Two-Stage Plans for Estimating the Inverse of a Monotone Function

Runlong TANG
Department of Statistics
Virginia Tech
Blacksburg, VA 24061
(runlongt@princeton.edu)

Moulinath BANERJEE
Department of Statistics
University of Michigan
Ann Arbor, MI 48109
(moulib@umich.edu)

George MICHAILDIS
Department of Statistics
University of Michigan
Ann Arbor, MI 48109
(gmichail@umich.edu)

Shawn MANKAD
Department of Decision
Operations and Information Technologies, Robert H. Smith
School of Business, University of Maryland
College Park, MD 20740
(smankad@umich.edu)

This study investigates two-stage plans based on nonparametric procedures for estimating an inverse regression function at a given point. Specifically, isotonic regression is used at stage one to obtain an initial estimate followed by another round of isotonic regression in the vicinity of this estimate at stage two. It is shown that such two-stage plans accelerate the convergence rate of one-stage procedures and are superior to existing two-stage procedures that use local parametric approximations at stage two when the available budget is moderate and/or the regression function is "ill-behaved." Both Wald- and likelihood ratio-type confidence intervals for the threshold value of interest are investigated and the latter are recommended in applications due to their simplicity and robustness. The developed plans are illustrated through a comprehensive simulation study and an application to car fuel efficiency data.

KEY WORDS: Experimental design; Isotonic regression; Two-stage plan.

1. INTRODUCTION

Threshold estimation is a canonical statistical estimation problem with numerous applications in science and engineering. Here is an interesting motivating example. In recent years, an important consideration for both car manufacturers and potential buyers in the United States is the fuel efficiency (FE) of the vehicle, expressed in miles per gallon (MPG). The National Highway Traffic Safety Administration (NHTSA) regulates the Corporate Average Fuel Economy (CAFE) standards to encourage automobile manufacturers to improve the average fuel efficiency of their fleets of vehicles. The CAFE standard for 2012 models is 29.8 MPG, set to increase to 34.3 MPG in 2016 according to the Environmental Protection Agency rule that came into effect in August 2012. To encourage higher FE, manufacturers are subject to a penalty if the average FE of their fleets falls below the CAFE standard. Moreover, a so-called gas guzzler tax is imposed on cars with low FE in accordance with the U.S. Energy Tax Act of 1978.

The data on fuel efficiency as a function of the vehicle’s horse power, which is a key component, for the 2012 models is shown in Figure 1 (for a detailed discussion of the data and CAFE standards, see Section 4). An expected decreasing relationship is observed and it is of interest to identify the horse power threshold at which the fuel efficiency meets the current CAFE, as well as the 2016 standard.

The data plot indicates that fitting a precise parametric model may be challenging, while it is rather straightforward to fit a monotonically decreasing nonparametric one and obtain the fuel efficiency threshold for the target values of around 30 and 40. Further, it is also desirable to assign a confidence interval around the estimate and an interesting question addressed in this article is whether an adaptive procedure can lead to improved precision for such threshold estimates.

The topic of using adaptive procedures in a design setting for threshold estimation models has been recently studied in the literature (Lan, Banerjee, and Michailidis 2009; Tang, Banerjee, and Michailidis 2011). Given our motivating data application, the basic model considered in this article is the same as that in Tang, Banerjee, and Michailidis (2011): \( Y = m(X) + \epsilon \), where the design point \( X \) takes values in \([a, b]\), the regression function \( m \) is monotone, for the sake of presentation henceforth assumed nondecreasing, and the random error \( \epsilon \) has mean 0 and finite variance \( \sigma^2 \). The quantity to be estimated is the threshold \( \theta_0 \), that is, \( m^{-1}(\theta_0) \) for some prespecified \( \theta_0 \); this is the value beyond which the mean response exceeds level \( \theta_0 \).

The employed two-stage adaptive procedure in Tang, Banerjee, and Michailidis (2011) works as follows: (i) in the first stage, it uses a portion \( p \) of the design budget to obtain an initial
estimate of $d_0$; (ii) in the second stage, the remaining portion $(1-p)$ of the budget is used to obtain more sample points in a small neighborhood of that estimate; (iii) finally, an improved estimate based on the second-stage data is constructed. This “zoom-in” sampling leads to accelerated convergence rates of the second stage estimators for $d_0$ compared to the standard ones that use all the data in one shot. Specifically, for the inverse regression problem, the rate of convergence can be accelerated from $n^{1/3}$ to $n^{1/2}$ by employing a local linear approximation (Tang, Banerjee, and Michailidis 2011), where $n$ denotes the total budget available. Hence, tighter confidence intervals can be constructed that have the correct nominal coverage with the same budget as standard one-stage procedures, or alternatively one can reduce the design budget and still have quality confidence intervals.

Our problem is closely related to dose-response (Rosenberger and Haines 2002; Dette, Neumeyer, and Pliz 2005; Bhattacharya and Lin 2010) and statistical calibration studies ( Osborne 1991) (for additional references see Tang, Banerjee, and Michailidis 2011)). While the above strategy—a first-stage estimate using isotonic regression, followed by a local linear approximation—gives a $\sqrt{n}$-consistent second-stage estimator, its success heavily hinges upon the (approximate) linearity of the regression function $m(\cdot)$ in the vicinity of $d_0$. Small departures from linearity do not adversely affect the results (especially when the budget is large enough to allow for significant “zooming-in” at the second stage), but severe departures are a totally different matter as illustrated next.

Consider a monotone regression function exhibiting strong nonlinearity around $d_0 = 0.5$, for example, given by $m(x) = (1/40)\sin(6\pi x) + 1/4 + (1/2)x + (1/4)x^2$ with $x \in [0, 1]$ (see the plot with label IsoSin in the left panel of Figure 4). The coverage rates and average lengths of the confidence intervals obtained from the two-stage adaptive strategy based on isotonic regression and a local linear approximation for selected total budget sizes ($n = 100, 300, 500$), varying portions $p$ allocated to the first stage and different noise levels ($\sigma = 0.1, 0.3, 0.5$) are depicted in Figure 2.

It can be seen that for the majority of portions $p$, the confidence intervals constructed from this adaptive strategy fail miserably in terms of coverage rates and exhibit relatively large lengths. Of interest is the fact that for large $p$, the coverage rates indeed approach the nominal level. In practice, however, it is not possible to choose an appropriate $p$ without prior information on $m$. Further, even for large $p$’s, the confidence intervals are excessively wide, especially for large noise levels.

In contrast, a two-stage adaptive strategy based on employing isotonic regression at both stages, which will be fully developed in this article, overcomes these difficulties. Such a strategy would be also desirable for the motivating data application, due to high variability in the vicinity of the CAFE thresholds, as seen in the scatterplots of Figure 1. In Figure 3, the coverage rates and average lengths using our new strategy are shown for the same settings as above, but with $p = 1/4$ (for more on this universal choice of $p$ see Section 2). It can be seen that this wholly non-parametric strategy overcomes the previous difficulties, proves robust to the level of local nonlinearity of the regression function $m$ and, as argued in Section 2, is easy to implement.

The remainder of the article is organized as follows: in Section 2 the adaptive procedure is introduced and the main results presented. Section 3 presents extensive simulations results, while an interesting application of the methodology to fuel efficiency data is shown in Section 4. Section 5 concludes. Proofs are sketched in the appendix.

## 2. TWO-STAGE ADAPTIVE PROCEDURES

### 2.1 An Overview of the Isotonic Regression Procedure

We provide a brief description of the one-stage isotonic regression procedure (OSIRP). Specifically, given $n$ fixed, random design points $\{X_i\}_{i=1}^n$ in $[a, b]$ distributed according to a
continuous design density \( g \) and the corresponding responses \( \{Y_i\}_{i=1}^n \), obtained from the proposed model assuming the design points have already been ordered, the constrained isotonic regression estimator of \( m(\cdot) \) is given by

\[
m_I(x) = m_I^I[1[x \in [a, X_1]] + \sum_{i=1}^{n-1} m_I^I[1[x \in [X_i, X_{i+1}]]] + m_n^I[1[x \in [X_n, b]]],
\]

where \( m_I^I \) is defined as \( \arg\min_{m_1 \leq m_2 \leq \cdots \leq m_n} \sum_{i=1}^n (Y_i - m_i)^2 \). This minimizer exists uniquely, has a nice geometric characterization as the slope of the greatest convex minorant of a stochastic process and is readily computable using the pool adjacent violators algorithm (PAVA) (see, e.g., Robertson, Wright, and Dykstra 1988). Then, for a prespecified value \( \theta_0 \in (m(a), m(b)) \), the one-stage isotonic regression estimator of \( d_0 \) is defined by

\[
d_I = m_I^{-1}(\theta_0) = \inf\{x \in [a, b] : m_I(x) \geq \theta_0\},
\]

where \( \inf\{\theta\} = b \). There is a mild assumption on the regression function and the design density, namely,

**Assumption A**: \( m \) is once continuously differentiable in a neighborhood of \( d_0 \) with positive derivative \( m'(d_0) \) and \( g \) is positive and continuous at \( d_0 \).

Under this assumption, the asymptotic distribution of \( d_I \) is given by (see Tang, Banerjee, and Michailidis 2011):

\[
n^{1/3}(d_I - d_0) \xrightarrow{d} C_{d_I} g(d_0)^{-1/3} Z,
\]

where \( C_{d_I} = [4\sigma^2/m'(d_0)^2]^{1/3} \) and \( Z \) follows the standard Chernoff distribution (Groeneboom and Wellner 2001). This result can be used to construct a \( 1 - \alpha \) Wald-type confidence interval for \( d_0 \):

\[
\left[ d_I \pm n^{-1/3} C_{d_I} g(d_0)^{-1/3} q(Z, 1 - \alpha/2) \right],
\]

where the hats denote consistent estimates and \( q(\xi, \tau) \) is the lower \( \tau \) quantile of a random variable \( \xi \).

An alternative is to construct confidence intervals through likelihood ratio (LR) testing. Specifically, the hypotheses of interest are

\[
H_0 : m^{-1}(\theta_0) = d_0 \leftrightarrow H_a : m^{-1}(\theta_0) \neq d_0.
\]

Then, the LR test statistic is given by

\[
2 \log \lambda_I = 2 \log \lambda_I(d_0) = 2 \left[ l_a(m_I, 0) - l_a(m_{1c}, 0) \right],
\]

where \( l_a(m, \sigma) = -(2\sigma^2)^{-1} \sum_{i=1}^n (Y_i - m(X_i))^2 \), \( m_{1c} \) is the constrained isotonic regression of \( m \) under \( H_0 \) and \( \hat{\sigma} \) a consistent estimate of \( \sigma \). It is known that \( m_{1c} \) uniquely exists (see Banerjee 2000). The asymptotic distribution of \( 2 \log \lambda_I \) under \( H_0 \) can be obtained from Banerjee (2009) (see Theorem 1.1 and also the first paragraph of Section 1.4, the concluding discussion) \( 2 \log \lambda_I \xrightarrow{D} \mathbb{D} \). Here \( \mathbb{D} \) is a “universal” random variable not depending on the parameters of the model (Banerjee and Wellner 2001). This result allows us to construct a \( 1 - \alpha \) LR-type confidence region for \( d_0 \):

\[
\{x \in [a, b] : 2 \log \lambda_I(x) \leq q(\mathbb{D}, 1 - \alpha)\}.
\]
The LR-type confidence region can be shown to be an interval and is typically asymmetric around $d_1$, unlike the Wald-type one. Its main advantage is that only $\sigma$ needs to be estimated for its construction, whereas for the Wald confidence interval, estimation of $m'(d_0)$ is also needed, a significantly more involved task.

Remark 2.1. The use of the term LR statistic in connection with (2.5) needs to be clarified. Under a normality assumption on the errors, $2 \log \lambda_1$ is, indeed, a proper likelihood ratio statistic; otherwise, it is more accurately a **residual sum of squares statistic** which can be interpreted as a “working likelihood ratio statistic” where the normal likelihood is used as a working likelihood. In this article, we do not assume normality of errors but continue to use the term LR statistic for $2 \log \lambda_1$ in the above sense.

2.2 Adaptive Two-Stage Procedures

As noted in Introduction, adaptive two-stage procedures can lead to accelerated convergence rates and hence to sharper confidence intervals for $d_0$. The main steps of such a two-stage fully nonparametric procedure are outlined next:

1. Denote by $p \in (0, 1)$ the sample proportion to be allocated in the first stage and by $n_1 = \lfloor np \rfloor$ and $n_2 = n - n_1$, the corresponding first- and second-stage sample sizes, respectively.

2. Generate the first-stage data $\{(X_{1,i}, Y_{1,i})\}_{i=1}^{n_1}$ with a design density $g_1$ on $[a, b]$. Then, compute a first-stage monotone nonparametric estimator $\hat{m}_1$ of $m$ and obtain the corresponding first-stage estimator $d_{1,t} = \hat{m}_1^{-1}(\hat{\theta}_0)$ of $d_0$ for a prespecified value $\theta_0$.

3. Specify the second-stage sampling interval $[L_1, U_1] = [d_{1,t} \pm C_1 n_1^{-3}] \cap [a, b]$ where $C_1 > 0$ and $0 < \gamma_1 < 1/3$; recall that $n_1^{1/3}$ is the convergence rate of $d_{1,t}$.

4. Obtain the second-stage data $\{(X_{2,i}, Y_{2,i})\}_{i=1}^{n_2}$ with a design density $g_2$ on $[L_1, U_1]$. Employ these data and a nonparametric procedure (which could be different from the one used previously) to compute a monotone second-stage estimator $\hat{m}_{2,t}$ and, as in the first stage, the corresponding $d_{2,t}$.

5. Construct confidence intervals for $d_0$ using the asymptotic distribution of $d_{2,t}$.

Remark 2.2. Choosing $\gamma_1 < 1/3$ ensures that the stage-two sampling interval contains $d_0$ with probability going to 1.

Remark 2.3. The key intuition behind the proposed two-stage procedure is that the resolution of the sampling design increases. Specifically, if one considers sampling based on a uniform grid, at the second stage the sampling interval has length $2C_1 n_1^{-3}$. Suppose that the second-stage design points are equispaced over this interval, the resolution of the resulting grid at which responses are measured is $O(n_1^{-3}/n_2) = O(n^{-1+\gamma_1})$ and this is driving factor for the realized rate of convergence of the two-stage estimator, as formally established in Proposition 1, in contrast to that obtained in the OSIRP ($n^{-1/3}$), since in the latter procedure the resolution of the sampling design is simply $O(n^{-1})$.

2.3 Asymptotic Properties of Two-Stage Estimators

We discuss the properties of the two-stage procedure, where isotonic regression is employed in both stages (henceforth, IR+IR).

**Proposition 1.** Consider the IR + IR procedure. Let the design density at the second stage be given by: $g_2(x) = (C_1 n_1^{-3})^{-1} \psi((x - d_{1,t})/C_1 n_1^{-3})$ where $\psi$ is a Lebesgue density on $[-1, 1]$ that is positive at 0 and continuous in a neighborhood of 0. Thus, $g_2$ is simply $\psi$ renormalized to the sampling interval at stage two. Assume that $m'$, the derivative of $m$, exists and is continuous in a neighborhood of $d_0$ and $m'(d_0) > 0$. Let $d_{2,t} = m_{2,t}^{-1}(\hat{\theta}_0)$ where $m_{2,t}$ is the isotonic estimator of $m$ constructed from the second-stage data. Then, $n^{(1+\gamma_1)/3}(d_{2,t} - d_0) \to C_{d_2} Z$, where $C_{d_2} = C_1 g(d_0)^{1/3} n^{\gamma_1/3}$.

From Proposition 1, a Wald-type $1 - \alpha$ asymptotic confidence interval for $d_0$ is given by

$$[d_{2,t} \pm n^{-1+\gamma_1/3} C_{d_2} q(Z, 1 - \alpha/2)].$$

Remark 2.4. A consequence of the accelerated rate of convergence obtained with the IR+IR strategy is that the asymptotic relative efficiency (ARE) of the two-stage estimator $d_{2,t}$ with respect to the one-stage estimator $d_{1,t}$ is

$$\text{ARE}(d_{2,t}, d_{1,t}) = \left(\frac{(1 - p)p^\gamma \psi(0)}{C_1 g(d_0)}\right)^{1/3} n^{\gamma_1/3},$$

which diverges to infinity.

Note that, in the generic description of the two-stage procedure above, we use a confidence interval for $d_0$ that relies on the asymptotic distribution of a point estimate computed at stage two. However, this is not the only way to proceed at stage two. Having collected the second-stage data at the beginning of Step 4, we can bypass point estimation altogether and construct a confidence interval using likelihood ratio inversion. This alternative possibility is discussed below. Also, as will be explained in the practical implementation, the construction of $[L_1, U_1]$ is achieved in practice by constructing a high probability confidence interval for $d_0$ from the stage one data. This also opens up the possibility of bypassing point estimates at stage one in favor of a likelihood ratio inversion based confidence interval, a point that we come to later.

An alternative LR-type CI can be constructed as follows: the LR-type test statistic at stage two for testing $H_0: d_0 = m^{-1}(\hat{\theta}_0)$ is

$$-2 \log \lambda_{2,t}(d_0) = 2 \left[ l_0(m_{2,t}, \hat{\sigma}) - l_0(m_{2,t}, \hat{\sigma}) \right],$$

where $l_0(m, \sigma) = -\frac{1}{2} \sum_{i=1}^{n_2} (Y_{2,i} - m(X_{2,i}))^2$, $m_{2,t}$ is the constrained estimator of $m$ under the null hypothesis $H_0$ and $\hat{\sigma}$ is a consistent estimate of $\sigma$.

**Proposition 2.** Under the assumptions of Proposition 1, and the null hypothesis $H_0$: $m^{-1}(\hat{\theta}_0) = d_0$ holding true, we have $2 \log \lambda_{2,t} \xrightarrow{d} \mathbb{D}$, where $\mathbb{D}$ is as before.
Finally, from Proposition 2, an LR-type \((1 - \alpha)\) asymptotic confidence interval for \(d_0\) is given by

\[
\{ x \in [a, b] : 2 \log \lambda_{2,1}(x) \leq q(\mathcal{D}, 1 - \alpha) \}. \tag{2.9}
\]

For the theoretical derivations in connection with Propositions 1 and 2, see the Appendix.

**Remark 2.5.** We have focused on the case of two-stage adaptive designs and the acceleration of the convergence rate by an IR+IR strategy. Obviously, one can extend it to multiple stages and continue using isotonic regression. As outlined in Section S1 in the supplement, it can be established that the convergence rate of such a procedure would come arbitrarily close to the \(\sqrt{n}\) parametric rate, if enough stages are employed, but would not achieve it.

### 2.4 Implementation Issues

We discuss, next, the main steps for implementing the IR+IR strategy in practice. Specifically, we address the following: (i) estimation of \(\sigma^2\), (ii) estimation of \(m'(d_0)\), (iii) determination of the first-stage sampling proportion \(p\), and (iv) the first-stage sampling proportion \(p\).

**Implementation of IR + IR:** For the estimation of \(\sigma^2\) at Stage 1, we employ the nonparametric estimator proposed by Gasser, Sroka, and Jennen-Steinmetz (1986), and for the estimation of \(m'(d_0)\), the local quadratic regression procedure proposed by Fan and Gijbels (1996); some details are provided in Section 3. Next comes the determination of the second-stage sampling interval. Recall, that the theoretical formula for the interval is given by

\[
[d_{1,1} \pm C_1 n_1^{\gamma_1}],
\]

with \(C_1>0\) and \(\gamma_1 \in (0, 1/3)\). While any such interval will contain \(d_0\) with probability going to 1 in the long run, in practice we would like to ensure that our prescribed sampling interval \([L_1, U_1]\) does trap \(d_0\) with high probability. The practical determination of \([L_1, U_1]\) is, therefore, achieved through a high probability confidence interval for \(d_0\) from Stage 1 data. Consider the the following intervals \(1 - \beta\) Wald-type confidence interval

\[
[d_{1,1} \pm n_1^{1/3} \hat{C}_{d_0} g(d_{1,1})^{1/3} q(Z, 1 - \beta/2)] \cap [a, b], \tag{2.10}
\]

where the computation of \(\hat{C}_{d_0}\) involves estimating both \(\sigma^2\) and \(m'(d_0)\) and where \(\beta\) is a small positive number such as 0.01. Using this, in practice, as \([L_1, U_1]\) amounts to choosing \(C_1\) and \(\gamma_1\) such that

\[
C_1 n_1^{-\gamma_1} = n_1^{1/3} \hat{C}_{d_0} g(d_{1,1})^{-1/3} q(Z, 1 - \beta/2).
\]

That is, \(\gamma_1 = 1/3\) and \(C_1 = \hat{C}_{d_0} g(d_{1,1})^{-1/3} q(Z, 1 - \beta/2)\). Although \(\gamma_1 = 1/3\) is not in \((0, 1/3)\) as required by our theoretical results, it nevertheless provides a good approximation in practice, since it lies at the boundary of that interval.

As far as the first-stage sampling proportion is concerned, one would like to choose this in such a way as to increase the precision of the second-stage isotonic estimator. Proposition 1 shows that the second-stage estimator is asymptotically unbiased and that its standard deviation is proportional to \((1 - p) p^{\gamma_1} - 1/3\). For a fixed \(p\), this is minimized when \(\log(1 - p) + \gamma_1 \log p\) is maximized, which happens when \(p = p_{opt} = \gamma_1/(1 + \gamma_1)\). As \(\gamma_1\) approaches 1/3, \(p_{opt}\) approaches 1/4. Thus, the optimal practical allocation of budget at Stage 1 is 25%.

From Stage-2 data, we can construct a confidence interval (CI) for \(d_0\) based on \(d_{2,1}\), following Proposition 1 in which case both \(m'(d_0)\) and \(\sigma^2\) need to be updated. Alternatively, we can use likelihood ratio inversion to get a CI of desired coverage for \(d_0\), following Proposition 2 in which case only the estimate of \(\sigma^2\) needs to be updated. Finally, a third option is to bypass the estimation of \(m'(d_0)\) altogether by prescribing a high probability LR-based confidence interval for \(d_0\) as \([L_1, U_1]\) in Stage 1 (see (2.6) and then using LR inversion at Stage 2 as well. While this procedure does not quite fall within the purview of our theoretical results it is a natural methodological choice and avoids the estimation of \(m'(d_0)\) altogether; furthermore, comparisons among these three approaches based on elaborate simulation studies demonstrate that it is superior to the other two in practice.

### 3. PERFORMANCE EVALUATION OF THE ADAPTIVE PROCEDURES

In this study, the following procedures are compared: (i) POSISRIP: practical one-stage procedure based on isotonic regression with Wald and LR CIs, (ii) PTSIRP-Wald: practical two-stage procedure based on isotonic regression for both stages and using Wald CIs both for selecting \(L_1; U_1\) and constructing the final CI, (iii) PTSIRP-LR: practical two-stage procedure similar to PTSIRPWald but employing LR CIs in both stages, and (iv) PABLTP: the procedure from Tang et al. (2011) that uses isotonic regression followed by a local linear approximation and bootstrapping for constructing CIs for \(d_0\). The use of the qualifier “Practical” before the various procedures above is to emphasize the point that they involve estimates of nuisance parameters, as explained below.

The simulation settings are as follows: the design space is the \([0, 1]\) interval and the regression functions considered: the sigmoid function \(m(x) = \exp(4(x - 0.5))/(1 + \exp(4(x - 0.5)))\), the quadratic function \(m(x) = x^2\) and the isotonic sine function \(m(x) = (1/40) \sin(6\pi x) + 1/4 + (1/2)x + (1/4)x^2\). The target point \(d_0\) is 0.4, 0.5, or 0.6, while the random error follows a \(N(0, \sigma^2)\) distribution, with \(\sigma\) taking values 0.1 and 0.3. The total sample size \(n\) ranges from 100 to 500 in increments of 100. All design densities \(g, g_1,\) and \(g_2\) are uniform, while the confidence level for all CIs is set to 0.95. The results presented are based on 1000 replicates. For PABLTP, we set the first stage sample proportions \(p = 0.7\) in order obtain accurate coverage rates for all functions. (Yet, as shown in Figure 2, in some cases good coverage rates are achieved at the cost of large average lengths.) For all other two-stage procedures, we set \(p\) to be the asymptotically optimal proportion of 0.25. The quantiles of \(\mathcal{D}\) and \(\mathcal{Z}\) for constructing the second-stage sampling intervals for PTSIRP are set to be 4 and 2, respectively, corresponding to \(\beta = 0.01\).

When estimating \(\sigma\) and \(m'(d_0)\) in the second stage, only Stage-2 points are used to stay strictly within the scope of the methods used for these purposes. With smaller budgets as in the real data example, we follow the natural practice of combining both stage samples for updating estimates of \(\sigma\) and \(m'(d_0)\), which makes second stage results more reliable.

To gain insight into the simulation results, we depict the plots of the functions under consideration together with their derivatives (see Figure 4).
The coverage rates and average lengths of the 95% confidence intervals for $d_0$ are shown in Figures 5 and 6.

It can be seen that for the quadratic and sigmoid functions, the proposed two-stage procedures perform well with the coverage being about the nominal level 95% for all $d_0$’s, sample sizes, and noise levels considered. Further, their average lengths are fairly comparable. In contrast, PABLTS shows inferior performance for larger noise and smaller sample sizes for the quadratic function.

The isotonic sine function proves the most challenging. The left panel of Figure 6 shows the coverage rates of the practical procedures. As discussed in the introduction and seen in the figure, this function exhibits strong nonlinearity causing the local linear approximation PABLTS to feature very poor coverage rates. Also, note that, for the case with $d_0 = 0.5$ the coverage rates of the confidence intervals from POSIRP-Wald and PTSIRP-Wald are consistently lower than 95%. This behavior is caused by inaccurate estimation of $m'(d_0)$ as illustrated in the online supplementary material (Section 2, Figure 1). The true value of $m'(d_0)$ is around 0.279 and the corresponding kernel estimators of $m'(d_0)$ are usually around 0.75, significantly larger than the true value. This makes the confidence interval far too short to cover $d_0$ and consequently, the coverage rates behave erratically.

Figure 5. The left and right panels show the coverage rates and average lengths of the 95% confidence intervals for $d_0$ from the practical procedures with the sigmoid and quadratic functions and different values of $\sigma$, $d_0$, and $n$. 
Full details for the estimation of $m'(d_0)$, which uses a local quadratic regression procedure, are available in Section 4 of Tang, Banerjee, and Michailidis (2011). An asymptotically optimal bandwidth, given in equation (3.20) on page 67 of Fan and Gijbels (1996), is employed for this purpose. This local bandwidth minimizes the asymptotic MSE, and indeed with large sample sizes, we find that $m'(d_0)$ is estimated accurately and the coverage rates approach the nominal level. Further emphasizing the importance of the derivative estimate and as illustrated in Figure 2 of the online supplementary material, if we repeat the procedures with perfect knowledge of $m'(d_0)$, then coverage rates are about the nominal level of 95% for the sample sizes considered.

Fortunately, for this wiggly isotonic sine function, POSIRP-LR and PTSIRP-LR have good coverage rates for all simulation cases. This indicates that LR-type confidence intervals are usually robust with different regression functions. The average lengths of the confidence intervals are shown in the right panel of Figure 6. Unsurprisingly, PTSIRP-LR achieves shorter average lengths since it is a two-stage procedure.

In summary, we find that when the underlying regression function is well-behaved, the more aggressive PABLTS performs well. However, the conservative but stable PTSIRP-LR offers a robust procedure that performs well, even when the underlying function exhibits strong nonlinearities.

4. AN APPLICATION TO FUEL EFFICIENCY STANDARDS

As discussed in Introduction, car fuel efficiency (FE) is an important issue for both manufacturers and consumers, due to new CAFE standards. Note that while the CAFE standards are regulated by the NHTSA, the vehicle FE is assessed by the Environmental Protection Agency (EPA). From 2008 onwards, the EPA measures the fuel efficiency of a vehicle in two testing modes: city and highway, taking into consideration different speeds and acceleration, as well as air conditioning usage and colder outside temperatures, in an effort to better approximate real-world fuel efficiency. From the unadjusted city and highway fuel efficiency, the unadjusted combined fuel efficiency is calculated as follows (see www.epa.gov):

$$\text{Combined FE} = \frac{1}{0.495/\text{City FE} + 0.351/\text{Highway FE}} + 0.15.$$  

The data for this study were extracted from the government website www.fueleconomy.gov that includes all FE data for all 2012 car models available to U.S. consumers. This dataset contains the unadjusted city, highway, and combined fuel efficiency for 3979 models, together with their horse power. We collected the horse power data for 1477 nonhybrid vehicles with automatic transmission gearboxes and natural aspiration engines (i.e., excluding turbo engines and plug-in hybrid vehicles), to have a relatively homogeneous dataset.

The objective of our analysis is to estimate the following model $FE = m(\text{HP})$ (or $\text{HP} = m^{-1}(FE)$ and then identify the horse power at which the combined FE is equal to 30 MPG, around the 2011 CAFE standard. Hence, we are interested in estimating $d_0 = m^{-1}(30)$.

The scatterplot in the left panel of Figure 1 shows the combined FE of these 1477 vehicles as a function of their horse power and indicates a decreasing relationship. Notice that there are multiple vehicle models with the same horse power, but different FE. To simplify the analysis, we add a small jitter to the original horse power to obtain a unique horsepower for every FE observation, whose scatterplot is given in the right panel of Figure 1. The jitter added is between ±1 to ensure that the ordering of samples by horsepower remains unchanged.

Given that this is an “observed” dataset, we will emulate the design setting (for a similar strategy see also Lan, Banerjee,
Figure 7. The top panels show one-stage procedures: POSIRP-Wald and POSIRP-LR. The bottom panels show two-stage procedures: PTSIRP, PTSIRP-LR and PABLTSP. Numbers denote first- and second-stage samples, vertical lines denote corresponding confidence intervals, and “X” marks the final point estimate.

The five procedures considered are: POSIRP-Wald, POSIRP-LR, PTSIRP-Wald, PTSIRP-LR, and PABLTSP. The fitted models are shown in Figure 7 and the confidence intervals obtained summarized in Table 1. It is interesting to note that only the LR-based procedures produce confidence intervals that cover the “true value.” The two-stage procedure PTSIRP-LR achieves much shorter interval lengths. The PTSIRP-Wald CIs are too short, resulting in their missing the “true value.”

Additional results from using different allocations in the two stages are provided in Figure 3 of the online supplementary material, where we see that in almost all cases PTSIRP-LR tends to cover the “true value” of 187 with better point estimates. POSIRP-Wald and PTSIRP-Wald continue to struggle due to estimation difficulties with \( m'(d_0) \).

Next, we perform another experiment to assess the reliability of the procedures with the FE dataset. We treat the data from the 1477 vehicles as the population, and sample from it according to different overall budgets of size 20, 30, 40, 50, 60, 70, 80, 90, 100. Given the modest budgets, we combine samples from both stages for estimation of auxiliary parameters and inversion of the likelihood ratio. The results, averaged over 500 repetitions for each budget size, are depicted in Figure 8. Note that POSIRP-Wald and PTSIRP-Wald still struggle to maintain coverage rates due to difficulties of auxiliary parameter estimation. The local linear approximation in PABLTSP also faces difficulties with such small budgets. On the other hand, POSIRP-LR and PTSIRP-LR maintain good coverage for all budgets. As noted before, PTSIRP-LR outperforms its one-stage counterpart with narrower intervals. Considering the overall budgets investigated in this experiment, PTSIRP-LR performs well with an extremely small fraction of the overall data, illustrating its utility in the context of very large datasets, a topic of further discussion in the Discussion section.
Table 1. Data analysis results for the five practical procedures

| Procedure      | Estimator | "Bias" | 95% CI         | Coverage | Length | n |
|----------------|-----------|--------|----------------|----------|--------|---|
| POSIRP-Wald    | 165.022   | 21.978 | [151.595, 178.450] | No       | 26.855 | 80|
| POSIRP-LR      | 165.022   | 21.978 | [135.887, 301.439] | Yes      | 165.552 | 80|
| PTSIRP-Wald    | 213.221   | 26.221 | [205.042, 221.400] | No       | 16.358 | 40,40|
| PTSIRP-LR      | 194.557   | 7.557  | [148.812, 225.955] | Yes      | 77.143 | 40,40|
| PABLTS TP      | 169.509   | 17.491 | [145.982, 172.964] | No       | 26.982 | 80|

Finally, we return to the task discussed in the introductory section of estimating the horse power at which the combined FE is equal to the upcoming 2016 CAFE standard of \( d_0 = m^{-1}(34) \). Employing isotonic regression on the entire sample yields a “true” value of \( d_0 \) of around 155, with corresponding 95% confidence interval [143.360, 166.052]. Following the same procedure and budget allocations as above, PTSIRP-LR yields \( \hat{d}_0 = 166.204 \) with corresponding 95% confidence interval [143.360, 166.052].
interval \([145.599, 175.450]\), as shown in Figure 9. PTSIRP-LR covers the “true” value with a reasonably sized interval, while using a small fraction of the overall budget.

The upshot of the analysis is that the two-stage LR-based procedure offers superior performance to its competitors even with smaller budgets and when the underlying function exhibits nonlinearities.

5. DISCUSSION AND CONCLUDING REMARKS

In this article, we considered the estimation of the inverse of a monotone regression function at a given point in a design setting. The results strongly suggest that a two-stage procedure using isotonic regression in both stages coupled with calculation of likelihood-ratio based confidence intervals is agnostic to the local structure in the vicinity of the parameter of interest, requires minimal tuning and exhibits superior performance.

The reader may wonder whether an alternative nonparametric procedure at stage one, with a faster than the \(n^{1/3}\) convergence rate of isotonic regression may offer advantages to the proposed strategy. We have investigated smoothed isotonic regression (Tang 2011) which, in a single stage, exhibits a convergence rate of \(n^{2/3}\) and when repeated in the second stage exhibits the same acceleration pattern as isotonic regression provided the bandwidth is appropriately chosen (for a detailed discussion of this subtle issue see Tang 2011). However, extensive numerical work shows that no significant performance gains are realized, compared to using isotonic regression in both stages, while at the same time a bandwidth parameter needs to be carefully specified. Indeed, a strategy based on isotonic regression in the first stage, followed by smooth isotonic regression in the second stage struggles with the estimation of the iso-sine function presented in Figure 2.

Another point that needs to be highlighted is that the methods of this article extend immediately to discrete response models. In many dose-response problems, the response to a dose is not a continuous random variable; rather, it is the number of individuals who showed a reaction to the dose. So, consider, for example, a Bernoulli dose-response model, where dose \(x\) is administered to an individual and the probability of this individual reacting to the dose, say \(m(x)\) is a monotone increasing function of \(x\): higher doses make it more likely for an individual to experience a reaction. Consider a budget of size \(n\). At Stage 1, \(n_1\) iid dose-response samples \(\{Y_i^{(1)}, X_i^{(1)}\}_{i=1}^{n_1}\) are recorded, where \(Y_i^{(1)}\) is 1 if the individual treated with dose level \(X_i^{(1)}\) reacted and 0 otherwise. Thus, \(Y_i^{(1)} | X_i^{(1)} \sim \text{Bernoulli}(m(X_i^{(1)}))\). The MLE of \(m\) under monotonicity constraints, say \(\hat{m}_1\), can be computed via an isotonic least squares problem that arises from the joint (Bernoulli) likelihood. As this is a Bernoulli monotone response model (see Banerjee 2007, p. 933), the point wise distribution of \(\hat{m}_1\) is described by the Chernoff random variable \(Z\), and letting \(d_{1,1} := \hat{m}_1^{-1}(d_0)\), it can be shown that \(n_1^{1/3} (d_{1,1} - d_0)\) converges to a multiple of \(Z\), as in this article. If the Stage-2 sample is collected as in this article, by restricting the covariate domain to \([d_{1,1} - K n_1^{-\gamma_1}, d_{1,1} + K n_1^{-\gamma_1}]\), and an updated estimate \(d_{2,1}\) constructed from the second-stage data, then, similar to what we have observed, \(n_2^{1/3} (d_{2,1} - d_0)\) converges (at an accelerated rate) to a multiple of \(Z\) again. The likelihood ratio-based procedures can be extended similarly and the optimal allocation continues to be 25% of the budget at Stage 1.

Finally, we should note that although the developed methodology applies to design settings (where the investigator can select the desired covariate and the corresponding response variable values), it can also prove useful in the context of very large datasets. Suppose that one is interested in estimating a threshold of a monotone function from a very large dataset that cannot be stored in its entirety in computer memory. In that case, one-stage estimation based on the entire dataset is computationally challenging, since it requires appropriate modification of the standard algorithms. However, by adopting the proposed adaptive design framework, one can overcome such computational difficulties, while still obtaining a high degree of accuracy. It is our belief that by going to multiple stages, if necessary, with judiciously chosen parameters, one can match the performance of the estimator based on all data that could be stored in computer memory, thus providing a computationally efficient procedure that avoids major modifications of existing algorithms. The latter claim is supported by the results of the experiment shown in Figure 8, and in Figure 4 of the online supplementary material, which indicates the PTSIRP-LR reduces computing time by substantial amounts at larger budgets compared to its one-stage counterpart.

APPENDIX A: PROOFS

We discuss Propositions 1 and 2 of the article.

A.1 Appendix of TSIRP

We introduce the following idealized two-stage isotonic regression procedure (ITSIRP) as follows:

1. Set the first-stage sample proportion \(p \in (0, 1)\) and let the first- and second-stage sample sizes be \(n_1 = \lceil np \rceil\) and \(n_2 = n - n_1\), respectively, where \(n\) is the total sample size.
2. Let the ideal second-stage sampling interval be \([L_i, U_i] = [d_0 \pm C_i n_i^{-1/3}]\) with \(C_i > 0\) and \(n_i > 0\).

3. Allocate the second-stage design points \([X_{2,i}]_{i=1}^{n_2}\) according to a Lebesgue density \(g_2\) on \([L_i, U_i]\), given by \(g_2(x) = (C n_i^{-1/3})^{-1} \phi((x - d_0)/C n_i^{-1/3})\). Denote the corresponding iid second-stage responses \([Y_{2,i}]_{i=1}^{n_2}\).

4. Compute the unconstrained isotonic regression \(m_{2,i}(\theta)\) and the constrained one \(m_{2,i}(\theta_0)\) under the null hypothesis \(m^{-1}(\theta_0) = d_0\) of \(m_{2,i}(\theta)\) over \([L_i, U_i]\) from the second-stage data.

5. Obtain \(d_{2,i} = m_{2,i}(\theta_0)\) and \(2 \log \lambda_{2,i} = 2 \log \lambda_{2,i}(d_0) = 2 \left[ \lambda_i(m_{2,i}(\theta), \delta) - \lambda_i(m_{2,i}(\delta), \delta) \right]\), the ideal second-stage isotonic regression based estimator of \(d_0\) and likelihood ratio statistic under \(H_0\): \(m^{-1}(\theta_0) = d_0\); here, as before, \(\delta\) is a consistent estimate of \(\sigma\).

Remark. Note that ITSIRP is similar to TSIRP except that in ITSIRP the second-stage sampling interval is centered at \(d_0\) instead of at \(d_{1,i}\); therefore, the sampling density at Stage 2 in ITSIRP is \(\psi^{m,n}\) renormalized to \([L_i, U_i]\), just as \(g_2\) is \(\psi\) renormalized to \([L_1, U_1]\) (see Proposition 1) in TSIRP. Since \(d_{1,i}\) converges to \(d_0\) at rate \(n_i^{1/3}\), which is faster than the rate at which \([L_i, U_i]\) is decreasing around \(d_{1,i}\) (since \(\gamma < 1/3\)), \([L_i, U_i]\) is essentially indistinguishable from its idealized counterpart \([L_1, U_1]\) and the asymptotic behavior of \(d_{2,i}\) will be identical to that of \(d_{2,i}\). Similarly, the asymptotic distribution of the idealized LRT, \(2 \log \lambda_{2,i}\), will be the same as that of \(2 \log \lambda_{2,i}\).

A rigorous proof of Proposition 1, formalizing the intuition above, relies on some recently developed theory for M-estimation via multistage sampling in Mallik (2013), ch. 6, where a number of two-stage estimation procedures (and, in particular, the two-stage isotonic regression problem) are investigated, and we direct the technically oriented reader to Section 6.3 of this work. Proposition 2 can be derived along similar lines. In this article, we provide a sketch of the derivations of the limiting distributions of the “idealized” (surrogate) quantities \(d_{2,i}\) and \(2 \log \lambda_{2,i}\). We first introduce the following quantities.

For positive constants \(a_1, a_2\), define the process \(X_{a_1, a_2}(t) \equiv a_1 W(t) + a_2 t^2\) where \(W(t)\) is two-sided Brownian motion on \(\mathbb{R}\), starting from 0. For a function \(f\) defined on \(\mathbb{R}\), let \(\text{scl}(f, I)\) denote the left slope of the greatest convex minorant of the restriction of \(f\) to the interval \(I\). Define \(g_{a_1, a_2}(t) = \text{scl}(X_{a_1, a_2}(t), t)\) and \(g_{a_1, a_2}(t) = \text{scl}(X_{a_1, a_2}(-\infty, 0])\). Define \(\Phi = \int g_{a_1, a_2}(t)^2 dt\) and recall that \(Z \equiv \argmin X_{a_1, a_2}(t)\) is the Chernoff random variable.

**Theorem A.1.** Under Assumption A, we have

\[
n^{(1+\gamma)/3}\{d_{2,i} - d_0\} \overset{d}{\rightarrow} C_{d_0} Z,
\]

where \(C_{d_0} = \left(\frac{C_i}{1-p/p^n\psi(0)}\right)^{1/3}\).

**Theorem A.2.** Under Assumption A and the null hypothesis \(H_0 : m^{-1}(\theta_0) = d_0\), \(2 \log \lambda_{2,i} \overset{d}{\rightarrow} D\).

**Proof-sketch of Theorem 3:** The following equivalence called the “switching relationship,” derived originally by Groeneboom, can be established by arguments similar to those leading to the last display on p. 298 of van der Vaart and Wellner (1996).

**Lemma A.1.** For \(r \in [d_0 \pm C_i n_i^{-1/3}]\) and \(x \in \mathbb{R}\),

\[
m_{2,i}(t) \leq u \iff \min_{x \in [d_0 \pm C_i n_i^{-1/3}]} \{ V_{2,i}(x) - s G_{2,i}(x) \} \geq T(t),
\]

where

\[
V_{2,i}(x) = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_{2,i} I(X_{2,i} \leq x),
\]

\[
G_{2,i}(x) = \frac{1}{n_2} \sum_{i=1}^{n_2} I(X_{2,i} \leq x);
\]

\(T(t)\) is the largest \(X_{2,i}\) less than or equal to \(t\).

As an easy consequence of the above, we have:

\[
d_{2,i} = m_{2,i}^{(1+\gamma)/3}(\theta_0) \approx \min \{ V_{2,i}(x) - \theta_0 G_{2,i}(x) \} + o_p(1/n) \equiv U_n + o_p(1/n).
\]

As \(n^{(1+\gamma)/3} = o(n)\), it follows that \(n^{(1+\gamma)/3}(d_{2,i} - d_0)\) is asymptotically equivalent to \(\hat{h}_n := n^{(1+\gamma)/3}(U_n - d_0)\). Next,

\[
\hat{h}_n \approx \arg \min_{x \in \mathbb{R}} \{ V_{2,i}(x) + n^{(1+\gamma)/3}(U_n - d_0) \},
\]

and

\[
d = (p/(1-p))^{1/3} \left( m'(d_0)/2 \right) \psi(0)/C_1.
\]

Since the process \(X_{c,d}(t)\) is almost surely continuous, diverges to \(\infty\) as \(h \rightarrow \pm \infty\) and has a unique minimum, and \(\hat{h}_n\) is \(O_p(1)\), as can be shown by an application of Theorem 3.2.5 of (van der Vaart and Wellner 1996), we conclude by Theorem 2.7 of Kim and Pollard (1990) that \(\hat{h}_n \approx \min X_{c,d}(t)\).

Now, from Problem 5 on p. 308 of van der Vaart and Wellner (1996), we have \(\min X_{c,d}(t) = (c/d)^{2/3} \min X_{1,i}(t)\), whence

\[
P \left( n^{(1+\gamma)/3}(d_{2,i} - d_0) \leq x \right) \rightarrow P \left( (c/d)^{2/3} \in \arg \min X_{1,i}(t) \leq x \right).
\]

Since

\[
(c/d)^{2/3} = \left( \frac{4a^2}{m'(d_0)^2} \cdot \frac{(1-p)^{1/3} C_1}{p^{1/3} \psi(0)} \right)^{1/3},
\]

we have

\[
n^{(1+\gamma)/3}(d_{2,i} - d_0) \overset{d}{\rightarrow} C_d \left( \frac{(1-p)^{1/3} C_1}{p^{1/3} \psi(0)} \right)^{1/3},
\]

which leads to \(n^{(1+\gamma)/3}(d_{2,i} - d_0) \overset{d}{\rightarrow} C_d Z\), the result in Theorem A.1.

**Proof-sketch of Theorem A.2:** For simplicity, we assume the second-stage sampling density is uniform on \([L_i, U_i]\). That is, \(g_2(x) = (2C_i n_i^{-1/3})^{-1}\) for \(x \in [L_i, U_i]\). Then, via a derivation similar to that on pp. 26–28 of Banerjee (2006) we deduce

\[
\left( n^{(1+\gamma)/3}(m_{2,i}(d_0 + x^n_{2,i}^{(1+\gamma)/3}) - m(d_0)) \right) \overset{d}{\rightarrow} \left( g_2^{(p)} d \right).
\]
where now, \( c = (2C \sigma^2(1-p)/p)^{1/2} \) and \( d = m'(d_0)/2 \). In fact, the weak convergence (A.4) holds not only finite dimensionally, but also in the normed linear space \( L_2([-K, K] \times L_2([-K, K])) \) for every \( K > 0 \), because of the monotonicity of both \( m_{ul} \) and \( m_{ol} \).

To derive the asymptotics for \( 2 \log \lambda_{ol} \), it suffices (by Slutsky’s theorem) to consider a tweaked version of this quantity with the \( \sigma^2 \) in the denominator replaced by the true \( \sigma^2 \). In what follows, we work with this version and continue to call it \( 2 \log \lambda_{ol} \). We have:

\[
2 \log \lambda_{ol} = 2 \left[ \frac{1}{2\sigma^2} \sum_{i=1}^{n_2} (Y_{2,i} - m_{ol}(X_{2,i}))^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n_2} (Y_{2,i} - m_{ul}(X_{2,i}))^2 \right] = \frac{1}{\sigma^2} \sum_{i=1}^{n_2} \left[ (Y_{2,i} - \theta_0) - (m_{ul}(X_{2,i}) - \theta_0) \right]^2 - \frac{1}{\sigma^2} \sum_{i=1}^{n_2} \left[ (Y_{2,i} - \theta_0) - (m_{ol}(X_{2,i}) - \theta_0) \right]^2 = \frac{2}{\sigma^2} \sum_{i=1}^{n_2} (Y_{2,i} - \theta_0)(m_{ol}(X_{2,i}) - \theta_0) + \frac{1}{\sigma^2} \sum_{i=1}^{n_2} \left[ (m_{ul}(X_{2,i}) - \theta_0)^2 - (m_{ol}(X_{2,i}) - \theta_0)^2 \right] = \frac{2}{\sigma^2} \sum_{i=1}^{n_2} (Y_{2,i} - m_{ul}(X_{2,i}))(m_{ol}(X_{2,i}) - \theta_0) + \frac{2}{\sigma^2} \sum_{i=1}^{n_2} (Y_{2,i} - m_{ol}(X_{2,i}))(m_{ol}(X_{2,i}) - \theta_0) + \frac{2}{\sigma^2} \sum_{i=1}^{n_2} (m_{ol}(X_{2,i}) - \theta_0)^2 + \frac{1}{\sigma^2} \sum_{i=1}^{n_2} \left[ (m_{ul}(X_{2,i}) - \theta_0)^2 - (m_{ol}(X_{2,i}) - \theta_0)^2 \right]
\]

where the last equation is a consequence of the fact that isotonic regression estimators are formed by averaging the responses over blocks of order statistics of the covariates, which ensures that

\[
\sum_{i=1}^{n_2} (Y_{2,i} - m_{ul}(X_{2,i}))(m_{ol}(X_{2,i}) - \theta_0)
\]

and

\[
\sum_{i=1}^{n_2} (Y_{2,i} - m_{ol}(X_{2,i}))(m_{ol}(X_{2,i}) - \theta_0)
\]

are both equal to 0.

Now denote \( \mathbb{P}_{n_2} \) as the empirical measure of the second-stage covariates \( \{X_{2,i}\}_{i=1}^{n_2} \) and \( \mathbb{P}_{n_2} \) as the corresponding uniform probability measure of \( X_{2,i} \). Let \( D_{n_2} \) denote the interval on which \( m_{ul} \) and \( m_{ol} \) differ. Then, we have

\[
2 \log \lambda_{ol} = \frac{n_2}{\sigma^2} \mathbb{P}_{n_2} \left[ \left( m_{ol}(x) - \theta_0 \right)^2 - \left( m_{ul}(x) - \theta_0 \right)^2 \right] \times \{ x \in D_{n_2} \} = T_1 + T_2,
\]

where

\[
T_1 = \frac{n_2}{\sigma^2} \left( \mathbb{P}_{n_2} - \mathbb{P}_{n_2} \right) \left[ \left( m_{ol}(x) - \theta_0 \right)^2 - \left( m_{ul}(x) - \theta_0 \right)^2 \right] \times \{ x \in D_{n_2} \},
\]

\[
T_2 = \frac{n_2}{\sigma^2} \mathbb{P}_{n_2} \left[ \left( m_{ol}(x) - \theta_0 \right)^2 - \left( m_{ul}(x) - \theta_0 \right)^2 \right] \times \{ x \in D_{n_2} \}.
\]

Since \((1 - 2\gamma)/3 < 1/2\), both \( m_{ul}(x) - \theta_0 \) and \( m_{ol}(x) - \theta_0 \) are \( O_P(n_2^{-1/6}) \) and

\[
T_1 = \frac{n_2}{\sigma^2} \mathbb{P}_{n_2} \left( \mathbb{P}_{n_2} - \mathbb{P}_{n_2} \right) \left[ \left( m_{ol}(x) - \theta_0 \right)^2 - \left( m_{ul}(x) - \theta_0 \right)^2 \right] \times \{ x \in D_{n_2} \},
\]

we can show that \( T_1 \) converges to 0 in probability by empirical process theory arguments.

Next, \( T_2 \) is given by

\[
\frac{n_2}{\sigma^2} \int_{D_{n_2}} \left[ \left( m_{ol}(x) - \theta_0 \right)^2 - \left( m_{ul}(x) - \theta_0 \right)^2 \right] \times \{ x \in D_{n_2} \}.
\]

The equality preceding the weak convergence above follows from the change of variable \( x = d_0 + t \gamma \), the weak convergence of the likelihood ratio statistic follows from the weak convergence result (A.4) in the \( L_2 \) sense and the fact that the set \( n_2^{1/6} (D_{n_2} - d_0) \) is contained in a compact set with (arbitrarily) high probability, eventually. For the very last equality, see, for example, the proof of Theorem 2.2 of Banerjee (2007). Thus, Theorem A.2 holds.

ACKNOWLEDGMENTS

The work of MB was supported by NSF grant DMS-1007751, NSA grant H98230-11-1-0166, and a Sokol Faculty Award, University of Michigan. The work of GM was supported by NSF grants DMS-1161838 and DMS-1228164.

[Received March 2013. Revised June 2014.]

REFERENCES

Banerjee, M. (2000), “Likelihood Ratio Inference in Regular and Non-Regular Problems,” PhD thesis, University of Washington. [397]

(2006), “Competing Statistics in Exponential Family Regression Models Under Certain Shape Constraints,” Technical Report, Department of Statistics, University of Michigan, available at: http://dept.stat.lsa.umich.edu/~moub/monomodal-expfamily.pdf. [405]

(2007), “Likelihood Based Inference for Monotone Response Models,” The Annals of Statistics, 35, 931–956. [404,406]

(2009), “Inference in Exponential Family Regression Models Under Certain Shape Constraints,” in Advances in Multivariate Statistical Meth-
ods, Statistical Science and Interdisciplinary Research, Singapore: World Scientific, pp. 249–272. [397]
Banerjee, M., and Wellner, J. (2001), “Likelihood Ratio Tests for Monotone Functions,” The Annals of Statistics, 29, 1699–1731. [397]
Bhattacharya, R., and Lin, L. (2010), “An Adaptive Nonparametric Method in Benchmark Analysis for Bioassay and Environmental Studies,” Statistics & Probability Letters, 80, 1947–1953. [396]
Dette, H., Neumeyer, N., and Pliz, K. (2005), “A Note on Nonparametric Estimation of the Effective Dose in Quantal Bioassay,” The Journal of the American Statistical Association, 100, 503–510. [396]
Fan, J., and Gijbels, I. (1996), Local Polynomial Modelling and Its Applications (Vol.66), Monographs on Statistics and Applied Probability, London: Chapman & Hall. [399,401]
Gasser, T., Sroka, L., and Jennen-Steinmetz, C. (1986), “Residual Variance and Residual Pattern in Nonlinear Regression,” Biometrika, 73, 625–633. [399]
Groeneboom, P., and Wellner, J. A. (2001), “Computing Chernoff’s Distribution,” Journal of Computational and Graphical Statistics, 10, 388–400. [397]
Kim, J., and Pollard, D. (1990), “Cube Root Asymptotics,” The Annals of Statistics, 18, 191–219. [405]
Lan, Y., Banerjee, M., and Michailidis, G. (2009), “Change-Point Estimation Under Adaptive Sampling,” The Annals of Statistics, 37, 1752–1791. [395,402]
Mallik, A. (2013), “Topics on Threshold Estimation, Multistage Methods and Random Fields,” PhD thesis, University of Michigan. [405]
Osborne, C. (1991), “Statistical Calibration: A Review,” International Statistical Review, 59, 309–336. [396]
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, Chichester: Wiley. [397]
Rosenberger, W. F., and Haines, L. M. (2002), “Competing Designs for Phase I Clinical Trials: A Review,” Statistics in Medicine, 21, 2757–2770. [396]
Tang, R. (2011), “Adaptive and Multistage Procedures for Inference on Monotone Regression Functions in Observed Data Studies and Design Settings,” PhD thesis, University of Michigan. [404]
Tang, R., Banerjee, M., and Michailidis, G. (2011), “A Two-Stage Hybrid Procedure for Estimating an Inverse Regression Function,” The Annals of Statistics, 39, 956–989. [395,396,397,399,401]
van der Vaart, A. W., and Wellner, J. A. (1996), Weak Convergence and Empirical Processes, Heidelberg: Springer. [405]