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 $x \in [0, 1]$. It is of interest to estimate $d_0 = f^{-1}(\theta_0)$ for some θ_0 in the interior of the range of f, given $f'(d_0) > 0$.

Depending on the nature of the problem, one usually first obtains an estimate of f and subsequently of d_0, either from observational data or from design studies [25]. In the latter case, one specifies a number of values for the design variable, and obtains the corresponding responses, which are then used to get the estimates.

Motivated by an engineering application, fully described in Section 5, we introduce a two-stage design for estimating d_0. Specifically, we consider a complex queueing system operating in discrete time under a throughput (average number

*Supported in part by NSF grant DMS-07-05288
†Supported in part by NIH grant 1RC1CA145444-01
AMS 2000 subject classifications: Primary 62G09, 62G20; secondary 62G07
Keywords and phrases: two-stage estimator, bootstrap, adaptive design, asymptotic properties

of customers processed per unit of time) maximizing resource allocation policy (for details see 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 can only be obtained via expensive simulations. The average delay as a function of the system's loading (number of customers arriving per unit of time) is depicted in Figure 1. The relationship between system loading and average delay can not be easily captured by a simple parametric model; hence, a nonparametric estimator might be more useful. In addition, given that the responses are obtained through simulation, only a relatively small number of simulation runs can be performed. It is of great interest for the system's operator to obtain accurate estimates of the loading corresponding to prespecified delay thresholds (e.g. 10 and 15 time units), so as to be able to decide whether to upgrade the available resources.

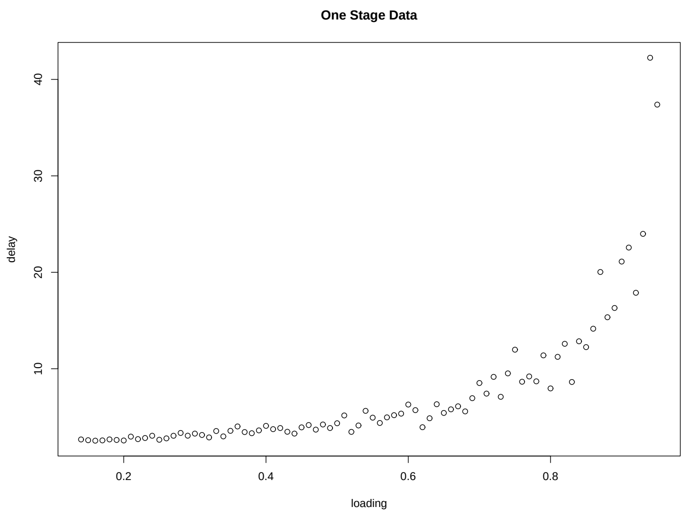

FIG 1. *The average delay as a function of system's loading.*

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. Before proceeding further, it is important to draw attention to the fact that our proposed two-stage method relies critically on the *reproducibility* of the experiment: i.e. at any stage, it is possible to sample responses from any pre-specified covariate value. While reproducibility in this sense is guaranteed for our motivating application, the two-stage procedure above is not applicable in the absence of adequate degree of control on the covariate. For example, if the covariate is time, the implementation of a two-stage procedure would require one to go back and sample from the past, which is impossible.

Isotonic regression is a conceptually natural and computationally efficient estimation method for shape-restricted problems [6, 31]. In the framework of regression, the asymptotic distribution for the isotonic regression estimator at a fixed point was first derived in Brunk [8], and then extended in Wright [37] and Leurgans [21]. The asymptotic distribution for the $\mathbb{L}_1$-distance between the isotonic estimator and the regression function was obtained in Durot [9], paralleling Groeneboom et al. [15] on a unimodal density, and then extended in Durot [10, 11]. Banerjee and Wellner [5] derive the asymptotic distribution for the inverse of the distribution function of the survival time at a given point in the current status model; the regression version of this result will be used to derive the asymptotics for the two stage procedures.

The inverse regression problem has been extensively studied in the context of different applications. For example, in statistical calibration, the goal is to estimate a scalar quantity d_0 from a model $Z = f(d_0) + \epsilon$, with Z observed. The information about the underlying function f comes from experimental data $\{Y_i, X_i\}$ that follow the same regression model; namely, $Y_i = f(X_i) + \epsilon_i$. Osborne [28] gives a comprehensive review of this topic and Gruet [17] provides a kernel based direct nonparametric estimator of d_0. It is clear that when $\epsilon = 0$, the calibration problem becomes the canonical problem described above.

Another active area is provided by the model-based dose-finding problems in toxicology and drug development, where d_0 corresponds to either the maximal tolerated dose or the effective dose with respect to a given maximal toxicity or an efficacy level. Possible dose levels are often prespecified. The dose-response relationship is usually assumed to be monotone and described either by parametric models (e.g. probit, logit [25], multihit [29], cubic logistic [24]), or by nonparametric models, for which kernel estimates [35] and isotonic regression [36] 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 [32] and Rosenberger and Haines [33] for comprehensive reviews on the subject). Stylianou and Flournoy [36] 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. [19] claim that the isotonic regression estimator still performs best for small to moderate sample sizes with several sequential designs from a family of up-and-down designs; Gezmu and Flournoy [14] recommend using smoothed isotonic regression with their group up-and down designs. All these partially motivate the use of isotonic regression in our two-stage procedure, though it should be noted that our approach is markedly different from the ones discussed above, owing to the different nature of the motivating application; in particular, ethical constraints that prevent administration of high dose-levels are absent in our situation.

In a nonparametric setting, one could also employ a fully sequential Robbins-Monro procedure [30] for finding d_0. This 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 [20], 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 [26].

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. Choose n increasing design points $\{x_{in}\}_{i=1}^n \in [0,1]$ and obtain the corre-

sponding responses that are generated according to $Y_{in} = f(x_{in}) + \epsilon_{in}$, $i = 1, 2, \cdots, n$, where f is in $\mathscr{F}_0$, a class of increasing real functions on $[0,1]$ with positive and continuous first derivatives in a neighborhood of d_0 and ϵ_{in} are independently and identically distributed (iid) random errors with mean zero and constant variance σ^2. Note that the subscript n will be suppressed from now on for simplicity of notation.

2. Obtain the isotonic regression estimate $\hat{f}$ of f from the data $\{(x_i, Y_i)\}_{i=1}^n$. (For details see Chapter 1 of Robertson et al. [31]).
3. Estimate d_0 by $\hat{d}_n^{(1)} = \hat{f}^{-1}(\theta_0) = \inf\{x \in [0,1] : \hat{f}(x) \geq \theta_0\}$, where $\theta_0 = f(d_0)$.

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

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

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

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

THEOREM 2.1. *If $f \in \mathscr{F}_0$ and (A1) holds,*

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

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 tails dwindling faster than those of the Gaussian and its quantiles have been tabled in Groeneboom and Wellner [16].

The proof of Theorem 2.1 follows by adaptations of the arguments from Theorem 1 in Banerjee and Wellner [5] to the current regression setting. Hence, an approximate confidence interval for d_0 with significance level $1 - 2\alpha$ can be constructed as follows

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

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

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

2.2. *Procedures Based On Two-Stage Sampling Designs.* We describe next a hybrid estimation procedure for estimating d_0 based on a two-stage sampling design. Suppose that the total budget consists of n doses that are going to be allocated in two stages.

1. Allocate $n_1 = np,\ p \in (0,1)$ design points and obtain the first-stage data $\{(x_i, Y_i)\}_{i=1}^{n_1}$, the isotonic regression estimate of f and the estimate $\hat{d}_{n_1}^{(1)}$ of d_0 as outlined in Section 2.1. Note that by np, we denote by $\lfloor np \rfloor$ or $\lfloor np \rfloor + 1$, depending on whether $n - \lfloor np \rfloor$ is even or not. Also, recall that the additional subscript n is suppressed.
2. Determine two second-stage sampling points L and U symmetrically around $\hat{d}_{n_1}^{(1)}$, where $L = \hat{d}_{n_1}^{(1)} - K n_1^{-\gamma}$ and $U = \hat{d}_{n_1}^{(1)} + K n_1^{-\gamma}$, 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 $\{\epsilon_i\}$.
4. Fit the second-stage data $\{(L, Y_i'), (U, Y_i'')\}$ with the linear model $y = \beta_0 + \beta_1 x$ using least squares. Denote the resulting intercept and slope estimates by $(\hat{\beta}_0, \hat{\beta}_1)$, respectively. Then, the second-stage (or two-stage) estimator of d_0 is given by $\tilde{d}_n^{(2)} = (\theta_0 - \hat{\beta}_0)/\hat{\beta}_1$.

Asymptotic properties of $\tilde{d}_n^{(2)}$ will be established in Subsection 3.1. For example, when f is in a subset of $\mathscr{F}_0$, denoted as $\mathscr{F}$, the third derivatives of whose elements are uniformly bounded around d_0, and $\gamma \in (1/4, 1/3)$, we have

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

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 $n^{1/3}$ convergence rate of the one-stage isotonic regression estimator.

However, when constructing confidence intervals from 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 $f \in \mathscr{F}$, $\gamma \in (0, 1/3)$ and all the absolute moments of the random error are finite, we have

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

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 $f \in \mathscr{F}$ and $\gamma \in (1/4, 1/3)$. In fact, the strong assumption on the random error can be replaced by a mild one that the sixth moment of the random error is finite, at the price of replacing strong consistency with weak consistency. Therefore, the bootstrapped procedure is theoretically validated under certain conditions.

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

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

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

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

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

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

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 three classes of underlying functions:

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

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

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

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

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

Based on Lemma 3.1, we obtain the asymptotic distribution of $\tilde{d}_n^{(2)}$ in the next theorem. It turns out that the asymptotic results with $f \in \mathscr{F}_1$ and $\mathscr{F}_2$ are the same for $\gamma > 1/4$. This implies that the nonlinearity of f at d_0 becomes asymptotically ignorable as the length of the neighborhood of d_0 shrinks fast enough.

THEOREM 3.2. *For $f \in \mathscr{F}$ and $\gamma \in (1/4, 1/2)$,*

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

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

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

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

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

where $C_1 = -K^2 p^{-2\gamma} f''(d_0)/[2f'(d_0)]$, $C_2 = \sigma/[f'(d_0)(1-p)^{1/2}]$, $C_3 = CC_2/K$, C is as given in Theorem 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 given $f \in \mathscr{F}$. On the other hand, for the boundary values of $\gamma = 1/4$ and $1/3$, there exists asymptotic bias in the former case (for $f \in \mathscr{F}_1$), while in the latter case the asymptotic variance increases. For $\gamma > 1/3$, the rate of convergence falls below $\sqrt{n}$, while for $\gamma < 1/4$ and $f \in \mathscr{F}_1$ the limit distribution of the two-stage estimate is degenerate and thus not conducive to inference. Hence, these results suggest selecting γ in the $(1/4, 1/3)$ range. Note that, the function class $\mathscr{F}_2$ achieves the $n^{1/2}$ rate of convergence for a slightly larger range of values for γ than $\mathscr{F}_1$. This is a consequence of the near linearity of f in the vicinity of d_0, which allows a good linear approximation of f with a relatively long interval $[L, U]$.

REMARK 3.3. The case of $\gamma < 1/8$ has been omitted for $f \in \mathscr{F}_2$, since it involves a Taylor expansion of f up to its fifth derivative. Nevertheless, in principle no other technical challenges are in play.

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

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

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

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

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

Let $\{\{x_{in}\}_{i=1}^n\}_{n=1}^\infty$ be a triangluar array of distinct design points in $[0,1]$ and ϵ a continuous random variable in $\mathbb{R}$ with mean 0 and finite variance $\sigma^2 > 0$. Now, there exists, on some probability space $(\Omega, \mathscr{A}, P)$, a set of random errors $\{\{\epsilon_{in}\}_{i=1}^n, \{\epsilon_{in}^{'}\}_{i=1}^n, \{\epsilon_{in}^{''}\}_{i=1}^n\}_{n=1}^\infty$ which are iid copies of ϵ. Then, suppressing the subscript n, $\{\{x_i\}_{i=1}^{n_1}, \{\epsilon_i(\omega)\}_{i=1}^{n_1}, \{\epsilon_i^{'}(\omega)\}_{i=1}^{n_2}, \{\epsilon_i^{''}(\omega)\}_{i=1}^{n_2}\}_{n=1}^\infty$ represents a fixed triangular array of real numbers for a fixed $\omega \in \Omega$, where $n = n_1 + 2n_2$ with n_1 and $2n_2$ denoting the first and second stage sample sizes.

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

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

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

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

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

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

REMARK 3.4. There exist triangular arrays of design points satisfying (A2). For example, with $x_i = i/n$ for $i = 1, 2, \cdots, n$ and all n, we have an array of discrete uniform designs on $[0,1]$. Let G be the uniform distribution function on $[0,1]$. Then, for this special array $\sup_{x\in[0,1]}|F_n(x)-G(x)| \leq 1/n$. Note that (A2) is stronger than (A1). A random variable with finite moment generating function in a small neighborhood of 0 satisfies (A4), such as a normal random variable. The assumptions (A2) to (A4) are essentially the fixed design versions of the assumptions for Lemma 1 of Durot [11], a modification of which enables us to identify a crucial boundary rate for the almost sure convergence of the isotonic regression estimator of d_0. This boundary rate plays a central role in the strong consistency of the bootstrapped estimator.

Next, we state results on the strong consistency of $\hat{\beta}_1$ and the conditional weak consistency of $\hat{\beta}_1^\star$ and then on strong consistency of the bootstrapped estimator. Note that $P^\star$ denotes the probability of the bootstrapped data conditional on the original data.

LEMMA 3.3. *If $f \in \mathscr{F}$, $\gamma \in (0, 1/2)$ and (A2) to (A4) hold,*

$$\hat{\beta}_1 \to f'(d_0),\ (P-a.s.),\ \ \hat{\beta}_1^\star \stackrel{P^\star}{\to} f'(d_0),\ (P-a.s.),$$

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

THEOREM 3.4. *If $f \in \mathscr{F}$, $\gamma \in (0, 1/3)$ and (A2) to (A4) hold,*

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

where C_2 and Z_1 are as in Theorem 3.2. That is,

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

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

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

THEOREM 3.5. *If $f \in \mathscr{F}$, $\gamma \in (0, 1/3)$ and (A1) and (A5) hold, for $t \in \mathbb{R}$,*

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

where C_2 and Z_1 are as in Theorem 3.2.

REMARK 3.5. Comparing Theorem 3.4 with Theorem 3.2, we see that, under the strong assumption (A5) on the random errors, the bootstrapped estimator is strongly consistent for $f \in \mathscr{F}$ and $\gamma \in (1/4, 1/3)$, which is exactly the γ-range of most interest. Further, if f is locally linear at d_0, i.e. $f \in \mathscr{F}_2$, the strong consistency continues to hold for $\gamma \in (1/8, 1/4]$. Similar conclusions on weak consistency hold by comparing Theorem 3.5 with Theorem 3.2, but under the milder assumption (A5) on the random errors.

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. [13], which is based on local linear fitting. Suppose the data $\{(x_i, Y_i)\}_{i=1}^n$ are already sorted in ascending order of x_i's. Then, we calculate

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

where $\tilde{\epsilon}_i = a_i Y_{i-1} + b_i Y_{i+1} - Y_i$, $c_i^2 = (a_i^2 + b_i^2 + 1)^{-1}$, $a_i = (x_{i+1} - x_i)/(x_{i+1} - x_{i-1})$ and $b_i = (x_i - x_{i-1})/(x_{i+1} - x_{i-1})$, for $i = 2, 3, \cdots, n-1$. An estimate of $f'(d_0)$ is obtained through the local quadratic regression estimator proposed by Fan and Gijbels [12], at the estimate $\hat{d}_n^{(1)}$. Specifically, let $K(\cdot)$ denote the Epanechnikov kernel function and $h > 0$ the bandwidth, so that $K_h(\cdot) = (1/h)K(\cdot/h)$. Further, let $\hat{\eta} = (\hat{\eta}_0, \hat{\eta}_1, \hat{\eta}_2)$ be given by

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

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

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

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

For the two-stage procedures, the tuning parameters 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 Wald-type 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 sample proportion p is not available; nevertheless, a practical qualitative rule of thumb dictates that p should decrease, while np should increase as the sample size increases. In our simulation study a number of different values for p are considered.

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

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

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

REMARK 4.1. Note that our method for choosing the tuning parameters γ, K brings in another subjective parameter β. However, the choice of β is guided by a rational principle, namely the requirement that the chosen interval contain the truth with high probability. The magnitude of β is related to how conservative the experimenter wants to be in the construction of the second stage interval which is fundamentally subjective. Also, our rule of thumb regarding the choice of p is based on the idea that with larger budgets smaller p's at stage one will still lead to reasonably precise sampling intervals at stage two, leaving a larger proportion of points for stage two and the possibility of more accurate conclusions.

The basic settings of the simulation study are as follows: two regression functions are considered, $f_1(x) = x^2 + x/5$ and $f_2(x) = e^{4(x-0.5)}/(1 + e^{4(x-0.5)})$ for $x \in [0, 1]$. The first-stage design points are drawn from a discrete uniform distribution on $[0, 1]$, i.e. $x_i = i/(n_1 + 1)$. Further, the target is set to $d_0 = 0.5$, the standard deviation of the random error 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 ranges from 0.2 to 0.8 in increments of 0.1. Finally, the levels of significance α and β are set to 0.025. Note that β is only required to be small and the specific choice of 0.025 is somewhat arbitrary. The following quantities are computed: coverage rates and average lengths of confidence intervals, and mean squared errors of estimators. The simulation programs and more results can be found on the first author's webpage: www.stat.lsa.umich.edu/∼rltang. In this paper, we show part of the results for saving space.

REMARK 4.2. Choosing γ and K via equation (4.1) is theoretically equivalent to having $\gamma = 1/3$ and $K = Cq_\beta$. Notice that strictly speaking, neither strong nor weak consistency for $\gamma = 1/3$ is expected to hold for the bootstrapped estimator. However, it is reasonable to expect that for realistic sample sizes, the performance of the bootstrap would be satisfactory, since $\gamma = 1/3$ is at the boundary of consistency. The obtained simulation results certainly vindicate this expectation. We would like to note that there are other bootstrap methods that could have been used, like the wild or residual bootstrap or the m out of n bootstrap, but it is not clear whether they would yield consistency at $\gamma = 1/3$. It would be interesting to explore some of these issues in future work.

4.1. *Comparison of Two-Stage Procedures.* By Theorem 2.1, from the first-stage data, an asymptotic $(1-2\beta)$ confidence interval for d_0 with the true parameter is given by:

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

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

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

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

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

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

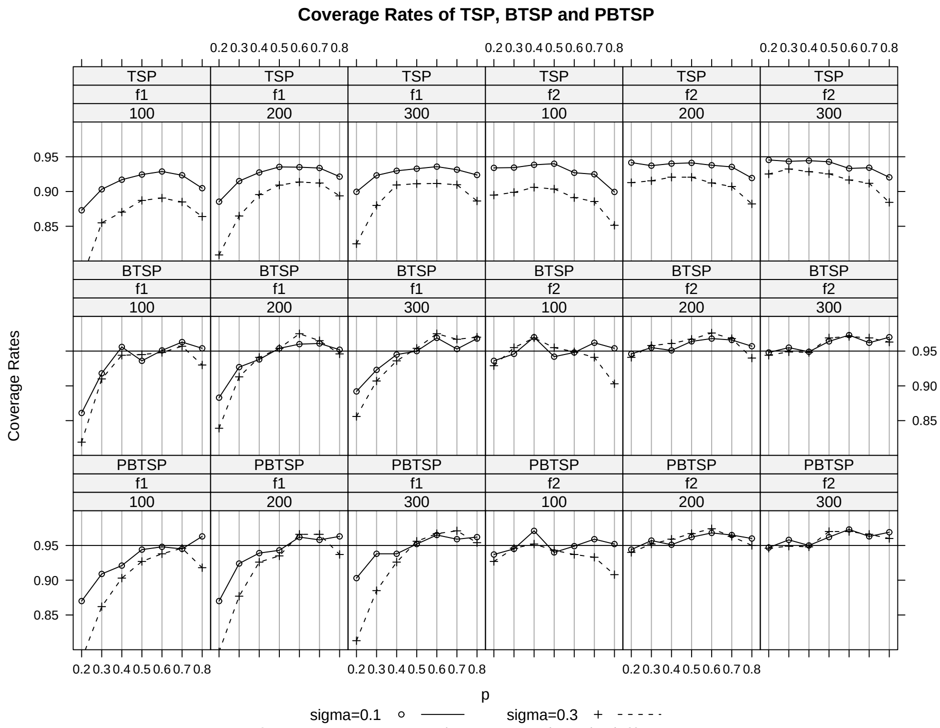

FIG 2. *Coverage Rate plot grouped with different σ's.*

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

The second row in Figure 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 the coverage rates achieve the nominal level with proper first-stage sample proportions, smaller values of which are preferred since both average lengths and mean square errors are usually increasing with p from simulation results not shown in this paper. It can be concluded that the BTSP exhibits superior performance to the TSP in terms of coverage rates, especially for settings with f_1, moderate noise and relatively small sample sizes.

Finally, the third row in Figure 2 depicts the coverage rates of the bootstrapped procedure, when both $f'(d_0)$ and σ are estimated from the first-stage data. The results based on 1000 replicates and 3000 bootstrap samples per replicate indicate a high level of agreement with those of the BTSP, which in turn suggests that the PBTSP is reliable in applications.

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

4.2. *Comparison of One- and Two-Stage Procedures.* We compare next the POSP and the PBTSP, in terms of coverage rates and average lengths of confidence intervals. We also compare the mean squared errors of the first- and second-stage estimates of d_0. The results for POSP are based on 5000 replications, while those of PBTSP on 1000 replications and 3000 bootstrap samples per replication, due to its computational intensity. It can be seen from Table 1 that both procedures usually perform well in terms of coverage rates. Further, under the PBTSP, confidence intervals usually have shorter average lengths, and the estimates for d_0 have smaller mean squared errors, with slightly more gains accruing in the f_2 case. However, it needs to be pointed out that both procedures suffer in the case with large noise and small to moderate sample sizes, especially for f_1.

TABLE 1
CR, AL and MSE stand for coverage rates, average lengths and mean squared errors of PBTSP while CR1, AL1 and MSE1 stand for those of POSP. ALR and MSER are the ratios of CR over CR1 and MSE over MSE1, respectively.

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

REMARK 4.3. One of the advantages of the bootstrap procedure, as pointed out 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 4.4. 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 (i.e. second derivatives) of the regression functions at these two points are about 0.76 and -0.76, respectively. The results are very close to those obtained for $d_0 = 0.5$.

REMARK 4.5. In PBTSP, the second stage sampling points L and U are identified through a Wald-type confidence interval constructed via estimating $f'(d_0)$ and 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 a testing-based confidence interval for d_0 obtained from the first-stage data. This testing-based interval is obtained as follows: consider testing the hypothesis $H_{0,d} : f^{-1}(\theta_0) = d$ vs $H_{1,d} : f^{-1}(\theta_0) \neq d$. Let $\hat{f}^{(1)}$ denote the usual isotonic estimator of f from the stage one data and $\hat{f}_d^{(1)}$ the constrained isotonic estimator under $H_{0,d}$. The residual sum of sqaures based test statistic is given by

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

where $\hat{\sigma}^2$ is a consistent estimate of σ^2. The inversion procedure assigns d to the confidence set if $RSS(d)$ falls below an appropriate threshold determined by a pre-specified quantile of its limit distribution $\mathbb{D}$ (when $d = d_0$ holds true), which is completely parameter-free and therefore enables the construction of the confidence set without the need for nuisance parameter estimation. The limit distribution of $RSS(d^0)$ can be derived by adapting Theorem 2 of [5] (where a likelihood ratio statistic is dealt with) to the residual sum of squares statistic in the nonparametric regression setting, but see also [3] and [4] for a unified treatment of likelihood ratio and residual sum of squares statistics in monotone function problems.

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

REMARK 4.6. 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 4.7. Our simulation results indicate that good choices for p are 0.5 for f_1 and 0.4 for f_2, respectively. Our practical recommendation is $p = 0.5$, whenever no prior information about the linearity of f around d_0 is available.

5. Data Application. We apply our methods to the engineering problem introduced at the beginning of this paper. We briefly describe the underlying system next: consider a complex queueing system comprising N first-in-first-out infinite capacity queues holding different classes of customers and a set of service resources. These resources are externally modulated by a stochastic process. The main issue is to allocate the available resources to the queue in an appropriate manner so as to maximize the system's throughput. This system represents a canonical model for wireless data/voice transmissions, in flexible manufacturing and in call centers (for more details see [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 2,000 events. Notice that our ability to simulate the system at any loading in order to obtain the response, allows us to easily implement the proposed two-stage procedure.

It is of interest to estimate $d_0 = f^{-1}(\theta_0)$ for $\theta_0 = 10$ and 15 units of delay, since around loadings corresponding to those levels the quality of service provided by the system exhibits a significant deterioration. For comparing the one- and two-stage procedures we fix a budget of $n = 82$. A fixed design wth spacing 0.01

was used in the interval $[0.14, 0.95]$ to obtain the one-stage data shown in Figure 1 (also in the left-panel plots of Figure 4). It can be seen that the response is heteroskedastic, but this does not affect the isotonic regression based estimation of f and thus of d_0. However, it impacts the construction of confidence intervals through the estimation of the variance at d_0. To overcome this issue, the variance function is estimated locally by the method proposed in [27]. More specifically, we compute the initial local variance estimates with the weights $(1/\sqrt{2}, -1/\sqrt{2})$ and the smoothed variance function by using *glkerns* in the R package *lokern* with an adaptive bandwidth, shown in the left panel of Figure 3.

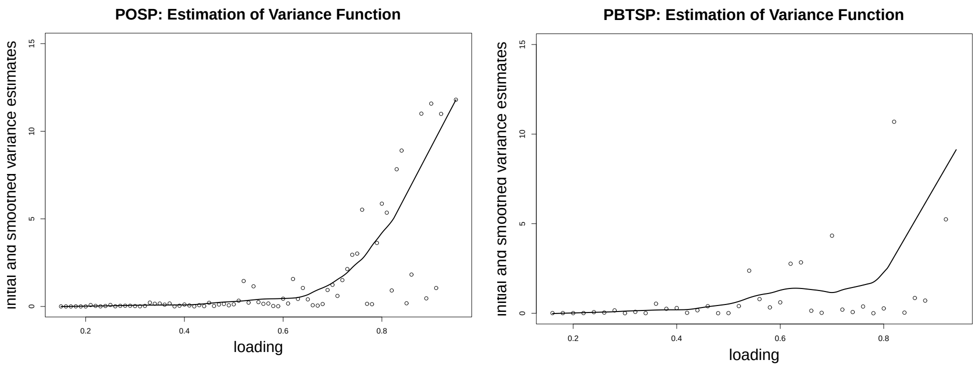

FIG 3. *Estimation of the Variance Function in POSP and PBTSP*

When implementing the two-stage procedure, we selected every other point from those used in the one-stage procedure ($p = 0.5$), thus resulting in a fixed design with spacing 0.02 on the interval $[0.14, 0.94]$. The initial local variance estimates and smoothed variance function with the first-stage data are shown in the right panel of Figure 3. After obtaining the $40 = 2 \times 20$ second-stage responses, the second-stage estimator of d_0 was computed using weighted least squares, with weights being the reciprocals of the estimated local variances at the corresponding sampling points.

The point estimates and the associated 95% confidence intervals from the POSP and the PBTSP are given in Table 2 and plotted in Figure 4. It can be seen that the point estimates are fairly similar. More significantly, the confidence intervals from PBTSP are much shorter than those from POSP, especially for the case $\theta_0 = 10$. This can be attributed to two factors: (i) the applicability of the linear model locally and (ii) the presence of a strong signal (small noise) for the design points around 0.8.

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

TABLE 2
Comparing POSP and PBTSP

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

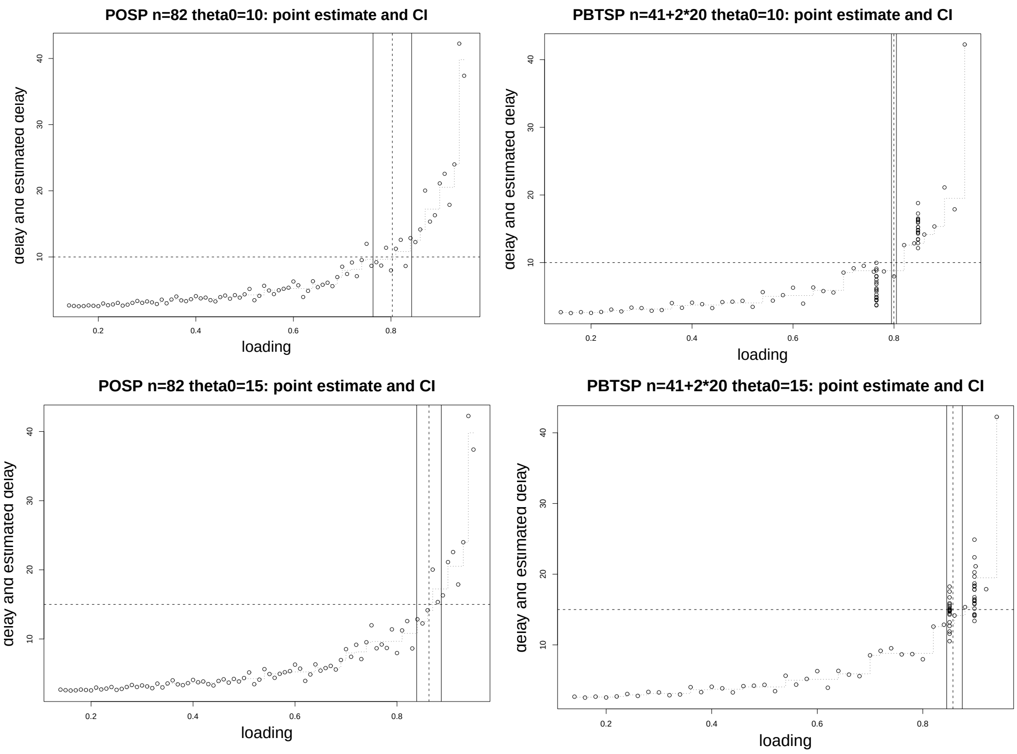

FIG 4. *Comparing POSP and PBTSP*

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

A bootstrapped version of the two-stage procedure is provided to overcome the difficulties posed by the requirement of estimating the derivative of the regression function at the unknown target point and the slow speed of convergence, especially with moderate sample sizes. Its asymptotic properties are also investigated and its strong consistency established (on this point see also Remark A.2).

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

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

Acknowledgements. The authors would like to thank the Editor, the Associate Editor and two anonymous referees for their constructive comments and suggestions. In particular, one of the referees pointed out a strategy and a reference to Durot (2008) that enabled us to strengthen the results on bootstrap consistency. Further, we would like to acknowledge Professor Michael Woodroofe's 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

In order to establish the strong consistency of the bootstrapped two-stage estimator, we need a rate of the almost sure convergence for the one-stage isotonic regression estimator $\hat{d}_n^{(1)}$ of d_0. The following lemma, which is the fixed design version of Lemma 1 in Durot 2008 ([11]), provides a useful tail probability for $\hat{d}_n^{(1)}$.

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

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

PROOF. It will be shown that (A2) implies

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

Recall that "$\lesssim$" denotes that the left side is less than a constant times the right side. Then, reworking the proof of Lemma 1 in Durot [11] for our fixed design setting and an increasing function, and replacing expression (13) in that Lemma with (A.1) ensures that all subsequent steps go through yielding the desired conclusion. To show (A.1) note that from (A2), we get $|G^{-1}(u) - G^{-1}(v)| \lesssim |u - v|$ for every $u, v \in [0, 1]$. Then,

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

gives (A.1) again by (A2). □

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

LEMMA A.2. *If (A2) to (A4) hold, for each $\alpha > 0$,*

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

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

PROOF. Use the notations K, q and η in Lemma A.1. Denote $K' = K\eta^{-3q/2}$ and $A_n = \{n^{1/3-\alpha}|\hat{d}_n^{(1)} - d_0| \geq \eta\}$. By Lemma A.1, $P(A_n) \leq K'n^{-3\alpha q/2}$ for each $\alpha > 0$. On the other hand, (A4) allows q to be arbitrarily large. Choosing $q > 2/(3\alpha)$ gives $\sum_{n=1}^{\infty} P(A_n) \leq K' \sum_{n=1}^{\infty} n^{-3\alpha q/2} < \infty$. Note that $\eta > 0$ is arbitrary. Therefore, $n^{1/3-\alpha}|\hat{d}_n^{(1)} - d_0|$ converges to 0 almost surely (see Corollary on Page 254–255 in Shiryaev [34]), which completes the proof. □

REMARK A.1. Note that Lemmas A.1 and A.2 hold for not only sequences, but also triangular arrays of design points and random errors.

REMARK A.2. The proof of Lemma A.2 implies $n^{1/3-\alpha}(\hat{d}_n^{(1)} - d_0) \overset{a.s.}{\to} 0$ for each $\alpha \in (0, 1/3)$ given $q > 2/(3\alpha)$. Then, $\mathbb{E}|\epsilon|^8 < \infty$ ensures $n^{\beta}(\hat{d}_n^{(1)} - d_0) \overset{a.s.}{\to} 0$ for each $\beta < 1/4$. However, this almost sure convergence result actually holds under a weaker condition $\mathbb{E}|\epsilon|^3 < \infty$ by Theorem in Makowski [22] and Remark 4 in Makowski [23]. This shows that it might be possible to weaken the assumption (A4) a little. Essentially, it means that it might be possible to weaken the condition on the random error in Lemma A.1. In fact, this possibility has been mentioned in Durot's papers on isotonic regression [9–11].

A.1. Proofs for Results in Subsection 3.1. For the simplicity of notation, from now on denote $\delta_d = \hat{d}_{n_1}^{(1)} - d_0$, $\epsilon_i^+ = \epsilon_i'' + \epsilon_i'$, $\epsilon_i^- = \epsilon_i'' - \epsilon_i'$, $f_{UL}^+ = f(U) + f(L)$, $f_{UL}^- = f(U) - f(L)$, $R_{UL}^+ = R_U + R_L$, $R_{UL}^- = R_U - R_L$, $R_{UL}'^+ = R_U' + R_L'$, $R_{UL}'^- = R_U' - R_L'$. Recall $Y_i^+ = Y_i'' + Y_i'$ and $Y_i^- = Y_i'' - Y_i'$.

PROOF OF LEMMA 3.1. Consider the following Taylor's expansions:

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

$$
\begin{aligned}
f(L) = f(\hat{d}_{n_1}^{(1)} - Kn_1^{-\gamma}) &= f(d_0) + f'(d_0)(\delta_d - Kn_1^{-\gamma}) \\
&+ (1/2)f''(d_0)(\delta_d - Kn_1^{-\gamma})^2 + R_L,
\end{aligned}
\tag{A.3}
$$

where $R_U = f'''(\xi_1)(\delta_d + Kn_1^{-\gamma})^3/6$, $R_L = f'''(\xi_2)(\delta_d - Kn_1^{-\gamma})^3/6$, ξ_1 lies between d_0 and $\hat{d}_{n_1}^{(1)} + Kn_1^{-\gamma}$ and ξ_2 lies between d_0 and $\hat{d}_{n_1}^{(1)} - Kn_1^{-\gamma}$. Since $\hat{d}_{n_1}^{(1)}$ converges to d_0 in probability by Theorem 2.1, so do ξ_1 and ξ_2.

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

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

From Theorem 2.1, $\delta_d \overset{P}{\to} 0$; and by the Lindeberg-Feller CLT for triangular arrays, for $\gamma \in (0, 1/2)$, $(n_1^{\gamma}/n_2)\sum_{i=1}^{n_2}\epsilon_i^- \overset{P}{\to} 0$. Next we show that $R_{UL}^-/(2Kn_1^{-\gamma}) \overset{P}{\to} 0$ for $\gamma \in (0, 1)$. Hence, for $\gamma \in (0, 1/2)$ we get $\hat{\beta}_1 \overset{P}{\to} f'(d_0)$. It suffices to show both $n_1^{\gamma}R_U$ and $n_1^{\gamma}R_L$ converge to 0 in probability for $\gamma \in (0, 1)$. We only show the former; the latter follows in an analogous manner.

From the definition of R_U, we have

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

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 $f'''(\cdot)$ is uniformly bounded around d_0 and $\xi_1 \to d_0$ in probability, $f'''(\xi_1) o_P(1) = o_P(1)$. This shows that $n_1^\gamma R_U$ converges to 0 in probability for $\gamma \in (0,1)$. Obviously, $R_U = o_P(1)$.

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

$$
\begin{aligned}
\hat\beta_0 &= (2n_2)^{-1}\sum_{i=1}^{n_2} Y_i^+ - \hat d_{n_1}^{(1)}\hat\beta_1 = f(d_0) + (1/2) f''(d_0)[\delta_d^2 + K^2 n_1^{-2\gamma}] \\
&+ f'(d_0)\delta_d + (1/2) R_{UL}^+ + (2n_2)^{-1}\sum_{i=1}^{n_2}\epsilon_i^+ - \hat d_{n_1}^{(1)}\hat\beta_1 \xrightarrow{P} f(d_0) - d_0 f'(d_0).
\end{aligned}
$$

Finally, for $\gamma \in (0, 1/2)$, the weak consistency of $\hat\beta_1$ and $\hat\beta_0$ gives $\tilde d_n^{(2)} = (\theta_0 - \hat\beta_0)/(\hat\beta_1) \xrightarrow{P} d_0$. □

PROOF OF THEOREM 3.2. First, suppose $f \in \mathscr{F}_1$. From (3.2), the definitions of Y_i' and Y_i'' and the Taylor expansions (A.2) and (A.3), we get

$$
\begin{aligned}
\tilde d_n^{(2)} - d_0 &= (1/\hat\beta_1)[f(d_0) - (2n_2)^{-1}\sum_{i=1}^{n_2} Y_i^+] + \delta_d \\
&= (1/f'(d_0))[f(d_0) - (2n_2)^{-1}\sum_{i=1}^{n_2} Y_i^+] + \delta_d \\
&+ (f'(d_0)\hat\beta_1)^{-1}(f'(d_0) - \hat\beta_1)[f(d_0) - (2n_2)^{-1}\sum_{i=1}^{n_2} Y_i^+] \\
&= S_1 + S_2 \times S_3
\end{aligned}
$$

where

$$
\begin{aligned}
S_1 &= -f''(d_0)(2f'(d_0))^{-1}(\delta_d^2 + K^2 n_1^{-2\gamma}) \\
&\quad -(2f'(d_0))^{-1} R_{UL}^+ - (2f'(d_0) n_2)^{-1}\sum_{i=1}^{n_2}\epsilon_i^+, \\
S_2 &= (f'(d_0)\hat\beta_1)^{-1}[f''(d_0)\delta_d + (2Kn_1^{-\gamma})^{-1} R_{UL}^- + (2Kn_1^{-\gamma} n_2)^{-1}\sum_{i=1}^{n_2}\epsilon_i^-], \\
S_3 &= f'(d_0)\delta_d + (1/2) f''(d_0)(\delta_d^2 + K^2 n_1^{-2\gamma}) + (1/2) R_{UL}^+ + (2n_2)^{-1}\sum_{i=1}^{n_2}\epsilon_i^+.
\end{aligned}
$$

Next consider the exact stochastic orders of the terms $S_1,\ S_2$ and S_3. We start with S_1. From Theorem 2.1, $\delta_d^2 = O_P(n^{-2/3})$; for $\gamma > 0$, $n_1^{-2\gamma} = O_P(n^{-2\gamma})$, $R_U = O_P(n^{-1}) + O_P(n^{-3\gamma})$, $R_L = O_P(n^{-1}) + O_P(n^{-3\gamma})$, and $n_2^{-1}\sum_{i=1}^{n_2}\epsilon_i^+ = O_P(n^{-1/2})$. Note that these are the exact rates of weak convergence. Then, for $\gamma \in (0, 1/2)$, $S_1 = T_1 + T_2 + o_P(n^{-2\gamma} \vee n^{-1/2})$, where

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

Thus, the possible main terms of S_1 are T_1 and T_2. In the same way, we can obtain the main terms of S_2 and S_3 and then those of $S_2 \times S_3$. Finally we have $S_1 + S_2 \times S_3 = T_1 + T_2 + T_3 + R$, where

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

It is easy to see that among the three rates $n^{-2\gamma}$, $n^{-1/2}$ and $n^{\gamma-5/6}$, the first, second or last one is slowest according as 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 Lindeberg-Feller CLT for triangular arrays, Slutsky's Lemma and the Continuous Mapping Theorem, and noting that $n_1^{1/3}\delta_d$ is independent of $n_2^{-1/2}\sum_{i=1}^{n_2}\epsilon_i^+$ and $n_2^{-1/2}\sum_{i=1}^{n_2}\epsilon_i^-$ and that ϵ_i^+ is uncorrelated with ϵ_i^-, we obtain the results of the five cases for $f \in \mathscr{F}_1$ 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

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

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

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

Note that $-C_2Z_1$ can be replaced by C_2Z_1 since $N(0, 1)$ and $-N(0, 1)$ have the same distribution. In similar fashion, we obtain the asymptotic results for the other four cases.

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

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

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

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

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

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

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

where

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

Similar to the previous argument on the exact weak convergence rates, $S_1 + S_2 \times S_3 = T_1 + T_2 + R'$, where

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

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

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 $b_1 \leq b_2 \in \mathbb{R}$. Note that $\zeta_n = B_{as}(b)$ for some $b > 0$ implies $\zeta_n \to 0$ almost surely. Denote $V_i^+ \equiv \epsilon_i^{\star +} = \epsilon_i^{''\star} + \epsilon_i^{'\star}$ and $V_i^- \equiv \epsilon_i^{\star -} = \epsilon_i^{''\star} - \epsilon_i^{'\star}$. Recall $Y_i^{\star +} = Y_i^{''\star} + Y_i^{'\star}$ and $Y_i^{\star -} = Y_i^{''\star} - Y_i^{'\star}$.

PROOF OF LEMMA 3.3. The proof of Lemma 3.1 establishes the weak consistency of $\hat{\beta}_1$ for the case $\gamma \in (0, 1/2)$. In fact, under the setting of the bootstrapped two-stage procedure, the strong consistency of $\hat{\beta}_1$ can be obtained.

From the proof of Lemma 3.1, it suffices to show δ_d, $(n_1^\gamma/n_2)\sum_{i=1}^{n_2} \epsilon_i^-$ and $R_{UL}^-/(2Kn_1^{-\gamma})$ converge to 0 almost surely. Lemma A.2 shows that δ_d converges to 0 almost surely, while Lemma A.5 establishes that $(n_1^\gamma/n_2)\sum_{i=1}^{n_2} \epsilon_i^-$ converges to 0 almost surely for $\gamma \in (0, 1/2)$. Thus, it suffices to show that both $n_1^\gamma R_U$ and $n_1^\gamma R_L$ converge to 0 almost surely for $\gamma \in (0, 1)$. Next, we show the former; the latter follows analogously.

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

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

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

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

where

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

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

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

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

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

For $\gamma \in (0, 1/2)$, gives that $T_{12} \to 0,\ (P - a.s.)$ by Lemma A.5 and $T_{11} \stackrel{P^\star}{\to} 0,\ (P - a.s.)$ by Lemma A.6 and Slutsky's Lemma. Thus, $T_1 \stackrel{P^\star}{\to} 0,\ (P - a.s.)$.

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

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

Thus, for $f \in \mathscr{F}$ and $\gamma \in (0, 1/2)$, $T_2 \to f'(d_0),\ (P - a.s.)$. Therefore, we get $\hat\beta_1^\star \stackrel{P^\star}{\to} f'(d_0),\ (P - a.s.)$. □

PROOF OF THEOREM 3.4. From (3.2) and (3.3),

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

where

$$T_1 = (f'(d_0)2n_2)^{-1} n^{1/2}\sum_{i=1}^{n_2}(Y_i^{\star +} - Y_i^+)$$

$$T_2 = n^{1/2}\Bigg[(1/\hat\beta_1^\star - 1/f'(d_0))(f(d_0) - (2n_2)^{-1}\sum_{i=1}^{n_2} Y_i^{\star +}) - (1/\hat\beta_1 - 1/f'(d_0))(f(d_0) - (2n_2)^{-1}\sum_{i=1}^{n_2} Y_i^{+})\Bigg].$$

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

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

where

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

and $s^2 = Var_\star[V_i^+]$, equal to that s^2 in the proof of Lemma 3.3.

Lemma A.4 gives $s^2 \to 2\sigma^2,\ (P - a.s.)$ and Lemma A.6 gives $\sum_{i=1}^{n_2}(V_i^+ - \nu_i^+)/(s\sqrt{n_2}) \stackrel{d^\star}{\to} Z_1,\ (P - a.s.)$. Note that $\sqrt{n}/\sqrt{n_2} \to \sqrt{2/(1-p)}$. Thus, Slutsky's lemma implies

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

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

$\square$

LEMMA A.3. *For $f \in \mathscr{F}$ and $\gamma \in (0, 1/3)$, $T_2 \overset{P^\star}{\to} 0,\ (P - a.s.)$.*

PROOF. Let

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

Then,

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

It will be shown that $T_{2i} \overset{P^\star}{\to} 0,\ (P - a.s.)$, $i = 1, 2, 3, 4$ for $\gamma \in (0, 1/3)$. Thus, by Lemma 3.3 and Slutsky' Lemma, the conclusion of this lemma holds.

We establish next the convergence of the terms T_{2i}. 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

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

First consider T_{21}. We have

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

where $T'_{21} = C_n \cdot I \cdot II$ and $C_n = n^{1/2} n_1^{\gamma} (2K n_2^{1/2})^{-1}$. Lemmas A.4 and A.6 give

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

Next, it will be shown that T'_{21} converges to 0 P-almost surely for $\gamma \in (0, 5/12)$. Then, an application of Slutsky's Lemma gives $T_{21} \xrightarrow{P^\star} 0,\ (P - a.s.)$.

With the notation introduced at the beginning of this subsection and by Lemmas A.5 and A.2, we have, for $\gamma > 0$, $n_1^{\gamma} = B_{as}(-\gamma)$, $(\delta_d) = B_{as}(1/3)$, $\sum_{i=1}^{n_2} (\epsilon''_i + \epsilon'_i)/n_2 = B_{as}(1/2)$ and $\sum_{i=1}^{n_2} (\epsilon''_i - \epsilon'_i)/n_2 = B_{as}(1/2)$. Both R_U and R_L are equal to $B_{as}(1) + B_{as}(\gamma + 2/3) + B_{as}(2\gamma + 1/3) + B_{as}(3\gamma)$. Thus we have $C_n = B_{as}(-\gamma)$, $I = B_{as}(1/3) + B_{as}(-\gamma)[B_{as}(1) + B_{as}(\gamma + 2/3) + B_{as}(2\gamma + 1/3) + B_{as}(3\gamma)] + B_{as}(1/2 - \gamma) = B_{as}(1/3) + B_{as}(2\gamma) + B_{as}(1/2 - \gamma)$ and $II = B_{as}(1/3) + [B_{as}(2/3) + B_{as}(2\gamma)] + (B_{as}(1) + B_{as}(\gamma + 2/3) + B_{as}(2\gamma + 1/3) + B_{as}(3\gamma)) + B_{as}(1/2) = B_{as}(1/3) + B_{as}(2\gamma)$. Thus, for $\gamma \in (0, 1/2)$,

$$\begin{aligned} T'_{21} &= C_n \cdot I \cdot II \\ &= B_{as}(-\gamma) \times [B_{as}(1/3) + (B_{as}(2\gamma)) + B_{as}(1/2 - \gamma)] \\ &\quad \times \{B_{as}(1/3) + B_{as}(2\gamma)\} \\ &= B_{as}(2/3 - \gamma) + B_{as}(1/3 + \gamma) + B_{as}(5/6 - 2\gamma) + B_{as}(3\gamma). \end{aligned}$$

It is easy to see that when $\gamma \in (0, 5/12)$, the above upper bounds $1/2 - \gamma$, $1/4 + \gamma$, $3/4 - 2\gamma$, and 3γ are all positive. This implies that T'_{21} converges to 0 P-almost surely for $\gamma \in (0, 5/12)$. Therefore, for $\gamma \in (0, 5/12)$, T_{21} converges to 0 in probability $(P - a.s.)$.

Similarly, we can show that T_{2i}, $i = 2, 3$ or 4, converges to 0 in probability $(P - a.s.)$, but with different intervals for 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 $\gamma \in (0, 1/3)$, T_{23} converges to 0 in probability $(P - a.s.)$. It is worthwhile to note that $\mathscr{F}$ can be considered directly because the $B_{as}(1/3 - \gamma)$ term in T_{23} does not depend on $f''(d_0)$. Since $1/3 < 5/12 < 1/2$, T_{2i} converges to 0 in probability $(P - a.s.)$ for $i = 1, 2, 3, 4$ and $\gamma \in (0, 1/3)$. Thus, for $f \in \mathscr{F}$ and $\gamma \in (0, 1/3)$, T_2 converges to 0 in probability $(P - a.s.)$. □

PROOF OF THEOREM 3.5. 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 $n_1^\gamma(\delta_d) \equiv (np)^\gamma(\hat{d}_{np}^{(1)} - d_0) \xrightarrow{P} 0$. It follows, by the relationship between convergence in probability and almost sure convergence (for example, see Theorem 20.5 in Billingsley [7]), 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 3.4, write $n_{k(i)}^{1/2}(\tilde{d}_{n_{k(i)}}^{(2)\star} - \tilde{d}_{n_{k(i)}}^{(2)})$ as $-T_1 + T_2$, where *both* T_1 *and* T_2 *are now indexed by* $n_{k(i)}$. It is then not difficult to show that the conditional distribution of T_1 converges to that of $C_2 Z_1$ P-almost-surely by replacing n, n_1 and n_2 by $n_{k(i)}$, $n_{k(i),1}$ and $n_{k(i),2}$ respectively, and essentially repeating the steps in Theorem 3.4.

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

A.3. Some Auxiliary Lemmas. First we state a special almost sure convergence result on a triangular array of iid mean zero random variables. For the general result, see Proposition in Hu et al. [18].

LEMMA A.4. *If a triangular array of random variables* $\{X_{ni}\}_{i=1}^{m_n}$ *for* $n \in \mathbb{N}$ *are iid copies of a mean* 0 *random variable* X *with* m_n *increases to* ∞ *as* n *goes to* ∞ *and* $\mathbb{E}|X|^{2p} < \infty$ *for some* $p \in [1, 2)$, $P(\lim_{n\to\infty} m_n^{-1/p} \sum_{i=1}^{m_n} X_{ni} = 0) = 1$.

Suppose a triangular array of random variables $\{\epsilon_{ni}\}_{i=1}^{m_n}$ for $n \in \mathbb{N}$ are iid copies of ϵ with mean 0, where m_n increases to ∞ as n goes to ∞. Then Lemma A.4 tells that $\bar{\epsilon}_n = (1/m_n)\sum_{i=1}^{m_n} \epsilon_{ni}$ and $(1/m_n)\sum_{i=1}^{m_n} \epsilon_{ni}^2$ converge to 0 and σ^2 almost surely given $\mathbb{E}\epsilon^2 < \infty$ and $\mathbb{E}\epsilon^4 < \infty$, respectively. Further, the following lemma shows that $n^{1/2}$ is an upper boundary rate of the almost sure convergence of $\bar{\epsilon}_n$.

LEMMA A.5. *If* $\mathbb{E}\epsilon^4 < \infty$, $P(\lim_{n\to\infty} m_n^\alpha \bar{\epsilon}_n = 0) = 1$ *for each* $\alpha < 1/2$.

PROOF. A direct application of Lemma A.4 gives that if $\mathbb{E}|\epsilon|^{2p} < \infty$ for some $p \in [1, 2)$, $P(\lim_{n\to\infty} m_n^{1-1/p}\bar{\epsilon}_n = 0) = 1$. On the other hand, $\mathbb{E}\epsilon^4 < \infty$ implies that $\mathbb{E}|\epsilon|^{2p} < \infty$ for every $p \in [1, 2)$. Thus, the conclusion follows. □

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. Note that the subscripts of these random variables indicating the sample size are suppressed for the simplicity of notation and that here "n" is understood as a dummy variable, not the total sample size. Recall $V_i^+ = \epsilon_i''^\star + \epsilon_i'^\star$, $\nu^+ = E_\star[V_i^+]$, $V_i^- = \epsilon_i''^\star - \epsilon_i'^\star$ and $\nu^- = E_\star[V_i^-]$, where $E_\star$ means the expectation conditioning on the second-stage data. Since $Var_\star[V_i^+] = Var_\star[V_i^-]$, we denote both as s^2. The following lemma shows that both V_i^+ and V_i^- are asymptotically normal P-almost surely.

LEMMA A.6. *If* $\mathbb{E}\epsilon^6 < \infty$, *we have*

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

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

PROOF. We only prove the former and the latter can be shown similarly. 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 $\eta > 0$ (see, for example, Theorem2 on Page 334 of Shiryaev [34]): $\sum_i^n E_\star[\xi_{ni}^2\{|\xi_{ni}| \geq \eta\}] \to 0,\ (P-a.s.)$. Note that

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

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(\epsilon_i'')^2 - (\frac{1}{n}\sum_{i=1}^{n}\epsilon_i'')^2 + \frac{1}{n}\sum_{i=1}^{n}(\epsilon_i')^2 - (\frac{1}{n}\sum_{i=1}^{n}\epsilon_i')^2 \to 2\sigma^2,\ (P-a.s.),$$

then it is sufficient to show $\overline{\lim}_{n\to\infty}E_\star|V_1^+ - \nu^+|^3 < \infty,\ (P-a.s.)$. Since

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

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

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

By Lemma A.4, both $(1/n)\sum_{i=1}^n|\epsilon_i''|^3$ and $(1/n)\sum_{i=1}^n|\epsilon_i'|^3$ converges almost surely under the assumption $\mathbb{E}\epsilon^6 < \infty$. Therefore, $\overline{\lim}_{n\to\infty}E_\star|V_1^+|^3 < \infty,\ (P-a.s.)$, which completes the proof. □

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. (2007). Likelihood based inference for monotone response models. *Annals of Statistics 35*, 931–956.

[4] BANERJEE, M. (2009). Inference in exponential family regression models under certain shape constraints. In *Advances in Multivariate Statistical Methods, Statistical Science and Interdisciplinary Research*. Vol. **4**. World Scientific, 249–272.

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

[6] BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M., AND BRUNK, H. D. (1972). *Statistical inference under order restrictions. The theory and application of isotonic regression*. John Wiley & Sons, London-New York-Sydney. Wiley Series in Probability and Mathematical Statistics. MR0326887

[7] 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

[8] 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

[9] DUROT, C. (2002). Sharp asymptotics for isotonic regression. *Probab. Theory Related Fields* **122**, 2, 222–240. MR1894068

[10] DUROT, C. (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *Ann. Statist.* **35**, 3, 1080–1104. MR2341699

[11] DUROT, C. (2008). Monotone nonparametric regression with random design. *Math. Methods Statist.* **17**, 4, 327–341. MR2483461

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

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

[14] GEZMU, M. AND FLOURNOY, N. (2006). Group up-and-down designs for dose-finding. *J. Statist. Plann. Inference* **136**, 6, 1749–1764. MR2255594

[15] GROENEBOOM, P., HOOGHIEMSTRA, G., AND LOPUHAÄ, H. P. (1999). Asymptotic normality of the l_1 error of the grenander estimator. *The Annals of Statistics 27*, 1316–1347.

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

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

[18] HU, T. C., MÓRICZ, F., AND TAYLOR, R. L. (1989). Strong laws of large numbers for arrays of rowwise independent random variables. *Acta Math. Hungar.* **54**, 1-2, 153–162. MR1015785

[19] 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.

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

[21] LEURGANS, S. (1982). Asymptotic distributions of slope-of-greatest-convex-minorant estimators. *Ann. Statist.* **10**, 1, 287–296. MR642740

[22] MAKOWSKI, G. (1975). A rate of convergence for a nondecreasing regression estimator. *Bull.*

Inst. Math. Acad. Sinica **3**, 1, 61–64. MR0397975
[23] MAKOWSKI, G. G. (1973). Laws of the iterated logarithm for permuted random variables and regression applications. *Ann. Statist. 1*, 872–887. MR0343358
[24] MORGAN, B. J. T. (1985). The cubic logistic model for quantal assay data. *Applied Statistics* **34**, 2, 105–113.
[25] MORGAN, B. J. T. (1992). *Analysis of Quantal Response Data*. Chapman & Hall.
[26] 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.
[27] MÜLLER, H.-G. AND STADTMÜLLER, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15**, 2, 610–625. MR888429
[28] OSBORNE, C. (1991). Statistical calibration: A review. *International Statistical Review 59*, 309–336.
[29] RAI, K. AND RYZIN, J. V. (1981). A generalized multihit dose-response model for low-dose extrapolation. *Biometrics 37*, 341–352.
[30] ROBBINS, H. AND MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Statistics 22*, 400–407. MR0042668
[31] 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
[32] ROSENBERGER, W. F. (1996). New directions in adaptive designs. *Statist. Sci. 11*, 137–149.
[33] ROSENBERGER, W. F. AND HAINES, L. M. (2002). Competing designs for phase i clinical trials: a review. *Stat. Med. 21*, 2757–2770.
[34] SHIRYAEV, A. N. (1995). *Probability*, Second ed. Springer-Verlag.
[35] STANISWALIS, J. G. AND COOPER, V. (1988). Kernel estimates of dose response. *Biometrics* **44**, 4, 1103–1119. MR981002
[36] 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
[37] WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9**, 2, 443–448. MR606630

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