_1}^{(1)}-d_0\right)^3\right. \\ &\quad \left.+\frac{f'''(d_0)}{2}\left(\hat{d}_{n_1}^{(1)}-d_0\right)K^2n_1^{-2\gamma}+\frac{1}{2}(R_U+R_L)+\frac{1}{2n_2}\sum_{i=1}^{n_2}\left(\epsilon_i''+\epsilon_i'\right)\right\}. \end{aligned}$$

Similar to the proof of Theorem 3.4, we have $S_1+S_2\times S_3=T_1+T_2+R$, where

$$\begin{aligned} T_1 &= -\frac{1}{2f'(d_0)n_2}\sum_{i=1}^{n_2}\left(\epsilon_i''+\epsilon_i'\right), \\ T_2 &= \frac{1}{\hat{\beta}_1}\left(\hat{d}_{n_1}^{(1)}-d_0\right)\frac{1}{2Kn_1^{-\gamma}n_2}\sum_{i=1}^{n_2}\left(\epsilon_i''-\epsilon_i'\right), \end{aligned}$$

and R is the sum of the remaining terms which converges to 0 faster than T_1 and T_2. Then, the result of the current theorem can be established by following steps similar to the proof of Theorem 3.2. □

A.2. Proofs for Results in Subsection 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 $b_1\le b_2\in\mathbb{R}$. Note that $\zeta_n=B_{as}(b)$ for some $b>0$ implies $\zeta_n\to 0$ almost surely.

PROOF OF LEMMA 3.5. The proofs of Lemmas 3.1 and 3.3 establish 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. Here we just show the strong consistency of $\hat{\beta}_1$ for $f \in \mathscr{F}_1$; the case for $f \in \mathscr{F}_2$ can be handled analogously.

From the proof of Lemma 3.1, it suffices to show $\hat{d}^{(1)}_{n_1} - d_0$, $(n_1^{\gamma}/n_2)\sum_{i=1}^{n_2}(\epsilon_i^{''} - \epsilon_i^{'})$ and $(R_U - R_L)/(2Kn_1^{-\gamma})$ converge to 0 almost surely. Lemma A.1 shows that $\hat{d}^{(1)}_{n_1} - d_0$ converges to 0 almost surely, while Lemma A.3, establishes that $(n_1^{\gamma}/n_2)\sum_{i=1}^{n_2}(\epsilon_i^{''} - \epsilon_i^{'})$ 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 $\gamma \in (0, 3/4)$. Next, we show the former; the latter follows analogously.

Since ξ_1 lies between d_0 and $\hat{d}^{(1)}_{n_1} + Kn_1^{-\gamma}$ and the latter converges to d_0 almost surely, we know ξ_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 A.1, the four terms within square brackets on the right-side of (A.4) are $B_{as}(3/4-\gamma)$, $B_{as}(1/2)$, $B_{as}(1/4+\gamma)$ and $B_{as}(2\gamma)$. Thus, by (A.4), we have that $n_1^{\gamma}R_U$ almost surely converges to 0 for $\gamma \in (0, 3/4)$.

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

Next, we establish the conditional weak consistency of $\hat{\beta}_1^{\star}$ for both $f \in \mathscr{F}_1$ and $f \in \mathscr{F}_2$. From (3.4), we get

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

where

$$T_1 = \frac{1}{2Kn_1^{-\gamma}n_2}\sum_{i=1}^{n_2}\left(\epsilon_i^{\prime\prime\star} - \epsilon_i^{\prime\star}\right),\ \ T_2 = \frac{1}{2Kn_1^{-\gamma}}\left(f(U) - f(L)\right).$$

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

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

$$V_i^- = \epsilon_i^{\prime\prime\star} - \epsilon_i^{\prime\star},\ \ \nu^- = E_{\star}[V_i^-] = \frac{1}{n_2}\sum_{i=1}^{n_2}\left(\epsilon_i^{\prime\prime} - \epsilon_i^{\prime}\right).$$

and

$$s^2 = Var_{\star}[V_i^-] = \frac{1}{n_2}\sum_{i=1}^{n_2}(\epsilon_i^{\prime\prime})^2 - \left(\frac{1}{n_2}\sum_{i=1}^{n_2}\epsilon_i^{\prime\prime}\right)^2 + \frac{1}{n_2}\sum_{i=1}^{n_2}(\epsilon_i^{\prime})^2 - \left(\frac{1}{n_2}\sum_{i=1}^{n_2}\epsilon_i^{\prime}\right)^2.$$

For $\gamma \in (0, 1/2)$, Lemma A.3 gives that $T_{12} \to 0,\ (P - a.s.)$. For $\gamma \in (0, 1/2)$, by Lemma A.6 and Slutsky's Lemma, we know $T_{11} \stackrel{P^{\star}}{\to} 0,\ (P - a.s.)$. Thus, for $\gamma \in (0, 1/2)$, we have $T_1 \stackrel{P^{\star}}{\to} 0,\ (P - a.s.)$.

Next, we consider T_2. If $f \in \mathscr{F}_1$, by the strong consistency of $\hat{d}_{n_1}^{(1)} - d_0$ and $(R_U - R_L)/(2Kn_1^{-\gamma})$, we have, for $\gamma \in (0, 3/4)$,

$$T_2 = f'(d_0) + f''(d_0)\left(\hat{d}_{n_1}^{(1)} - d_0\right) + \frac{1}{2Kn_1^{-\gamma}}(R_U - R_L) \to f'(d_0),\ (P - a.s.).$$

If $f \in \mathscr{F}_2$, similarly, for $\gamma \in (0, 1)$, we have

$$\begin{aligned} T_2 &= f'(d_0) + \frac{f'''(d_0)}{2}\left(\hat{d}_{n_1}^{(1)} - d_0\right)^2 \\ &\quad + \frac{f'''(d_0)}{6}K^2 n_1^{-2\gamma} + \frac{1}{2Kn_1^{-\gamma}}(R_U - R_L) \to f'(d_0),\ (P - a.s.). \end{aligned}$$

Thus, for f belonging either $\mathscr{F}_1$ or $\mathscr{F}_2$ and for $\gamma \in (0, 1/2)$, we have $T_2 \to f'(d_0),\ (P - a.s.)$. Therefore, we get $\hat{\beta}_1^\star \xrightarrow{P^\star} f'(d_0),\ (P - a.s.)$. ☐

PROOF OF THEOREM 3.6. From (3.2) and (3.3), we have

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

where

$$T_1 = \frac{\sqrt{n}}{f'(d_0)2n_2}\sum_{i=1}^{n_2}\left[(Y_i''^\star + Y_i'^\star) - (Y_i'' + Y_i')\right]$$

and

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

By the definitions of $Y_i', Y_i'', Y_i'^\star, Y_i''^\star$, we have

$$T_1 = \frac{\sqrt{n}}{f'(d_0)2n_2}\sum_{i=1}^{n_2}\left[(\epsilon_i''^\star + \epsilon_i'^\star) - (\epsilon_i'' + \epsilon_i')\right] = \frac{s\sqrt{n}}{2f'(d_0)\sqrt{n_2}}\sum_{i=1}^{n_2}\frac{V_i^+ - \nu^+}{s\sqrt{n_2}},$$

where

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

and

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

The SLLN gives $s^2 \to 2\sigma^2,\ (P-a.s.)$. By Lemma A.6, we get $\sum_{i=1}^{n_2}(V_i^+ - \nu_i^+)/(s\sqrt{n_2}) \xrightarrow{d^\star} Z_1,\ (P-a.s.)$. Note that $\sqrt{n}/\sqrt{n_2} \to \sqrt{2/(1-p)}$. Thus, Slutsky's Lemma implies

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

In Lemma A.2 following this proof, we show that for $\gamma \in (0, 1/4)$,

$$T_2 \xrightarrow{P^\star} 0,\ (P-a.s.).$$

Therefore, another application of Slutsky's Lemma proves this theorem.

$\square$

LEMMA A.2. *For* $\gamma \in (0, 1/4)$, $T_2 \xrightarrow{P^\star} 0,\ (P-a.s.)$.

PROOF. Let

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

we then have

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

We will show that $T_{2i} \xrightarrow{P^\star} 0,\ (P-a.s.),\ i = 1, 2, 3, 4$ for $\gamma \in (0, 1/4)$. Thus, by Lemma 3.5 and Slutsky' Lemma, we know the lemma holds.

We establish next the convergence of the terms T_{2i}. First assume f belongs to $\mathscr{F}_1$. From (3.1), (3.4), the definitions of Y_i', Y_i'', $Y_i'^\star$ and $Y_i''^\star$, and the Taylor's expansions of $f(L)$ and $f(U)$ ((A.2) and (A.3)), we have

$$A = \frac{n_1^\gamma}{2Kn_2}\sum_{i=1}^{n_2}\left[(\epsilon_i''^\star - \epsilon_i'^\star) - (\epsilon_i'' - \epsilon_i')\right] = \frac{n_1^\gamma s}{2K\sqrt{n_2}} n_2^{-1/2}\sum_{i=1}^{n_2}\left[\frac{V_i^- - \nu^-}{s}\right],$$

$$B = \frac{1}{2n_2}\sum_{i=1}^{n_2}\left[(\epsilon_i''^\star + \epsilon_i'^\star) - (\epsilon_i'' + \epsilon_i')\right] = \frac{s}{2\sqrt{n_2}} n_2^{-1/2}\sum_{i=1}^{n_2}\left[\frac{V_i^+ - \nu^+}{s}\right],$$

$$I = \hat{\beta}_1 - f'(d_0) = f''(d_0)(\hat{d}_{n_1}^{(1)} - d_0) + \frac{n_1^\gamma}{2K}(R_U - R_L) + \frac{n_1^\gamma}{2Kn_2}\sum_{i=1}^{n_2}\left(\epsilon_i'' - \epsilon_i'\right),$$

and

$$\begin{aligned} II &= f(d_0) - \frac{1}{2n_2}\sum_{i=1}^{n_2}\left(Y_i'' + Y_i'\right) \\ &= -f'(d_0)(\hat{d}_{n_1}^{(1)} - d_0) - \frac{f''(d_0)}{2}\left[(\hat{d}_{n_1}^{(1)} - d_0)^2 + K^2 n_1^{-2\gamma}\right] \\ &\quad -\frac{1}{2}(R_U + R_L) - \frac{1}{n_2}\sum_{i=1}^{n_2}\left(\epsilon_i'' + \epsilon_i'\right). \end{aligned}$$

Thus, we have

$$T_{21} = n^{1/2} A \cdot I \cdot II = T_{21}' s n_2^{-1/2}\sum_{i=1}^{n_2}\left[\frac{V_i^- - \nu^-}{s}\right],$$

where $T_{21}' = C_n \cdot I \cdot II$ and $C_n = n^{1/2} n_1^\gamma (2K\sqrt{n_2})^{-1}$.

SLLN and Lemma A.6 give

$$s \to \sqrt{2}\sigma,\ (P - a.s.),\ \ n_2^{-1/2}\sum_{i=1}^{n_2}\left[\frac{V_i^- - \nu^-}{s}\right] \xrightarrow{d^\star} Z_2,\ (P - a.s.).$$

Next it will be shown that for $\gamma \in (0, 3/8)$, T_{21}' converges to 0 P-almost surely. Then, an application of Slutsky's Lemma gives $T_{21} \xrightarrow{P^\star} 0,\ (P - a.s.)$.

For easing the presentation, we introduce some notation regarding the upper bound of the almost sure convergence rate of a sequence of random variables; this is in analogy to the usual $o_P(1)$ and $O_P(1)$ definitions.

With this notation and by Lemmas A.3 and A.1, we have, For $\gamma > 0$, $n_1^\gamma = B_{as}(-\gamma)$, $(\hat{d}_{n_1}^{(1)} - d_0) = B_{as}(1/4)$, $\sum_{i=1}^{n_2}(\epsilon_i'' + \epsilon_i')/n_2 = B_{as}(1/2)$ 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 $B_{as}(3/4) + B_{as}(\gamma + 1/2) + B_{as}(2\gamma + 1/4) + B_{as}(3\gamma)$. Thus we have $C_n = B_{as}(-\gamma)$, $I = B_{as}(1/4) +$

$B_{as}(-\gamma)(B_{as}(3/4)+B_{as}(\gamma+1/2)+B_{as}(2\gamma+1/4)+B_{as}(3\gamma))+B_{as}(1/2-\gamma) = B_{as}(1/4)+B_{as}(2\gamma)+B_{as}(1/2-\gamma)$ and $II = B_{as}(1/4)+[B_{as}(1/2)+B_{as}(2\gamma)]+(B_{as}(3/4)+B_{as}(\gamma+1/2)+B_{as}(2\gamma+1/4)+B_{as}(3\gamma))+B_{as}(1/2) = B_{as}(1/4)+B_{as}(2\gamma)$. Thus, we have

$$\begin{aligned} T'_{21} &= C_n \cdot I \cdot II \\ &= B_{as}(-\gamma) \times [B_{as}(1/4) + (B_{as}(2\gamma)) + B_{as}(1/2-\gamma)] \\ &\quad \times \{B_{as}(1/4) + B_{as}(2\gamma)\} \\ &= B_{as}(1/2-\gamma) + B_{as}(1/4+\gamma) + B_{as}(3/4-2\gamma) + B_{as}(3\gamma). \end{aligned}$$

It is easy to see that when $\gamma \in (0, 3/8)$, the above five upper bounds $1/2-\gamma$, $1/4+\gamma$, $3/4-2\gamma$, and 3γ are all positive. This implies that for $\gamma \in (0, 3/8)$, T'_{21} converges to 0 P-almost surely. Therefore, for $\gamma \in (0, 3/8)$, 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 γ. 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 $\gamma \in (0, 1/4)$, T_{23} converges to 0 in probability $(P - a.s.)$. Since $1/4 < 3/8 < 1/2$, we have, for $\gamma \in (0, 1/4)$, T_{2i} converges to 0 in probability $(P - a.s.)$ for $i = 1, 2, 3, 4$. Thus when f is from $\mathscr{F}_1$, for $\gamma \in (0, 1/4)$, T_2 converges to 0 in probability $(P - a.s.)$.

Next assume f belongs to $\mathscr{F}_2$. The derivation is basically the same as the case $f \in \mathscr{F}_1$. Note that for this case we use the Taylor expansions (A.5) and (A.6). It turns out that the γ intervals for which T_{2i}, $i = 1, 2, 3, 4$ converge to 0 in probability $(P - a.s.)$ are exactly the same as those in the former case. Thus, when f is from $\mathscr{F}_2$, for $\gamma \in (0, 1/4)$, T_2 converges to 0 in probability $(P - a.s.)$ as well. ☐

PROOF OF THEOREM 3.7. 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 2.1, we know that $n_1^{\gamma}(\hat{d}_{n_1}^{(1)} - d_0) \equiv (np)^{\gamma}(\hat{d}_{np}^{(1)} - d_0) \xrightarrow{P} 0$. It follows, by the relationship between convergence in probability and almost sure convergence (for example, see Theorem 20.5 in Billingsley [4]), that there exists $\{n_{k(i)}\}_{i=1}^{\infty}$, a further subsequence of $\{n_k\}$, such that $n_{k(i),1}^{\gamma}(\hat{d}_{n_{k(i),1}}^{(1)} - d_0) \to 0,\ (P - a.s.)$. It now suffices to show that

$$n_{k(i)}^{1/2}(\tilde{d}_{n_{k(i)}}^{(2)\star} - \tilde{d}_{n_{k(i)}}^{(2)}) \xrightarrow{d^\star} C_2 Z_1,\ (P - a.s.).$$

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 3.6, 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 mimicking the steps in Theorem 3.6.

It remains to show that $T_2 \xrightarrow{P^\star} 0$ $(P - a.s.)$. The proof of this follows from that of Lemma A.2 by replacing n, n_1 and n_2 by $n_{k(i)}$, $n_{k(i),1}$ and $n_{k(i),2}$ respectively, and noting $\hat{d}^{(1)}_{n_{k(i),1}} - d_0 = B_{as}(1/3)$.

More specifically, for $f \in \mathscr{F}_1$ consider $T_{21} = C_{n_{k(i)}} \cdot I \cdot II$. Note that $C_{n_{k(i)}} = B_{as}(-\gamma)$, $I = B_{as}(1/3) + B_{as}(-\gamma)(B_{as}(1) + B_{as}(\gamma + 2/3) + B_{as}(2\gamma + 1/3) + B_{as}(3\gamma)) + B_{as}(1/2 - \gamma) = B_{as}(1/3) + B_{as}(2\gamma) + B_{as}(1/2 - \gamma)$ and $II = B_{as}(1/3) + [B_{as}(2/3) + B_{as}(2\gamma)] + (B_{as}(1) + B_{as}(\gamma + 2/3) + B_{as}(2\gamma + 1/3) + B_{as}(3\gamma)) + B_{as}(1/2) = B_{as}(1/3) + B_{as}(2\gamma)$. Then, $T_{21} = B_{as}(2/3 - \gamma) + B_{as}(1/3 + \gamma) + B_{as}(5/6 - 2\gamma) + B_{as}(3\gamma) + B_{as}(1/2)$. Thus, $T_{21} \xrightarrow{P^\star} 0$ $(P - a.s.)$ for $\gamma \in (0, 5/12)$. Similarly, T_{22} and T_{24} converge to 0 in probability $(P - a.s.)$ for $\gamma \in (0, 1/2)$ and T_{23} converges to 0 in probability $(P - a.s.)$ for $\gamma \in (0, 1/3)$. Because $1/3 < 5/12 < 1/2$, T_2 converges to 0 in probability $(P - a.s.)$ for $\gamma \in (0, 1/3)$.

For $f \in \mathscr{F}_2$, it turns out that the γ intervals for which $T_{2i}, i = 1, 2, 3, 4$ converge to 0 in probability $(P - a.s.)$ are exactly the same as those in the former case. Thus, when f is from $\mathscr{F}_2$, T_2 also converges to 0 in probability $(P - a.s.)$ for $\gamma \in (0, 1/3)$. □

A.3. Some Auxiliary Lemmas. Suppose $\{\epsilon_i\}_{i=1}^{\infty}$ are the first-stage random errors defined in Subsection 3.2. Let $\bar{\epsilon}_n = (1/n)\sum_{i=1}^{n} \epsilon_i$. The following simple lemma shows that $n^{1/2}$ is an upper boundary of the almost sure convergence rate of $\bar{\epsilon}_n$.

LEMMA A.3. *For $\alpha < 1/2$, we have*

$$P\left(\lim_{n\to\infty} n^{\alpha}\bar{\epsilon}_n = 0\right) = 1.$$

PROOF. Using the Law of the Iterated Logarithm (See, for example, Page 397 of Shiryaev (1995) [24]) yields

$$P\left(\limsup_{n\to\infty}\left(\frac{n}{2\sigma^2 \log\log n}\right)^{1/2}\bar{\epsilon}_n = 1\right) = 1.$$

and the result follows. □

Suppose $\{X_i\}_{i=1}^{\infty}$ is the sequence of first-stage design points defined in Subsection 3.2. Denote the *maximal spacing* of the first n design points by

$$\Delta_n = \max\{X_{(1)}, X_{(2)} - X_{(1)}, \cdots, X_{(n)} - X_{(n-1)}, 1 - X_{(n)}\}.$$

Then the following lemma provides a almost sure order of Δ_n.

LEMMA A.4. *We have $\Delta_n = O(\log n/n)$ P-almost surely.*

PROOF. We have $m = \inf_{u\in[0,1]} g(u) > 0$ since g is continuous and positive on $[0,1]$. Denote by G the cumulative distribution function of X. Then G is strictly increasing on $[0,1]$, with $G(0) = 0$ and $G(1) = 1$. Letting $H(v) = G^{-1}(v)$ for $v \in [0,1]$, it can be seen that H is strictly increasing on $[0,1]$, with $H(0) = 0$ and $H(1) = 1$. Furthermore, we have $H'(v) = 1/g(H(v)) \le 1/m$ for $v \in [0,1]$. Then, we have $H(v_2) - H(v_1) \le M(v_2 - v_1)$ for $v_1 \le v_2 \in [0,1]$, where $M = 1/m$; i.e. H satisfies a Lipschitz condition with a positive constant M.

Let $U_i = G(X_i)$ for $i = 1,2,\cdots$. Then $\{U_i\}_i^\infty$ are iid copies of $U[0,1]$. Since G is increasing, given n, we have $U_{(i)} = G(X_{(i)})$ for $i = 1,2,\cdots,n$. By the established Lipschitz condition on H, we have $X_{(i)} - X_{(i-1)} = H(U_{(i)}) - H(U_{(i-1)}) \le M(U_{(i)} - U_{(i-1)})$ for $i = 2,\cdots,n$. Since $H(0) = 0$ and $H(1) = 1$, we have $X_{(1)} = H(U_{(1)}) - H(0) \le MU_{(1)}$ and $1 - X_{(n)} = H(1) - H(U_{(n)}) \le M(1 - U_{(i-1)})$. Thus, we have $\Delta_n \le M\Delta_n^U$, where

$$\Delta_n^U = \max\{U_{(1)}, U_{(2)} - U_{(1)}, \cdots, U_{(n)} - U_{(n-1)}, 1 - U_{(n)}\}.$$

From Devroye (1981) [6], we know

$$\limsup_{n\to\infty} \frac{n\Delta_n^U - \log n}{2\log_2 n} = 1,\ (P - a.s.).$$

Thus, we have $\Delta_n^U = O(\log n/n)$ almost surely. Since $0 \le \Delta_n \le C\Delta_n^U$, we have $\Delta_n = O(\log n/n)$ almost surely. □

Next, we establish a result on the rate of the almost sure convergence of the one-stage isotonic regression estimator of f is established. We derive this result in the framework of Subsection 3.2, adapting generic notation as introduced before and used in Lemma A.1. When n increases, more data points are sequentially taken out from the two fixed sequences of first-stage design points and responses and $\hat{f}^{(1)}$ is subsequently updated.

LEMMA A.5. *Suppose $f \in \mathscr{F}_0$ and is Lipschitz continuous. Further, assume $\{\alpha_i\}_{i=1}^\infty$ is an arbitrary sequence of real numbers converging to ∞. Then, for $0 < a < b < 1$, we have*

$$P\left(\lim_{n\to\infty} \frac{1}{\alpha_n}\left(\frac{n}{\log\log n(\log n)^2}\right)^{1/4} \sup_{x\in[a,b]} |\hat{f}^{(1)}(x) - f(x)| = 0\right) = 1.$$

PROOF. This proof essentially runs along the lines of that of Corollary 2.5 on Page 181 of Brunk (1970) [5].

Let $A = \{\omega_1 \in \Omega_1 : \Delta_n = O(\log n/n)\}$ and

$$B = \left\{\omega = (\omega_1, \omega_2) \in \Omega : \lim_{n\to\infty} \frac{1}{\alpha_n} \left(\frac{n}{\log\log n(\log n)^2}\right)^{1/4} \times \sup_{x\in[a,b]} |\hat{f}^{(1)}(x) - f(x)| = 0\right\}.$$

Lemma A.4 shows that $P_1(A) = 1$.

We will establish that $P_2(B_{\omega_1}) = 1$ for each $\omega_1 \in A$. Assuming this result holds, we get

$$P(B) = \int_{\Omega_1} P_2(B_{\omega_1}) P_1(d\omega_1) = \int_A P_2(B_{\omega_1}) P_1(d\omega_1) = \int_A P_1(d\omega_1) = 1.$$

which proves the lemma.

Next, we consider $P_2(B_{\omega_1})$ for $\omega_1 \in A$. The main argument is based on Lemmas 1 and 2 and Theorem in Makowski [13]. Suppose f is Lipschitz continuous with a constant $K > 0$. Given $\omega_1 \in A$, for simplicity, denote $x_i = \tilde{X}_i(\omega_1)$ for $i = 1, 2, \cdots$. Now $Y_i = Y_i(\omega_2) = f(x_i) + \tilde{\epsilon}_i(\omega_2)$, for each $\omega_2 \in \Omega_2$ and $i = 1, 2, \cdots$. That is, given $\omega_1 \in A$, the randomness of Y_i comes from ϵ_i. For a given sample size $n \in \mathbb{N}$, the data is given by $\{(x_i, Y_i)\}_{i=1}^n$, from which we can obtain the sorted data $\{(x_{n,i}, Y_{n,i})\}_{i=1}^n$ in ascending order of $\{x_i\}_{i=1}^n$.

Denote the maximal spacing by generically $\Delta_n = \max\{x_{n,1}, x_{n,2} - x_{n,1}, \cdots, x_{n,n} - x_{n,n-1}, 1 - x_{n,n}\}$, the partial sum of the responses by $S_{n,j} = \sum_{i=1}^j Y_{n,i}$, and the maximum of the absolute values of the centered partial sums by $R'_n = \max_{1\le j\le n} |S_{n,j} - P_2 S_{n,j}|$. Let $s_n^2 = Var[S_{n,n}]$; then $s_n^2 = n\sigma^2 = O(n)$. Note that this variance is calculated under probability measure P_2. For each $\delta > 0$, set $c = c(n) = (2\delta s_n \Delta_n/K)^{1/2}$ and $d = 1 - c$. Note that $c \to 0$. Then, Lemma 1 of Makowski [13] establishes the following result:

$$\{R'_n < \delta s_n\} \subset \left\{\sup_{c\le x\le d} \left|\hat{f}^{(1)}(x) - f(x)\right| < 2(2K\delta s_n\Delta_n)^{1/2} + K\Delta_n\right\}. \tag{A.7}$$

This result connects $\sup_{c\le x\le d} \left|\hat{f}^{(1)}(x) - f(x)\right|$ with R'_n. From Remark 4 on Pages 878-879 of Makowski (1973) [14], we conclude $\{\epsilon_i\}_{i=1}^\infty$ satisfies Condition P' on Page 877 of Makowski (1973) [14]. Thus, by Theorem 3 of Makowski (1973) [14] or Lemma 2 of Makowski (1975) [13], we have the following result:

$$P_2\left(\limsup_{n\to\infty} \frac{R'_n}{(2s_n^2 \log\log s_n^2)^{1/2}} \le K_1\right) = 1, \tag{A.8}$$

where $K_1 = 2\sqrt{6}$. This result establishes a law of iterated logarithm for R'_n.

Then, analogously to the Theorem of Makowski (1975) [13] , for $0 < a < b < 1$, we will show

$$P_2\left(\limsup_{n\to\infty}\left(\frac{n}{\log\log n(\log n)^2}\right)^{1/4}\sup_{a\le x\le b}\left|\hat{f}^{(1)}(x)-f(x)\right|\le K_2\right)=1, \tag{A.9}$$

for a positive constant K_2.

Since $\alpha_n \to \infty$, by (A.9), we have

$$P_2\left(\lim_{n\to\infty}\frac{1}{\alpha_n}\left(\frac{n}{\log\log n(\log n)^2}\right)^{1/4}\sup_{a\le x\le b}\left|\hat{f}^{(1)}(x)-f(x)\right|=0\right)=1.$$

Notice that the event in the above expression is exactly B_{ω_1} for $\omega_1 \in A$, which establishes the lemma.

Now consider the expression given in (A.9). Note that conclusion (A.8) is equivalent to the statement that for every $\lambda > 0$,

$$\lim_m P_2\left(\bigcup_{n\ge m}\left\{R'_n \ge (K_1+\lambda)(2s_n^2\log\log s_n^2)^{1/2}\right\}\right)=0. \tag{A.10}$$

For given n, let $\delta = \delta(n) = (K_1+\lambda)(2\log\log s_n^2)^{1/2}$. Then by (A.7), when n is sufficiently large, there exist positive constants K_3, K_4 and K_5 such that

$$\begin{aligned}
&\left\{R'_n \ge (K_1+\lambda)(2s_n^2\log\log s_n^2)^{1/2}\right\}\\
\supset\;&\left\{\sup_{c\le x\le d}|\hat{f}^{(1)}(x)-f(x)|\ge 2(2K\delta s_n\Delta_n)^{1/2}+K\Delta_n\right\}\\
\supset\;&\left\{\sup_{a\le x\le b}|\hat{f}^{(1)}(x)-f(x)|\ge 2(2K\delta s_n\Delta_n)^{1/2}+K\Delta_n\right\}\\
\supset\;&\left\{\sup_{a\le x\le b}|\hat{f}^{(1)}(x)-f(x)|\ge K_3\left[\log\log n(\log n)^2/n\right]^{1/4}+K_4\log n/n\right\}\\
\supset\;&\left\{\sup_{a\le x\le b}|\hat{f}^{(1)}(x)-f(x)|\ge K_5\left[\log\log n(\log n)^2/n\right]^{1/4}\right\}
\end{aligned}$$

Then, by (A.10), we have

$$\lim_{m\to\infty}P_2\left(\bigcup_{n\ge m}\left(\frac{n}{\log\log n(\log n)^2}\right)^{1/4}\sup_{a\le x\le b}\left|\hat{f}^{(1)}(x)-f(x)\right|\ge K_5\right)=0,$$

which establishes (A.9) with $K_2 > K_5$.

□

Suppose $\{\epsilon_i'\}_{i=1}^n$, $\{\epsilon_i''\}_{i=1}^n$, $\{\epsilon_i^{'\star}\}_{i=1}^n$ and $\{\epsilon_i^{''\star}\}_{i=1}^n$ are the second-stage random errors and the corresponding bootstrapped ones defined in Subsection 3.2. Let $V_i^+ = \epsilon_i''^\star + \epsilon_i'^\star$, $\nu^+ = E_\star[V_i^+]$, $V_i^- = \epsilon_i''^\star - \epsilon_i'^\star$ and $\nu^- = E_\star[V_i^-]$, where $E_\star$ means the expectation conditioning on the second-stage data. It is easy to verify $Var_\star[V_i^+] = Var_\star[V_i^-]$. Thus we denote both as s^2. The following lemma shows that both V_i^+ and V_i^- are asymptotically normal P-almost surely.

LEMMA A.6. *Assuming the above notations and assumptions, we have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{V_i^+ - \nu^+}{s} \xrightarrow{d^\star} Z,\ (P-a.s.),\quad \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{V_i^- - \nu^-}{s} \xrightarrow{d^\star} Z,\ (P-a.s.),$$

where Z follows a $N(0,1)$ distribution.

PROOF. We only prove the case for V_i^+, since that for V_i^- follows along similar lines. Let $\xi_{ni} = (V_i^+ - \nu^+)/(\sqrt{n}s)$, for $i = 1, 2, \cdots, n$, and $S_n = \sum_{i=1}^n \xi_{ni}$. It is easy to see that $E_\star[\xi_{ni}] = 0$ and $Var_\star[S_n] = 1$. Thus, it suffices to check that the following Lindeberg condition holds for each $\epsilon > 0$ (see, for example, Theorem2 on Page 334 of Shiryaev (1995) [24]):

$$\sum_i^n E_\star[\xi_{ni}^2\{|\xi_{ni}| \geq \epsilon\}] \to 0,\ (P-a.s.).$$

We have

$$\begin{aligned}\sum_i^n E_\star[\xi_{ni}^2\{|\xi_{ni}| \geq \epsilon\}] &= E_\star\left[\left(\frac{V_1^+ - \nu^+}{s}\right)^2\left\{\left|\frac{V_1^+ - \nu^+}{s}\right| \geq \sqrt{n}\epsilon\right\}\right] \\ &\leq \frac{1}{\sqrt{n}\epsilon}\frac{1}{|s|^3}E_\star|V_1^+ - \nu^+|^3.\end{aligned}$$

Since

$$s^2 = \frac{1}{n}\sum_{i=1}^n(\epsilon_i'')^2 - \left(\frac{1}{n}\sum_{i=1}^n \epsilon_i''\right)^2 + \frac{1}{n}\sum_{i=1}^n(\epsilon_i')^2 - \left(\frac{1}{n}\sum_{i=1}^n \epsilon_i'\right)^2,$$

we have $s \to \sqrt{2}\sigma,\ (P-a.s.)$, by the Strong Law of Large Numbers (SLLN.) Then, it is sufficient to show $\overline{\lim}_{n\to\infty} E_\star|V_1^+ - \nu^+|^3 < \infty,\ (P-a.s.)$. We have

$$\begin{aligned}E_\star|V_1^+ - \nu^+|^3 &\leq E_\star\left[|V_1^+|^3 + |\nu^+|^3 + 3|V_1^+|^2|\nu^+| + 3|V_1^+||\nu^+|^2\right] \\ &= E_\star|V_1^+|^3 + 3|\nu^+|E_\star|V_1^+|^2 + 3|\nu^+|^2E_\star|V_1^+| + |\nu^+|^3.\end{aligned}$$

Since $\nu^+ = \frac{1}{n}\sum_{i=1}^{n_2}(\epsilon_i'' + \epsilon_i') \to 0,\ (P - a.s.)$, by the SLLN, it suffices to show $\overline{\lim}_{n\to\infty} E_\star |V_1^+|^k < \infty,\ (P - a.s.)$, for $k = 1, 2, 3$. It suffices to show the case where $k = 3$. Since $(a + b)^3 \le 4(a^3 + b^3)$ for nonnegative a and b, we have

$$E_\star |V_1^+|^3 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|\epsilon_i'' + \epsilon_j'|^3 \le 4\left(\frac{1}{n}\sum_{i=1}^{n}|\epsilon_i''|^3 + \frac{1}{n}\sum_{i=1}^{n}|\epsilon_i'|^3\right).$$

By SLLN, we have $\frac{1}{n}\sum_{i=1}^{n}|\epsilon_i''|^3 \to \tau,\ (P - a.s.)$ and $\frac{1}{n}\sum_{i=1}^{n}|\epsilon_i'|^3 \to \tau,\ (P - a.s.)$. Since $\tau < \infty$, finally we have $\overline{\lim}_{n\to\infty} E_\star |V_1^+|^3 \le 8\tau < \infty,\ (P - a.s.)$. $\square$

REFERENCES

[1] ANBAR, D. (1977). A modified Robbins-Monro procedure approximating the zero of a regression function from below. *Ann. Statist.* **5**, 1, 229–234. MR0428642

[2] BAMBOS, N. AND MICHAILIDIS, G. (2004). Queueing and scheduling in random environments. *Adv. in Appl. Probab.* **36**, 1, 293–317. MR2036345

[3] BANERJEE, M. AND WELLNER, J. A. (2005). Confidence intervals for current status data. *Scand. J. Statist.* **32**, 3, 405–424. MR2204627

[4] BILLINGSLEY, P. (1995). *Probability and measure*, Third ed. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York. A Wiley-Interscience Publication. MR1324786

[5] BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)*. Cambridge Univ. Press, London, 177–197. MR0277070

[6] DEVROYE, L. (1981). Laws of the iterated logarithm for order statistics of uniform spacings. *Ann. Probab.* **9**, 5, 860–867. MR628878

[7] FAN, J. AND GIJBELS, I. (1996). *Local polynomial modelling and its applications*. Monographs on Statistics and Applied Probability, Vol. **66**. Chapman & Hall, London. MR1383587

[8] GASSER, T., SROKA, L., AND JENNEN-STEINMETZ, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73**, 3, 625–633. MR897854

[9] GROENEBOOM, P. AND WELLNER, J. A. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10**, 2, 388–400. MR1939706

[10] GRUET, M.-A. (1996). A nonparametric calibration analysis. *Ann. Statist.* **24**, 4, 1474–1492. MR1416643

[11] IVANOVA, A., MONTAZER-HAGHIGHI, A., MOHANTY, S. G., AND DURHAM, S. D. (2003). Improved up-and-down designs for phase I trials. *Stat. Med. 22*, 69–82.

[12] LAI, T. L. (2003). Stochastic approximation. *Ann. Statist.* **31**, 2, 391–406. Dedicated to the memory of Herbert E. Robbins. MR1983535

[13] MAKOWSKI, G. (1975). A rate of convergence for a nondecreasing regression estimator. *Bull. Inst. Math. Acad. Sinica* **3**, 1, 61–64. MR0397975

[14] MAKOWSKI, G. G. (1973). Laws of the iterated logarithm for permuted random variables and regression applications. *Ann. Statist. 1*, 872–887. MR0343358

[15] MORGAN, B. J. T. (1985). The cubic logistic model for quantal assay data. *Applied Statistics* **34**, 2, 105–113.

[16] MORGAN, B. J. T. (1992). *Analysis of Quantal Response Data*. Chapman & Hall.

[17] MÜLLER, H.-G. AND SCHMITT, T. (1990). Choice of number of doses for maximum likelihood estimation of the ED50 for quantal dose-response data. *Biometrics 46*, 117–129.

[18] OSBORNE, C. (1991). Statistical calibration: A review. *International Statistical Review 59*, 309–336.

[19] RAI, K. AND RYZIN, J. V. (1981). A generalized multihit dose-response model for low-dose extrapolation. *Biometrics 37*, 341–352.

[20] ROBBINS, H. AND MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Statistics 22*, 400–407. MR0042668

[21] ROBERTSON, T., WRIGHT, F. T., AND DYKSTRA, R. L. (1988). *Order restricted statistical inference*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester. MR961262

[22] ROSENBERGER, W. F. (1996). New directions in adaptive designs. *Statist. Sci. 11*, 137–149.

[23] ROSENBERGER, W. F. AND HAINES, L. M. (2002). Competing designs for phase I clinical trials: a review. *Stat. Med. 21*, 2757–2770.

[24] SHIRYAEV, A. N. (1995). *Probability*, Second ed. Springer-Verlag.

[25] STANISWALIS, J. G. AND COOPER, V. (1988). Kernel estimates of dose response. *Biometrics* **44**, 4, 1103–1119. MR981002

[26] STYLIANOU, M. AND FLOURNOY, N. (2002). Dose finding using the biased coin up-and-down design and isotonic regression. *Biometrics* **58**, 1, 171–177. MR1891376

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF MICHIGAN
1085 SOUTH UNIVERSITY
ANN ARBOR, MI 48109-1107
E-MAIL: rltang@umich.edu
moulib@umich.edu
gmichail@umich.edu
URL: http://www.stat.lsa.umich.edu/∼rltang
http://www.stat.lsa.umich.edu/∼moulib
http://www.stat.lsa.umich.edu/∼gmichail