Fractional order statistic approximation for
nonparametric conditional quantile inference

Matt Goldman*    David M. Kaplan*

September 28, 2015; first version December 15, 2011

Abstract

Using and extending fractional order statistic theory, we characterize the $O(n^{-1})$ coverage probability error of the previously proposed confidence intervals for population quantiles using $L$-statistics as endpoints in Hutson (1999). We derive an analytic expression for the $n^{-1}$ term, which may be used to calibrate the nominal coverage level to get $O(n^{-3/2} \log(n))$ coverage error. Asymptotic power is shown to be optimal. Using kernel smoothing, we propose a related method for nonparametric inference on conditional quantiles. This new method compares favorably with asymptotic normality and bootstrap methods in theory and in simulations. Code is provided for both unconditional and conditional inference.

JEL classification: C21

Keywords: fractional order statistics, high-order accuracy, inference-optimal bandwidth, kernel smoothing.

1 Introduction

Quantiles contain information about a distribution’s shape. Complementing the mean, they capture heterogeneity, inequality, and other measures of economic interest. Nonparametric conditional quantile models further allow arbitrary heterogeneity across regressor values. This paper concerns nonparametric inference on quantiles and conditional quantiles. In

*Goldman: Microsoft, mattgold@microsoft.com. Kaplan (corresponding author): Department of Economics, University of Missouri, kaplandm@missouri.edu. We thank the co-editor (Oliver Linton), associate editor, referees, and Yixiao Sun for helpful comments and references, and Patrik Guggenberger and Andres Santos for feedback improving the clarity of presentation. Thanks also to Brendan Beare, Karen Messer, and active audience members at seminars and conferences. Thanks to Ruixuan Liu for providing and discussing code from [Fan and Liu (2015)]. This paper was previously circulated as parts of "IDEAL quantile inference via interpolated duals of exact analytic $L$-statistics" and "IDEAL inference on conditional quantiles."
particular, we characterize the high-order accuracy of both Hutson's (1999) $L$-statistic-based confidence intervals (CIs) and our new conditional quantile CIs based thereon.

Conditional quantiles appear across diverse topics because they are fundamental statistical objects. Studies of wages have looked at experience (Hogg, 1975), union membership (Chamberlain, 1994), and inequality (Buchinsky, 1994). Other examples include models of infant birthweight (Abrevaya, 2001), demand for alcohol (Manning et al., 1995), school quality and student outcomes (Eide and Showalter, 1998), and Engel curves (Alan et al., 2005; Deaton, 1997, pp. 81–82), which we examine in our empirical application.

Previously, Hutson (1999) proposed CIs for quantiles based on fractional order statistics, using $L$-statistics (linearly interpolating two order statistics) as endpoints. Although they worked well and were used in practice, the CIs lacked formal proofs of accuracy and precision. We provide these by deriving coverage probability error (CPE) and asymptotic power against local alternatives, which is shown to be optimal.

The theoretical results we develop en route contribute to the fractional order statistic literature and provide the basis for inference on other objects of interest explored in Goldman and Kaplan (2014) and Kaplan (2014). In particular, Theorem 2 tightly links the distributions of $L$-statistics from the observed and ‘ideal’ (unobserved) fractional order statistic processes, also showing a connection with the Brownian bridge process. Additionally, Lemma 7 provides precise PDF and PDF derivative approximations of the Dirichlet distribution (that governs ideal fractional order statistics) in terms of the multivariate normal.

We derive an analytic leading $n^{-1}$ term in the CPE for Hutson (1999) and provide a new calibration to achieve $O(n^{-3/2} \log(n))$ CPE. This is analogous to the Ho and Lee (2005a) analytic calibration of the $L$-statistic CIs in Beran and Hall (1993), which similarly improves CPE from $O(n^{-1})$ to $O(n^{-3/2})$.

High-order accuracy is most valuable in small samples. These may occur in experiments or when the population itself is small (e.g., the 50 states in the U.S.). Perhaps most importantly, even with moderate or large samples, nonparametric analysis can entail small local sample
sizes. For example, if \( n = 1024 \) and there are five binary regressors, then the smallest local sample size cannot exceed \( 1024 / 2^5 = 32 \).

Our quantile CIs are closer to equal-tailed in finite samples than most. Symmetric and equal-tailed CIs are first-order equivalent due to sample quantiles’ asymptotic normality, but the higher-order differences can be important in finite samples, especially away from the median. Viewing the CI as an interval estimator, this equal-tailed property is equivalent to median unbiasedness, the virtues of which are discussed more in Section 3. Our theoretical framework is not specific to equal-tailed CIs; modifications to achieve alternative optimal properties are left to future work.

For nonparametric conditional quantile inference, we introduce an \( L \)-statistic-based kernel method that smooths over continuous conditioning variables and also allows discrete conditioning variables. Achieving CPE-optimal inference balances the aforementioned analytic CPE of our unconditional method and CPE from bias due to smoothing over continuous regressors. We derive the optimal CPE and bandwidth rates, as well as a plug-in bandwidth when the conditioning vector contains a single continuous component.

Our \( L \)-statistic method has theoretical and computational advantages over methods based on normality or an unsmoothed bootstrap. The theoretical bottleneck for our approach is the need to use a uniform kernel to apply our exact unconditional results. Nonetheless, when the conditioning vector has one or two continuous components, our CPE is of smaller order than that of normality or bootstrap methods, even if they assume infinite smoothness of the conditional quantile function while we only assume two derivatives; see Section 4.3 for details. Our method also computes more quickly than existing methods (of reasonable accuracy), handling even more challenging tasks in 10–15 seconds instead of minutes; see Table 8.

Recent complementary work of Fan and Liu (2015) also concerns a “direct method” of nonparametric inference on conditional quantiles. They use a limiting Gaussian process to derive first-order accuracy in a general setting, whereas we use the finite-sample Dirichlet
process to achieve high-order accuracy in an iid setting. Fan and Liu (2015) also provide uniform (over $X$) confidence bands. We suggest a confidence band from interpolating a growing number of joint CIs, although it will take additional work to rigorously justify. A different, more ad hoc confidence band described in Section 6 seems to perform well in practice.

If applied to a local constant estimator with a uniform kernel and the same bandwidth, the Fan and Liu (2015) approach is less accurate than ours due to the normal (instead of beta) reference distribution and integer (instead of interpolated) order statistics in equation (6). However, with other estimators like local polynomials, the Fan and Liu (2015) method is not necessarily less accurate. We compare further in our simulations. One open question is whether using our beta reference and interpolation can improve accuracy for the general Fan and Liu (2015) method beyond the local constant estimator with a uniform kernel; our Lemma 3 shows this will at least retain first-order accuracy.

The order statistic approach to quantile inference uses the idea of the probability integral transform, which dates back to R. A. Fisher (1932), Karl Pearson (1933), and Neyman (1937). For continuous $X_i \overset{iid}{\sim} F(\cdot)$, $F(X_i) \overset{iid}{\sim} \text{Unif}(0, 1)$. Each order statistic from such an iid uniform sample has a known beta distribution for any sample size $n$. Order statistics may be regarded as sample $u$-quantiles with index $u \in (0, 1)$ such that $(n + 1)u$ is an integer. However, for a given $n$, $(n + 1)u$ is fractional for almost all $u \in (0, 1)$.

Though unobserved for non-integer indices, ‘ideal’ fractional order statistics\footnote{Technically, these are not statistics since they are not functions of the observed sample, but rather theoretical constructs. Nonetheless, we follow the literature’s naming convention.} for an iid uniform sample jointly follow a particular Dirichlet process (Stigler, 1977), the marginal of which is a beta distribution. We show that the $L$-statistic linearly interpolating consecutive order statistics is well approximated by this beta distribution, with only $O(n^{-1})$ error in CDF. For example, the sampling distribution of the average of the 4th and 5th order statistics is well approximated by the beta distribution of the 4.5th ideal fractional order statistic. This $O(n^{-1})$ bound on interpolation error is the key to formally justifying the $L$-statistic methods.
examined here and in other papers. Although $O(n^{-1})$ is an asymptotic claim, the interpolated fractional order statistic cannot lie far from its unobserved counterpart even in small samples, which is an advantage over methods more sensitive to asymptotic approximation error.

Many other approaches to one-sample quantile inference have been explored. With Edgeworth expansions, Hall and Sheather (1988) and Kaplan (2015) obtain two-sided $O(n^{-2/3})$ CPE. With bootstrap, smoothing is necessary for high-order accuracy. This increases the computational burden and requires good bandwidth selection in practice. See Ho and Lee (2005b), §1 for a review of bootstrap methods. Bartlett-corrected smoothed empirical likelihood (Chen and Hall, 1993) also achieves nice theoretical properties, but with the same caveats.

Other order statistic-based CIs dating back to Thompson (1936) are surveyed in David and Nagaraja (2003, §7.1). Most closely related to Hutson (1999) is Beran and Hall (1993). Like Hutson (1999), Beran and Hall (1993) linearly interpolate order statistics for CI endpoints, but with an interpolation weight based on the binomial distribution. Although their proofs use expansions of the Rényi (1953) representation instead of fractional order statistic theory, their $n^{-1}$ CPE term ($d_1$ in their appendix, or $d_2$ for two-sided CIs) is identical to that for Hutson (1999) other than the different weight value. Prior work (e.g., Bickel, 1967; Shorack, 1972) has established asymptotic normality of $L$-statistics and convergence of the sample quantile process to a Gaussian limit process, but the methods of Beran and Hall (1993) and Hutson (1999) achieve higher-order accuracy.

The most apparent difference between the two-sided CIs of Beran and Hall (1993) and Hutson (1999) is that the former are “symmetric” in terms of order statistic index (but not length), whereas the latter are equal-tailed. This also allows Hutson (1999) to be computed further into the tails; e.g., when $n = 99$, Hutson (1999) can compute a two-sided 95% CI for quantiles in the range [0.036, 0.05], whereas Beran and Hall (1993) cannot. Additionally,
Goldman and Kaplan (2014) extend our fractional order statistic framework to justify high-order accurate confidence intervals for two-sample quantile differences, interquantile ranges, and other objects. Such extensions may be possible through the Rényi representation framework, but we are unaware of such results, and the resulting methods would not have the simple interpretation of using a Dirichlet reference distribution for the linearly interpolated order statistics.

For nonparametric conditional quantile inference, in addition to the aforementioned Fan and Liu (2015) approach, Chaudhuri (1991) derives the pointwise asymptotic normal distribution of a local polynomial estimator. Qu and Yoon (2015) propose modified local linear estimators of the conditional quantile process that converge weakly to a Gaussian process, and they suggest using a conservative type of bias correction (that only enlarges a CI) to deal with the asymptotic bias present when using a bandwidth of the MSE-optimal rate. Alternatively, Fan and Liu (2015) may be applied to the Qu and Yoon (2015) estimator, addressing bias either through undersmoothing (as in the former paper’s simulations) or direct bias correction (as in the latter). For comparison with the uniform (over quantiles) results in Qu and Yoon (2015), the extension of our probability integral transform approach to uniform confidence bands for the (unconditional) quantile process is in Goldman and Kaplan (2015).

Section 2 contains our theoretical results on fractional order statistic approximation, which are applied to unconditional quantile inference in Section 3. Section 4 concerns our new conditional quantile inference method. An empirical application and simulation results are in Sections 5 and 6 respectively. Proofs absent from the text are collected in Appendix A, and implementation notes are in Appendix B, further details of both are in the supplemental appendix, along with additional simulations.

Notationally, \( \phi(\cdot) \) is the standard normal PDF, \( \Phi(\cdot) \) is the standard normal CDF, \( \approx \) should be read as “is equal to, up to smaller-order terms”; \( \asymp \) as “has exact (asymptotic) rate/order of” (same as “big theta” Bachmann–Landau notation, \( \Theta(\cdot) \)); and \( A_n = O(B_n) \) as usual, \( \exists k < \infty \text{ s.t. } |A_n| \leq B_nk \) for sufficiently large \( n \). Acronyms used are those for
cumulative distribution function (CDF), confidence interval (CI), coverage probability (CP), coverage probability error (CPE), and probability density function (PDF).

2 Fractional order statistic theory

In this section, we introduce notation and present our core theoretical results linking unobserved ‘ideal’ fractional $L$-statistics with their observed counterparts.

Given an iid sample $\{X_i\}_{i=1}^n$ of draws from a continuous CDF denoted $F(\cdot)$, interest is in $Q(p) \equiv F^{-1}(p)$ for some $p \in (0, 1)$, where $Q(\cdot)$ is the quantile function. Let $u \in (0, 1)$,

$$k \equiv \lfloor u(n + 1) \rfloor, \quad \epsilon \equiv u(n + 1) - k,$$

where $\lfloor \cdot \rfloor$ is the floor function and $\epsilon$ is the interpolation weight. The sample $L$-statistic commonly associated with $Q(u)$ is

$$\hat{Q}_X^L(u) \equiv (1 - \epsilon)X_{n:k} + \epsilon X_{n:k+1},$$

(1)

where $X_{n:k}$ denotes the $k$th order statistic (i.e., $k$th smallest sample value). While $Q(u)$ is latent and nonrandom, $\hat{Q}_X^L(u)$ is a random variable, and $\hat{Q}_X^L(\cdot)$ is a stochastic process, both of which are observed for arguments varying over $[1/(n + 1), n/(n + 1)]$.

Let $\Xi_n \equiv \{k/(n + 1)\}_{k=1}^n$ denote the set of quantiles corresponding to the observed order statistics. If $u \in \Xi_n$, then no interpolation is necessary and $\hat{Q}_X^L(u) = X_{n:k}$. As detailed in Section 3, application of the probability integral transform yields exact coverage probability of a CI endpoint $X_{n:k}$ for $Q(p)$: $P(X_{n:k} < F^{-1}(p)) = P(U_{n:k} < p)$, where $U_{n:k} \equiv F(X_{n:k}) \sim \beta(k, n + 1 - k)$ is equal in distribution to the $k$th order statistic from $U_i \overset{iid}{\sim} \text{Unif}(0, 1)$, $i = 1, \ldots, n$ (Wilks 1962 8.7.4). However, we also care about $u \notin \Xi_n$, in which case $k$ is fractional. To better handle such fractional order statistics, we will present a tight link between the marginal distributions of the stochastic process $\hat{Q}_X^L(\cdot)$ and those of the analogous

---

Footnote: $^3$ $F$ will often be used with a random variable subscript to denote the CDF of that particular random variable. If no subscript is present, then $F(\cdot)$ refers to the CDF of $X$. Similarly for the PDF $f(\cdot)$.
'ideal' (I) process

\[ \hat{Q}_X^I(\cdot) \equiv F^{-1}\left(\hat{Q}_F^I(\cdot)\right), \]  

(2)

where \( \hat{Q}_F^I(\cdot) \) is the ideal (I) uniform (U) fractional order statistic process. We use a tilde in \( \hat{Q}_X^I(\cdot) \) and \( \hat{Q}_F^I(\cdot) \) instead of the hat like in \( \hat{Q}_X^L(\cdot) \) to emphasize that the former are unobserved, whereas the latter is computable from the sample data.

This \( \hat{Q}_F^I(\cdot) \) in (2) is a Dirichlet process \( \text{[Ferguson, 1973; Stigler, 1977]} \) on the unit interval with index measure \( \nu([0, t]) = (n + 1)t \). Its univariate marginals are

\[ \hat{Q}_F^I(u) = U_{n:(n+1)u} \sim \beta((n + 1)u,(n + 1)(1 - u)). \]  

(3)

The marginal distribution of \( \left( \hat{Q}_F^I(u_1), \hat{Q}_F^I(u_2) - \hat{Q}_F^I(u_1), \ldots, \hat{Q}_F^I(u_k) - \hat{Q}_F^I(u_{k-1}) \right) \) for \( u_1 < \cdots < u_k \) is Dirichlet with parameters \( (u_1(n + 1), (u_2 - u_1)(n + 1), \ldots, (u_k - u_{k-1})(n + 1)) \).

Note that \( \hat{Q}_X^I(u) \) coincides with \( \hat{Q}_X^L(u) \) for all \( u \in \Xi_n \), differing only in its interpolation between these points. Theorem 1 shows \( \hat{Q}_X^I(\cdot) \) and \( \hat{Q}_X^L(\cdot) \) to be closely linked in probability.

**Theorem 1.** For any fixed \( \delta > 0 \) and \( m > 0 \), define \( U^\delta \equiv \{ u \in (0, 1) \mid \forall t \in (u - m, u + m), f(F^{-1}(t)) \geq \delta \} \) and \( U_\alpha^\delta \equiv U^\delta \cap \left[ \frac{1}{n+1}, \frac{n}{n+1} \right] \); then, \( \sup_{u \in U_\alpha^\delta} |\hat{Q}_X^I(u) - \hat{Q}_X^L(u)| = O_p(n^{-1}[\log n]) \).

Although Theorem 1 motivates approximating the distribution of \( \hat{Q}_X^L(u) \) by that of \( \hat{Q}_X^I(u) \), it is not relevant to high-order accuracy. In fact, its result is achieved by any interpolation between \( X_{n:k} \) and \( X_{n:k+1} \), not just \( \hat{Q}_X^L(u) \); in contrast, the high-order accuracy we establish in Theorem 4 is only possible with precise interpolations like \( \hat{Q}_X^L(u) \).

Next, we consider the multivariate distribution of a vector of quantiles with fixed dimension \( J < \infty \), where \( u_j \) denotes an element in column vector \( \mathbf{u} \in (0, 1)^J \). The joint PDF of \( \{\hat{Q}_F^I(u_j)\}_{j=1}^J \) evaluated at \( \{x_j\}_{j=1}^J \) is Dirichlet with parameters \( (n + 1)(u_j - u_{j-1}) \):

\[ \frac{\Gamma((n + 1)(u_j - u_{j-1}))}{\prod_{j=1}^{J+1} \Gamma((n + 1)(u_j - u_{j-1}))} \left( \prod_{j=1}^{J+1} (x_j - x_{j-1})^{(n + 1)(u_j - u_{j-1}) - 1} \right), \]

where \( u_0 \equiv 0, x_0 \equiv 0, u_{J+1} \equiv 1, \) and \( x_{J+1} \equiv 1 \). Note that the Dirichlet more directly
describes the distribution of spacings \( \tilde{Q}_U^I(u_j) - \tilde{Q}_U^I(u_{j-1}) \), rather than values \( \tilde{Q}_U^I(u_j) \).

We also consider the Gaussian approximation to the sampling distribution of order statistics. It is well known that the centered and scaled empirical process for standard uniform random variables converges to a Brownian bridge. For a Brownian bridge process \( B(\cdot) \), we define on \( u \in (0, 1) \) the additional stochastic processes

\[
\tilde{Q}_B^U(u) \equiv u + n^{-1/2} B(u) \quad \text{and} \quad \tilde{Q}_X^B(u) \equiv F^{-1}(\tilde{Q}_U^B(u)).
\]

Like the realizations of \( \tilde{Q}_X^I(\cdot) \) and \( \tilde{Q}_U^I(\cdot) \), the realizations of these processes do not correspond to observed data, but the distribution of \( \tilde{Q}_X^B(u) \) provides a convenient approximation to the distributions from the other processes. The vector \( \tilde{Q}_U^I(u) \) has an ordered Dirichlet distribution (i.e., the spacings between consecutive \( \tilde{Q}_U^I(u_j) \) follow a joint Dirichlet distribution), while \( \tilde{Q}_U^B(u) \) is multivariate Gaussian.

Lemma 7 in the appendix proves the close relationship between multivariate Dirichlet and Gaussian PDFs and PDF derivatives for all values outside the tails (parts i and ii), which are shown to have rapidly diminishing probability (parts iii and iv). To the best of our knowledge, these are new results, and they facilitate the primary theoretical contributions of our paper.

We use Lemma 7 to prove the close distributional link among linear combinations of ideal, interpolated, and Gaussian-approximated fractional order statistics in Theorem 2. Specifically, for arbitrary weight vector \( \psi \in \mathbb{R}^J \) with non-zero elements \( \psi_j \), we seek to (distributionally) approximate

\[
L^L \equiv \sum_{j=1}^{J} \psi_j \tilde{Q}_X^L(u_j) \quad \text{by} \quad L^I \equiv \sum_{j=1}^{J} \psi_j \tilde{Q}_X^I(u_j),
\]

or alternatively by

\[
L^B \equiv \sum_{j=1}^{J} \psi_j \tilde{Q}_X^B(u_j).
\]

For convenience and without loss of generality, we normalize \( \psi_1 = 1 \).

Our assumptions for this section are now presented, followed by the main theoretical
result. Assumption $A2$ ensures that the first three derivatives of the quantile function are uniformly bounded in neighborhoods of the quantiles, $u_j$, which helps bound remainder terms in the proofs. We use \textbf{bold} for vectors, \underline{underline} for matrices, and $\phi_{\Sigma}(\cdot)$ for the multivariate normal PDF with mean $0$ and covariance matrix $\Sigma$.

**Assumption A1.** Sampling is iid: $X_i \overset{iid}{\sim} F$, $i = 1, \ldots, n$.

**Assumption A2.** For each quantile $u_j$, the PDF $f(\cdot)$ (corresponding to CDF $F(\cdot)$ in A1) satisfies (i) $f(F^{-1}(u_j)) > 0$; (ii) $f''(\cdot)$ is continuous in some neighborhood of $F^{-1}(u_j)$, i.e., $f \in C^2(U_\delta(F^{-1}(u_j)))$ with $U_\delta(x)$ denoting some $\delta$-neighborhood of point $x \in \mathbb{R}$.

**Theorem 2.** Define $\mathbf{V}$ as the $J \times J$ matrix such that $V_{i,j} = \min\{u_i, u_j\}(1 - \max\{u_i, u_j\})$, and let

$$A \equiv \text{diag}\{f(F^{-1}(u))\}, \quad \mathbf{V}_\psi \equiv \psi'(A^{-1}\mathbf{V}_\psi A^{-1})\psi, \quad \mathbf{X}_0 \equiv \sum_{j=1}^J \psi_j F^{-1}(u_j).$$

As in (1), the $\epsilon_j \equiv (n+1)u_j - \lfloor (n+1)u_j \rfloor$ are interpolation weights. Let Assumption $A1$ hold, and let $A2$ hold at $\bar{u}$. Given the definitions in (1), (2), and (4), the following results hold uniformly over $\mathbf{u} = \bar{u} + o(1)$.

(i) For a given constant $K$,

$$P\left(L^L < \mathbf{X}_0 + n^{-1/2}K\right) - P\left(L^I < \mathbf{X}_0 + n^{-1/2}K\right) = K \exp\left\{-K^2/(2\mathbf{V}_\psi^3)\right\} \left[\sum_{j=1}^J \frac{\psi_j^2 \epsilon_j (1 - \epsilon_j)}{f[F^{-1}(u_j)]^2}\right] n^{-1} + O(n^{-3/2}\log(n)),$$

where the remainder is uniformly $O(n^{-3/2}\log(n))$ over all $K$.

(ii) Uniformly over $K$,

$$\sup_{K \in \mathbb{R}} \left[ P\left(L^L < \mathbf{X}_0 + n^{-1/2}K\right) - P\left(L^I < \mathbf{X}_0 + n^{-1/2}K\right) \right] = e^{-1/2} \left[\sum_{j=1}^J \frac{\psi_j^2 \epsilon_j (1 - \epsilon_j)}{f[F^{-1}(u_j)]^2}\right] n^{-1} + O(n^{-3/2}\log(n)),$$
and
\[
\sup_{K \in \mathbb{R}} \left| P \left( L^B < X_0 + n^{-1/2} K \right) - P \left( L^I < X_0 + n^{-1/2} K \right) \right| = O(n^{-1/2} \log(n)).
\]

The additional result for \( L^B \) in part (ii) follows directly from the Dirichlet PDF approximation in Lemma 8(iii), where the cubic term in the Taylor expansion around the mode is \( O(n^{-1/2} \log(n)) \).

By the Cramér–Wold device, the vector \( \hat{Q}^L_X(u) \) converges in distribution to \( \tilde{Q}^L_X(u) \) up to an \( O(n^{-1}) \) term and to \( \tilde{Q}^B_X(u) \) up to an \( O(n^{-1/2} \log(n)) \) term. This could allow reliable inference on more general finite-dimensional functionals of the quantile process. In the remainder of this paper, attention is restricted to the class of linear combinations as in (4).

As shown in the proof, the CDF error in Theorem 2(ii) is proportional to
\[
\max \left\{ \left[ \min_j \{u_j\} \right]^{-1}, \left[ \min_j \{1 - u_j\} \right]^{-1} \right\}.
\]
This is consistent with the well-known additional difficulties of constructing confidences intervals for “extreme quantiles” instead of “central quantiles.” We may remove this specific term via analytic calibration, though a similar term likely remains.

### 3 Quantile inference: unconditional

For any quantile of interest \( p \) and confidence level \( 1 - \alpha \), we define \( u^h(\alpha) \) and \( u^l(\alpha) \) as solutions to
\[
\alpha = P \left( \hat{Q}^L_U(u^h(\alpha)) < p \right), \quad \alpha = P \left( \hat{Q}^L_U(u^l(\alpha)) > p \right),
\]
with \( \hat{Q}^L_U(u) \sim \beta((n + 1)u, (n + 1)(1 - u)) \) from (3). These parallel (7) and (8) in Hutson (1999).

One-sided CI endpoints are \( \hat{Q}^L_X(u^h) \) or \( \hat{Q}^L_X(u^l) \), i.e., linearly interpolated fractional order

\footnote{A similar approximation error may follow from using the Bahadur remainder in Portnoy (2012).}
statistics with index \((n + 1)u^h\) or \((n + 1)u^l\). Two-sided CIs replace \(\alpha\) with \(\alpha/2\) in (5) and use both endpoints. This use of \(\alpha/2\) yields the equal-tailed property; more generally, for \(t \in (0, 1)\), \(t\alpha\) and \((1-t)\alpha\) can be used. Figure 1 visualizes an example. The beta distribution's mean is \(u^h\) (or \(u^l\)). Decreasing \(u^h\) increases the probability mass in the shaded region below \(u\), while increasing \(u^h\) decreases the shaded region, and vice-versa for \(u^l\). Consequently, solving (5) is a simple numerical search problem.

Figure 1: Example of one-sided CI endpoint determination, \(n = 11\), \(p = 0.65\), \(\alpha = 0.1\). Left: the lower one-sided endpoint \(u^l\) is such that the shaded region’s area is \(P\left(\tilde{Q}^l_U(u^l) > p\right) = \alpha\). Right: similarly, \(u^h\) solves \(P\left(\tilde{Q}^h_U(u^h) < p\right) = \alpha\).

The CI endpoints converge to \(p\) at a root-\(n\) rate and may be approximated by quantiles of a normal distribution, as in Lemma 3.

**Lemma 3.** Let \(z_{1-\alpha}\) denote the \((1 - \alpha)\)-quantile of a standard normal distribution, \(z_{1-\alpha} \equiv \Phi^{-1}(1 - \alpha)\). From the definitions in (5), the values \(u^l(\alpha)\) and \(u^h(\alpha)\) can be approximated as

\[
u^l(\alpha) = p - n^{-1/2}z_{1-\alpha}\sqrt{p(1-p)} - \frac{2p - 1}{6n}(z_{1-\alpha}^2 + 2) + O(n^{-3/2}),
\]

\[
u^h(\alpha) = p + n^{-1/2}z_{1-\alpha}\sqrt{p(1-p)} - \frac{2p - 1}{6n}(z_{1-\alpha}^2 + 2) + O(n^{-3/2}).
\]

We first consider a lower one-sided CI for \(Q(p)\) based on the iid sample \(\{X_i\}_{i=1}^n\). Using
the $1 - \alpha$ CI from Hutson (1999) is
\[
\left( -\infty, \hat{Q}_L^X(u^h(\alpha)) \right).
\] (6)

Coverage probability is
\[
P\left\{ Q(p) \in \left( -\infty, \hat{Q}_L^X(u^h(\alpha)) \right) \right\} = P\left( \hat{Q}_X^L(u^h(\alpha)) > Q(p) \right)
\]
\[
\equiv P\left( \hat{Q}_X^L(u^h(\alpha)) > Q(p) \right) + \frac{\epsilon_h(1 - \epsilon_h)z_{1-\alpha} \exp\left\{ -z_{1-\alpha}^2/2 \right\}}{\sqrt{2\pi}u^h(\alpha)(1 - u^h(\alpha))} n^{-1} + O(n^{-3/2} \log(n))
\]
\[
= 1 - \alpha - \frac{\epsilon_h(1 - \epsilon_h)z_{1-\alpha} \phi(z_{1-\alpha})}{p(1 - p)} n^{-1} + O(n^{-3/2} \log(n))
\]
where $\phi(\cdot)$ is the standard normal PDF and the $n^{-1}$ term is non-negative. Similar to the Ho and Lee (2005a) calibration, we can remove the analytic $n^{-1}$ term with the calibrated CI
\[
\left( -\infty, \hat{Q}_X^L \left[ u^h \left( \alpha + \frac{\epsilon_h(1 - \epsilon_h)z_{1-\alpha} \phi(z_{1-\alpha})}{p(1 - p)} n^{-1} \right) \right] \right).
\] (7)

which has CPE of $O(n^{-3/2} \log(n))$. We follow convention and define CPE $\equiv$ CP $- (1 - \alpha)$, where CP is the actual coverage probability and $1 - \alpha$ the desired confidence level.

By parallel argument, Hutson’s (1999) uncalibrated upper one-sided and two-sided CIs also have $O(n^{-1})$ CPE, or $O(n^{-3/2} \log(n))$ with calibration. For the upper one-sided case, again using (5), the $1 - \alpha$ Hutson CI and our calibrated CI are respectively given by
\[
\left( \hat{Q}_X^L[u^l(\alpha)], \infty \right),
\] (8)
and for equal-tailed two-sided CIs,
\[
\left( \hat{Q}_X^L\left[ u^l \left( \alpha/2 + \frac{\epsilon_h(1 - \epsilon_h)z_{1-\alpha/2} \phi(z_{1-\alpha})}{p(1 - p)} n^{-1} \right) \right], \infty \right),
\] (9)
In all cases, the order $n^{-1}$ CPE term is non-negative (indicating over-coverage).

For relatively extreme quantiles $p$ (given $n$), the $L$-statistic method cannot be computed because the $(n + 1)$th (or zeroth) order statistic is needed. In such cases, our code uses the Edgeworth expansion-based CI in Kaplan (2015). Alternatively, if bounds on $X$ are known a priori, they may be used in place of these “missing” order statistics to generate conservative CIs. As $n \to \infty$, the range of computable quantiles approaches $(0, 1)$.

To summarize CI construction, as in Hutson (1999) and implemented in our code:

1. Parameters: determine the sample size $n$, quantile of interest $p \in (0, 1)$, and coverage level $1 - \alpha$. Optionally, use the calibrated $\alpha$ in (7), (8), or (11).

2. Endpoint index computation: solve for $u^h$ (for two-sided or lower one-sided CIs) and/or $u^l$ (for two-sided or upper one-sided CIs) from (5), using (3). For a two-sided CI, replace $\alpha$ in (5) with $\alpha/2$.

3. CI construction: compute the lower and/or upper endpoints, respectively $\hat{Q}^L_{X}(u^l)$ and $\hat{Q}^L_{X}(u^h)$, using (1).

The hypothesis tests corresponding to all the foregoing CIs achieve optimal asymptotic power against local alternatives. The sample quantile is a semiparametric efficient estimator, so it suffices to show that power is asymptotically first-order equivalent to that of the test based on asymptotic normality. This is suggested by Lemma 3 and shown formally in the appendix. Theorem 4 collects all of our results on coverage and power.

**Theorem 4.** Let $z_\alpha$ denote the $\alpha$-quantile of the standard normal distribution, and let $\epsilon_h = (n + 1)u^h(\alpha) - \lfloor (n + 1)u^h(\alpha) \rfloor$ and $\epsilon_l = (n + 1)u^l(\alpha) - \lfloor (n + 1)u^l(\alpha) \rfloor$. Let Assumption A1 hold, and let A2 hold at quantile of interest $p$. Then, we have the following.

(i) The one-sided lower and upper CIs in (6) and (8) have coverage probability

$$1 - \alpha + \frac{\epsilon(1 - \epsilon)z_{1-\alpha} \phi(z_{1-\alpha})}{p(1-p)} n^{-1} + O(n^{-3/2} \log(n)),$$

with $\epsilon = \epsilon_h$ for the former and $\epsilon = \epsilon_l$ for the latter.
(ii) The equal-tailed, two-sided CI in (10) has coverage probability

\[
1 - \alpha + \frac{[\epsilon_h(1 - \epsilon_h) + \epsilon_\ell(1 - \epsilon_\ell)]z_{1-\alpha/2}\phi(z_{1-\alpha})}{u(1 - u)}n^{-1} + O(n^{-3/2}\log(n)).
\]

(iii) The calibrated one-sided lower, one-sided upper, and two-sided equal-tailed CIs given in (7), (8), and (11), respectively, have \(O(n^{-3/2}\log(n))\) CPE.

(iv) Against local sequence \(D_n = Q(p) + \kappa n^{-1/2}\), asymptotic power of lower one-sided (l), upper one-sided (u), and equal-tailed two-sided (t) CIs is

\[
\mathcal{P}_n^l(D_n) \to \Phi(z_\alpha + S), \quad \mathcal{P}_n^u(D_n) \to \Phi(z_\alpha - S), \quad \mathcal{P}_n^t(D_n) \to \Phi(z_{\alpha/2} + S) + \Phi(z_{\alpha/2} - S),
\]

where \(S \equiv \kappa f(F^{-1}(p))/\sqrt{p(1 - p)}\).

The equal-tailed property of our two-sided CIs is a type of median unbiasedness. If a CI for scalar parameter \(\theta\) has lower endpoint \(\hat{L}\) and upper endpoint \(\hat{H}\), then an equal-tailed CI is unbiased under the loss function \(L(\theta, \hat{L}, \hat{H}) = \max\{0, \theta - \hat{H}, \hat{L} - \theta\}\), following the definition in equation (5) of [Lehmann 1951]. Intuitively and mathematically, this is similar to a median unbiased point estimator being above or below \(\theta\) with equal probability. This median unbiased property may be desirable (e.g., [Andrews and Guggenberger 2014] footnote 11), although it is different than the usual “unbiasedness” where a CI is the inversion of an unbiased test.

More generally, in (10), we could replace \(u^l(\alpha/2)\) and \(u^h(\alpha/2)\) by \(u^l(t\alpha)\) and \(u^h((1 - t)\alpha)\) for \(t \in [0, 1]\), where \(t = 1/2\) is currently used. Different values of \(t\) may achieve different optimal properties, which we leave to future work.
4 Quantile inference: conditional

4.1 Setup and bias

Let $Q_{Y|X}(u; x)$ be the conditional $u$-quantile function of scalar outcome $Y$ given conditioning vector $X \in \mathcal{X} \subset \mathbb{R}^d$, evaluated at $X = x$. The sample $\{Y_i, X_i\}_{i=1}^n$ is drawn iid. If the conditional CDF $F_{Y|X}(\cdot)$ is strictly increasing and continuous at $u$, then $F_{Y|X}(Q_{Y|X}(u; x); x) = u$.

For interior point $x_0$ and quantile $p \in (0, 1)$, interest is in $Q_{Y|X}(p; x_0)$. Without loss of generality, let $x_0 = 0$.

If $X$ is discrete so that $P(X = 0) > 0$, we can take all the observations with $X_i = 0$ and compute a CI from the corresponding $Y_i$ values, using the method in Section 3. If there are $N_n$ observations with $X_i = 0$, then the CPE from Theorem 4 is $O(N_n^{-1})$ (uncalibrated). This $O(N_n^{-1})$ is $O(n^{-1})$ even with dependence like strong mixing among the $X_i$ (so that $N_n$ is almost surely of order $n$) as long as we have independent draws of $Y_i$ from the same $Q_{Y|X}(\cdot; 0)$.

If $X$ is continuous (or $N_n$ too small), we must include observations with $X_i \neq 0$. If $X$ contains mixed continuous and discrete components, then we can apply our method for continuous $X$ to each subsample corresponding to each unique value of the discrete subvector of $X$. The asymptotic rates are unaffected by the presence of discrete variables (although the finite-sample consequences may deserve more attention), so we focus on the case where all components of $X$ are continuous. Before proceeding, we present our assumptions and definitions for this section. We continue using the normalization $x_0 = 0$.

Definition 1 (local smoothness). Following Chaudhuri (1991, pp. 762–3): if, in a neighborhood of the origin, function $g(\cdot)$ is continuously differentiable through order $k$, and its $k$th derivatives are uniformly Hölder continuous with exponent $\gamma \in (0, 1]$, then $g(\cdot)$ has local smoothness of degree $s = k + \gamma$.

Assumption A3. Sampling of $(Y_i, X'_i)'$ is iid, for continuous scalar $Y_i$ and continuous vector $X_i \in \mathcal{X} \subseteq \mathbb{R}^d$. The point of interest $X = 0$ is in the interior of $\mathcal{X}$, and the quantile of interest
is $p \in (0, 1)$.

**Assumption A4.** The marginal density of $X$, denoted $f_X(\cdot)$, satisfies $f_X(0) > 0$ and has local smoothness $s_X = k_X + \gamma_X > 0$.

**Assumption A5.** For all $u$ in a neighborhood of $p$, $Q_{Y|X}(u; \cdot)$ (as a function of the second argument) has local smoothness $s_Q = k_Q + \gamma_Q > 0$.

**Assumption A6.** As $n \to \infty$, the bandwidth satisfies (i) $h \to 0$, (ii) $nh^d/\lfloor \log(n) \rfloor^2 \to \infty$.

**Assumption A7.** For all $u$ in a neighborhood of $p$ and all $x$ in a neighborhood of the origin, $f_{Y|X}(Q_{Y|X}(p; 0); 0)$ is uniformly bounded away from zero.

**Assumption A8.** For all $y$ in a neighborhood of $Q_{Y|X}(p; 0)$ and all $x$ in a neighborhood of the origin, $f_{Y|X}(y; x)$ has a second derivative in its first argument ($y$) that is uniformly bounded and continuous in $y$, having local smoothness $s_Y = k_Y + \gamma_Y > 2$.

For continuous $X_i$, the “effective sample” includes observations with $X_i \in C_h$, where $h$ is the bandwidth and $C_h = [-h, h]$ if $d = 1$. For $d > 1$, $C_h$ is a hypersphere or hypercube, similar to Chaudhuri (1991, pp. 762–3). Definition 2 refers to

$$C_h \equiv \{ x : x \in \mathbb{R}^d, \| x \| \leq h \},$$

$$N_n \equiv \# \{ Y_i : X_i \in C_h, 1 \leq i \leq n \}.\quad (12)$$

**Definition 2** (effective sample). Let $h$ denote the bandwidth and $p \in (0, 1)$ the quantile of interest. The “effective sample” consists of $Y_i$ values from observations with $X_i$ inside the window $C_h \subset \mathbb{R}^d$ defined in (12), wherein $\| \cdot \|$ denotes any norm on $\mathbb{R}^d$ (e.g., $L_2$ norm to get hypersphere $C_h$, $L_\infty$ norm for hypercube). The “effective sample size” is $N_n$ as in (13). Additionally, let $Q_{Y|X}(p; C_h)$ be the $p$-quantile of $Y$ given $X \in C_h$, satisfying $p = P(Y < Q_{Y|X}(p; C_h) \mid X \in C_h)$.

---

5Our $s_Q$ corresponds to variable $p$ in Chaudhuri (1991); Bhattacharya and Gangopadhyay (1990) use $s_Q = 2$ and $d = 1$. 

17
Given fixed values of \( n \) and \( h \), Assumption \( A3 \) implies that the \( Y_i \) in the effective sample are independent and identically distributed, which is needed to apply Theorem \( 4 \). However, they do not have the quantile function of interest, \( Q_{Y|X}(\cdot; 0) \), but rather the slightly biased \( Q_{Y|X}(\cdot; C_h) \). This is like drawing a global (any \( X_i \)) iid sample of wages, \( Y_i \), and restricting it to observations in Japan (\( X \in C_h \)) when our interest is only in Tokyo (\( X = 0 \)); our restricted \( Y_i \) constitute an iid sample from Japan, but the \( p \)-quantile wage in Japan may differ from that in Tokyo. Assumptions \( A4 \)–\( A6(i) \) and \( A8 \) are necessary for the calculation of this bias, \( Q_{Y|X}(p; C_h) - Q_{Y|X}(p; 0) \), in Lemma \( 5 \). Assumptions \( A6(ii) \) and \( A7 \) (and \( A3 \)) ensure \( N_n \overset{a.s.}{\to} \infty \). Assumptions \( A7 \) and \( A8 \) are conditional versions of Assumptions \( A2(i) \) and \( A2(ii) \), respectively. Their uniformity ensures uniformity of the remainder term in Theorem \( 4 \), accounting for the fact that the local sample’s distribution, \( F_{Y|C_h} \), changes with \( n \) (through \( h \) and \( C_h \)).

From \( A6(i) \), asymptotically \( C_h \) is entirely contained within the neighborhoods implicit in \( A4 \)–\( A5 \) and \( A8 \). This in turn allows us to examine only a local neighborhood around quantile of interest \( p \) (e.g., as in \( A5 \)) since the CI endpoints converge to the true value at a \( \sqrt{N_n} \) rate. Reassuringly, the optimal bandwidth rate derived in Section \( 4.2 \) turns out to be inside the bounds in \( A6 \).

**Definition 3** (steps for conditional \( L \)-statistic method). Given \( \{X_i, Y_i\}_{i=1}^n \) where vector \( X_i \) contains only continuously distributed variables, bandwidth \( h > 0 \), quantile \( p \in (0, 1) \), and desired coverage probability \( 1 - \alpha \), first \( C_h \) and \( N_n \) are calculated as in Definition \( 2 \). Using the \( Y_i \) from observations with \( X_i \in C_h \), a \( p \)-quantile CI is constructed as in \( \text{Hutson (1999)} \) if computable or \( \text{Kaplan (2015)} \) otherwise. If additional discrete conditioning variables exist, this method is repeated separately for each combination of discrete conditioning values (e.g., once for males and once for females). This procedure may be repeated for any number of \( x_0 \). For the bandwidth, we recommend the formulas in Section \( 4.4 \).

The \( X_i \) being iid helps guarantee that \( N_n \) is almost surely of order \( nh^d \). The \( h^d \) comes from the volume of \( C_h \). Larger \( h \) lowers CPE via \( N_n \) but raises CPE via bias. This tradeoff
determines the optimal rate at which \( h \to 0 \) as \( n \to \infty \). Using Theorem 4 and additional results on CPE from bias below, we determine the optimal value of \( h \).

Since our conditional quantile CI uses the subsample of \( Y_i \) with \( X_i \in C_h \) rather than \( X_i = 0 \), our CI is constructed for the biased conditional quantile \( Q_{Y|X}(p; C_h) \) (from Definition 2) rather than for \( Q_{Y|X}(p; 0) \). The bias characterized in Lemma 5 is the difference between these two population conditional quantiles.

**Lemma 5.** Define \( b \equiv \min\{s_Q, s_X+1, 2\} \) and \( B_h \equiv Q_{Y|X}(p; C_h) - Q_{Y|X}(p; 0) \). If Assumptions \( A4, A5, A6(i), \) and \( A8 \) hold, then the bias is of order

\[
|B_h| = O(h^b). \tag{14}
\]

With \( A4 \) and \( A5 \) strengthened to \( k_X \geq 1 \) and \( k_Q \geq 2 \), the bias is \( O(h^2) \) with remainder \( o(h^2) \).

With further strengthening to \( k_X \geq 2 \) and \( k_Q \geq 3 \), the bias remains \( O(h^2) \), but the remainder sharpens to \( o(h^3) \).

With \( d = 1 \), and defining

\[
Q_{Y|X}^{(0,1)}(p; 0) \equiv \left. \frac{\partial}{\partial x} Q_{Y|X}(p; x) \right|_{x=0}, \quad Q_{Y|X}^{(0,2)}(p; 0) \equiv \left. \frac{\partial^2}{\partial x^2} Q_{Y|X}(p; x) \right|_{x=0}, \quad \xi_p \equiv Q_{Y|X}(p; 0),
\]

\[
f_{Y|X}^{(0,1)}(y; 0) \equiv \left. \frac{\partial}{\partial x} f_{Y|X}(y; x) \right|_{x=0}, \quad f_{Y|X}^{(1,0)}(\xi_p; 0) \equiv \left. \frac{\partial}{\partial y} f_{Y|X}(y; 0) \right|_{y=\xi_p},
\]

and similarly for \( F_{Y|X}^{(0,1)}(\xi_p; 0) \) and \( F_{Y|X}^{(0,2)}(\xi_p; 0) \), the bias is

\[
B_h = \frac{h^2}{6} \left\{ 2Q_{Y|X}^{(0,1)}(p; 0)f_X'(0)/f_X(0) + Q_{Y|X}^{(0,2)}(p; 0) \\
+ 2f_{Y|X}^{(0,1)}(\xi_p; 0)Q_{Y|X}^{(0,1)}(p; 0)/f_{Y|X}(\xi_p; 0) \\
+ f_{Y|X}^{(1,0)}(\xi_p; 0) \left[ Q_{Y|X}^{(0,1)}(p; 0) \right]^2/f_{Y|X}(\xi_p; 0) \right\} + R
\]

\[
= -h^2 \frac{f_X(0)f_{Y|X}^{(0,2)}(\xi_p; 0)}{6f_X(0)f_{Y|X}(\xi_p; 0)} + 2f_X'(0)F_{Y|X}^{(0,1)}(\xi_p; 0) + R,
\]

where \( R = o(h^2) \) or \( R = o(h^3) \) as discussed.

The dominant term in the bias when \( k_X \geq 1 \), \( k_Q \geq 2 \), and \( d = 1 \) is the same as the
bias in Bhattacharya and Gangopadhyay [1990], who derive it using different arguments. The maximum value \( b = 2 \) (making the bias at least \( O(h^2) \)) is due to our implicit use of a uniform kernel, which is a second-order kernel. Consequently, there is no benefit to having smoothness greater than \((s_Q, s_X, s_Y) = (2, 1, 1) + \epsilon\) for arbitrarily small \( \epsilon > 0 \).

### 4.2 Optimal CPE order

The CPE-optimal bandwidth minimizes the sum of the two dominant high-order CPE terms. It must be small enough to control the \( O(h^b + N_nh^{2b}) \) (two-sided) CPE from bias, but large enough to control the \( O(N_n^{-1}) \) CPE from applying the unconditional \( L \)-statistic method. As before, \( \text{CPE} \equiv \text{CP} - (1 - \alpha) \), so CPE is positive when there is over-coverage and negative when there is under-coverage. This means the equivalent hypothesis test is size distorted when CPE is negative. The following theorem summarizes optimal bandwidth and CPE results.

**Theorem 6.** Let Assumptions A3–A8 hold, and define \( b \equiv \min\{s_Q, s_X + 1, 2\} \). The following results are for the method in Definition 3 when Hutson (1999) is used.

For a one-sided CI, the bandwidth \( h^* \) minimizing CPE has rate \( h^* \asymp n^{-3/(2b+3d)} \). This corresponds to overall CPE of \( O(n^{-2b/(2b+3d)}) \).

For two-sided inference, the optimal bandwidth rate is \( h^* \asymp n^{-1/(b+d)} \), and the optimal CPE is \( O(n^{-b/(b+d)}) \). If Kaplan (2015) is used instead of Hutson (1999), then \( h^* \asymp n^{-5/(5d+6b)} \) and CPE is \( O(n^{-5b/(5d+6b)}) \).

Using the calibration in Section 3, the nearly (up to \( \log(n) \)) CPE-optimal two-sided bandwidth rate is \( h^* \asymp n^{-5/(4b+5d)} \), yielding CPE of \( O(n^{-6b/(4b+5d)} \log(n)) \). The nearly optimal calibrated one-sided bandwidth rate is \( h^* \asymp n^{-2/(b+2d)} \), yielding CPE of \( O(n^{-3b/(2b+4d)} \log(n)) \).

If used in a semi-nonparametric (e.g., partially linear) model like in Fan and Liu (2015), depending on \( d \), the CPE-optimal bandwidth and CPE rates may differ from Theorem 6.

For example, with \( d = 1, b = 2 \), and a two-sided CI, the bias from Lemma 5 with \( h \asymp n^{-1/3} \)
is $O(h^2) = O(n^{-2/3})$. Since there is larger-order bias from plugging in a root-$n$ normal estimator of the parametric part of the model, the CPE-optimal $h$ will be smaller, and the overall CPE larger. Precise characterization of the optimal $h$ and CPE rates in seminonparametric models would require a statement about $\hat{\beta}$ such as $P(|\hat{\beta} - \beta| > n^\delta) = O(n^n)$, where the new bias order is $n^\delta$ and the overall CPE is of order $n^n$ or larger. For now, we note that our method is at least first-order accurate in such models, with Lemma 3 explicitly connecting our method to results in Fan and Liu (2015).

4.3 CPE comparison with other methods

Theorem 6 suggests that for the most common values of dimension $d$ and most plausible values of smoothness $s_Q$, our method is preferred to inference based on asymptotic normality (or, equivalently, unsmoothed bootstrap) with a local polynomial estimator. Only our uncalibrated method is compared here.

The only chance for normality to yield smaller CPE is to greatly reduce bias by using a very large local polynomial. However, our method has smaller CPE when $d = 1$ or $d = 2$ even if $s_Q = \infty$.

In finite samples, the normality approach may have additional error from estimation of the asymptotic variance, which includes the probability density of the error term at zero as discussed in Chaudhuri (1991, pp. 764–766). Additionally, large local polynomials cannot perform well unless there is a large local sample size.

For the local polynomial approach, the CPE-optimal bandwidth and optimal CPE can be derived from the results in Chaudhuri (1991). We balance the CPE from the bias with additional CPE from the Bahadur remainder from his Theorem 3.3(ii). The CPE from bias is of order $N_n^{1/2} h^s_Q$. Chaudhuri (1991, Thm. 3.3) gives a Bahadur-type expansion of the local polynomial quantile regression estimator that has remainder $R_n \sqrt{N_n} = O(N_n^{-1/4})$ (up to log terms) as in Bahadur (1966), but recently Portnoy (2012) has shown that the CPE is nearly $O(N_n^{-1/2})$ in such cases. Solving $N_n^{1/2} h^{s_Q} \approx N_n^{-1/2} \approx (nh^d)^{-1/2}$ yields $h^* \approx n^{-1/(s_Q+d)}$.
and optimal CPE (nearly) $O(\sqrt{N_n}B_n) = O(N_n^{-1/2}) = O(n^{-s_Q/(2s_Q+2d)})$.

Figure 2: Two-sided CPE comparison between new (“L-stat”) method and the local polynomial asymptotic normality method based on Chaudhuri (1991). Left: with $s_Q = 2$, writing CPE as $n^\kappa$, comparison of $\kappa$ for different methods and different values of $d$. Right: required smoothness $s_Q$ for the local polynomial normality-based CPE to match that of L-stat, as well as the corresponding number of terms in the local polynomial, for different $d$.

As illustrated in the left panel of Figure 2, if $s_Q = 2$ (one Lipschitz-continuous derivative), then the optimal CPE from asymptotic normality is nearly $O(n^{-2/(4+2d)})$, which is always larger than our method’s CPE. With $d = 1$, this is $n^{-1/3}$, significantly larger than our $n^{-2/3}$ (one-sided: $n^{-4/7}$). With $d = 2$, $n^{-1/4}$ is larger than our $n^{-1/2}$ (one-sided: $n^{-2/5}$). It remains larger for all $d$, even for one-sided inference, since the bias is the same for both methods and the unconditional $L$-statistic inference is more accurate than normality.

From another perspective, as in the right panel of Figure 2, what amount of smoothness (and local polynomial degree) is needed for asymptotic normality to match our method’s CPE? For the most common cases of $d \in \{1, 2\}$ (one-sided: $d = 1$), normality is worse even with infinite smoothness. With $d = 3$ (one-sided: $d = 2$), normality needs $s_Q \geq 12$ to match our CPE. If $n$ is large, then maybe such a high degree ($k_Q \geq 11$) local polynomial will be appropriate, but often it is not. Interaction terms are required, so an 11th-degree polynomial has $\sum_{T=d-1}^{k_Q+d-1} \binom{T}{d-1} = 364$ terms. As $d \to \infty$, the required smoothness approaches $s_Q = 4$.
(one-sided: 8/3) from above, though again the number of terms in the local polynomial grows with \( d \) as well as \( kQ \) and may thus still be prohibitive in finite samples.

Basic bootstraps claim no refinement over asymptotic normality, so the foregoing discussion also applies to such methods. Without using smoothed or \( m \)-out-of-\( n \) bootstrap, which require additional smoothing parameter choice and computation time, Studentization offers no CPE improvement. Gangopadhyay and Sen (1990) examine the bootstrap percentile method for a uniform kernel (or nearest neighbor) estimator with \( d = 1 \) and \( sQ = 2 \), but they only show first-order consistency. Based on their (2.14), the optimal error seems to be \( O(n^{-1/3}) \) when \( h \asymp n^{-1/3} \) (matching our derivation from Chaudhuri (1991)) to balance the bias and remainder terms, improved by Portnoy (2012) from \( O(n^{-2/11}) \) CPE when \( h \asymp n^{-3/11} \).

### 4.4 Plug-in bandwidth

Keeping track of terms more carefully, we can solve for not just the rate of the CPE-optimal bandwidth, but its value. To avoid recursive dependence on \( \epsilon \) (the interpolation weight), we fix its value. This does not achieve the theoretical optimum, but it remains close even in small samples and seems to work well in practice. The CPE-optimal bandwidth value derivation is shown for \( d = 1 \) in Appendix B, a plug-in version of which is implemented in our code. For reference, the plug-in bandwidth expressions are collected here. Standard normal quantiles are denoted, for example, \( z_{1-\alpha} \) for the \((1 - \alpha)\)-quantile such that \( \Phi(z_{1-\alpha}) = 1 - \alpha \). We let \( \hat{B}_h \) denote the estimator of bias term \( B_h \); \( \hat{f}_X \) the estimator of \( f_X(x_0) \); \( \hat{f}'_X \) the estimator of \( f'_X(x_0) \); \( \hat{F}_{Y|X}^{(0,1)} \) the estimator of \( F_{Y|X}^{(0,1)}(\xi_p; x_0) \); and \( \hat{F}_{Y|X}^{(0,2)} \) the estimator of \( F_{Y|X}^{(0,2)}(\xi_p; x_0) \), where \( \xi_p \equiv Q_{Y|X}(p; x_0) \).

We recommend the following when \( d = 1 \).
• For one-sided inference, let

\[ \hat{h}_{+-} = n^{-3/7} \left( \frac{2z_{1-\alpha}}{3 \left[ p(1-p)\hat{f}_X \right]^{1/2} \left\{ \hat{f}_X \hat{F}_{Y|X}^{(0,2)} + 2\hat{f}_X \hat{F}_{Y|X}^{(0,1)} \right\} } \right)^{2/7}, \tag{15} \]

\[ \hat{h}_{++} = -0.770\hat{h}_{+-}. \tag{16} \]

For lower one-sided inference, \( \hat{h}_{+-} \) should be used if \( \hat{B}_h < 0 \), and \( \hat{h}_{++} \) otherwise.

For upper one-sided inference, \( \hat{h}_{++} \) should be used if \( \hat{B}_h < 0 \), and \( \hat{h}_{+-} \) otherwise.

Alternatively, always using \( |\hat{h}_{++}| \) is simpler but will sometimes be conservative.

• For two-sided inference on the median,

\[ \hat{h} = n^{-1/3} \left\| \hat{f}_X \hat{F}_{Y|X}^{(0,2)} + 2\hat{f}_X \hat{F}_{Y|X}^{(0,1)} \right\|^{-1/3}. \tag{17} \]

• For two-sided inference with general \( p \in (0,1) \), and equivalent to (17) with \( p = 1/2 \),

\[ \hat{h} = n^{-1/3} \left( \frac{(\hat{B}_h/|\hat{B}_h|)(1-2p) + \sqrt{(1-2p)^2 + 4}}{2 \left\| \hat{f}_X \hat{F}_{Y|X}^{(0,2)} + 2\hat{f}_X \hat{F}_{Y|X}^{(0,1)} \right\|} \right)^{1/3}. \tag{18} \]

We also suggest (and implement in our code) shifting toward a larger bandwidth as \( n \to \infty \). Once the absolute differences in CPE are small, a larger bandwidth (that still maintains \( o(1) \) CPE) is preferable since it yields shorter CIs. However, the usual question of how big \( n \) needs to be remains to be formalized. Currently, we use a coefficient of \( \max\{1, n/1000\}^{5/60} \) that keeps the CPE-optimal bandwidth for \( n \leq 1000 \) and then moves toward a \( n^{-1/20} \) undersmoothing of the MSE-optimal bandwidth rate, as in Fan and Liu (2015).

### 4.5 Joint confidence intervals and a possible uniform band

Beyond the pointwise CIs in Definition 3 the usual Bonferroni adjustment \( \alpha/m \) gives joint CIs over \( m \) different values of \( x_0 \). If the \( C_h \) are mutually exclusive, then the local samples are independent due to Assumption A3 and the adjustment can be refined to \( 1 - (1 - \alpha)^{1/m} \),
which maintains exact joint coverage (not conservative like Bonferroni).

The number of such mutually exclusive windows grows as $n \to \infty$, which may be helpful for constructing a uniform confidence band. If Assumption A5 is strengthened to a uniform bound on the second derivative (wrt $X$) of the conditional quantile function, then the local bandwidths have a common asymptotic rate, $h$, and there can be order $1/h$ mutually exclusive local samples. The error from a linear approximation of the conditional quantile function over a span of $h$ is $O(h^2)$. This is smaller than the length of the CIs, which is order $1/\sqrt{nh^d} = h$ when using the CPE-optimal bandwidth rate $h \asymp n^{-1/(2+d)}$ for two-sided CIs from Theorem 6 with $b = 2$.

The growing number of CIs is not problematic given independent sampling, but the fact that $\alpha \to 0$ (at rate $h$) has not been rigorously treated. We leave such proof to future work, suggesting due caution in applying the uniform band until then. As discussed in Section 6 a Hotelling (1939) tube-based calibration of $\alpha$ also appears to yield reasonable uniform confidence bands, but it has even less formal justification.

5 Empirical application

We present an application of our $L$-statistic inference to Engel (1857) curves. Code is available from the latter author’s website, and the data are publicly available.

Banks et al. (1997) argue that a linear Engel curve is sufficient for certain categories of expenditure, while adding a quadratic term suffices for others. Their Figure 1 shows non-parametrically estimated mean Engel curves (budget share $W$ against log total expenditure $\ln(X)$) with 95% pointwise CIs at the deciles of the total expenditure distribution, using a subsample of 1980–1982 U.K. Family Expenditure Survey (FES) data.

We present a similar examination, but for quantile Engel curves in the 2001–2012 U.K. Living Costs and Food Surveys (Office for National Statistics and Department for Environment, Food and Rural Affairs, 2012), which is a successor to the FES. We examine the same
four categories as in the original analysis: food; fuel, light, and power (“fuel”); clothing and footwear (“clothing”); and alcohol. We use the subsample of households with one adult male and one adult female (and possibly children) living in London or the South East, leaving 8,528 observations. Expenditure amounts are adjusted to 2012 nominal values using annual CPI data.

Table 1: L-statistic 99% CIs for various quantiles (p) of the budget share distribution, for different categories of expenditure described in the text.

<table>
<thead>
<tr>
<th>Category</th>
<th>p = 0.5</th>
<th>p = 0.75</th>
<th>p = 0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>food</td>
<td>(0.1532,0.1580)</td>
<td>(0.2095,0.2170)</td>
<td>(0.2724,0.2818)</td>
</tr>
<tr>
<td>fuel</td>
<td>(0.0275,0.0289)</td>
<td>(0.0447,0.0470)</td>
<td>(0.0692,0.0741)</td>
</tr>
<tr>
<td>clothing</td>
<td>(0.0135,0.0152)</td>
<td>(0.0362,0.0397)</td>
<td>(0.0697,0.0761)</td>
</tr>
<tr>
<td>alcohol</td>
<td>(0.0194,0.0226)</td>
<td>(0.0548,0.0603)</td>
<td>(0.1012,0.1111)</td>
</tr>
</tbody>
</table>

Table 1 shows unconditional L-statistic CIs (Hutson, 1999) for various quantiles of the budget share distributions for the four expenditure categories. (Due to the sample size, calibrated CIs are identical at the precision shown.) These capture some population features, but the conditional quantiles are of more interest.

Figure 3 is comparable to Figure 1 of Banks et al. (1997) but with 90% joint (over all nine expenditure levels) CIs instead of 95% pointwise CIs, alongside quadratic quantile regression estimates. Joint CIs are more intuitive for assessing the shape of a function since they jointly cover all corresponding points on the true curve with 90% probability, rather than any given single point. The CIs are interpolated only for visual convenience. Although some of the joint CI shapes do not look quadratic, the only cases where the quadratic fit lies outside one of the intervals are for alcohol at the conditional median and clothing at the conditional upper quartile, and neither is a radical departure. With a 90% confidence level and 12 confidence sets, we would not be surprised if one or two did not cover the true quantile Engel curve completely. Importantly, the CIs are relatively precise, too; the linear fit is rejected in 8 of 12 cases. Altogether, this evidence suggests that the benefits of a quadratic (but not linear)
Figure 3: Joint (over the nine expenditure levels) 90% confidence intervals for quantile Engel curves: food (top left), fuel (top right), clothing (bottom left), and alcohol (bottom right).
approximation probably outweigh the cost of approximation error.

The $L$-statistic joint CIs (“L-stat”) in Figure 4 are the same as those for $p = 0.5$ and $p = 0.9$ in Figure 3, but now a nonparametric (instead of quadratic) conditional quantile estimate is shown, along with joint CIs from Fan and Liu (2015). The L-stat CIs are generally shorter, but there can be exceptions, as seen especially in the bottom right graph. Of course, shorter is not better if coverage probability is sacrificed; we explore both properties in the simulations of Section 6.

Figure 4: Joint (over the nine expenditure levels) 90% confidence intervals for quantile Engel curves, $p = 0.5$ and $p = 0.9$: food (top left), fuel (top right), clothing (bottom left), and alcohol (bottom right).
6 Simulation study

Code for our $L$-statistic methods in R is available on the latter author’s website\footnote{In the code, if $N_n$ is not large enough to compute Hutson (1999) at quantile $p$, then the method in Kaplan (2015) is used instead. If $N_n$ is too small even for Kaplan (2015), then a method based on extreme value theory is recommended instead.} as is the simulation code.

6.1 Unconditional simulations

We compare two-sided unconditional CIs from the following methods: “L-stat” from Section 3 originally in Hutson (1999); “BH” from Beran and Hall (1993); “Norm” using the sample quantile’s asymptotic normality and kernel-estimated variance; “K15” from Kaplan (2015); and “BStsym,” a symmetric Studentized bootstrap (99 draws) with bootstrapped variance (100 draws). (Other bootstraps were consistently worse in terms of coverage: (asymmetric) Studentized bootstrap, and percentile bootstrap with and without symmetry.)

Unless otherwise noted, $\alpha = 0.05$ and 10,000 replications of each DGP were simulated.

Overall, L-stat and BH have the most accurate coverage probability (CP), avoiding under-coverage while maintaining shorter length than other methods achieving at least 95% CP. Near the median, L-stat and BH are nearly identical. In the tails, L-stat is shorter than BH for some distributions and comes closer to being equal-tailed. Farther into the tails, BH is not computable, but L-stat maintains a favorable combination of CP and length compared with other methods.

Table 2 shows nearly exact CP for both L-stat and BH when $n = 25$ and $p = 0.5$. The normal CI can be slightly shorter, but it under-covers. The bootstrap has only slight under-coverage, and K15 none, but their CIs are longer than L-stat and BH. In the Supplemental Appendix, results are shown for additional distributions and for $n = 9$, but the qualitative points are the same.

Table 3 shows a case in the lower tail with $n = 99$ where BH cannot be computed (because it needs the zeroth order statistic). Even then, L-stat’s CP remains almost exact, and it is
Table 2: Coverage probability and median CI length, $1 - \alpha = 0.95$; various $n$, $p$, and distributions of $X_i (F)$ shown in table. “Too high” is the proportion of simulations in which the lower endpoint was above the true value, $F^{-1}(p)$, and “too low” is the proportion when the upper endpoint was below $F^{-1}(p)$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$p$</th>
<th>$F$</th>
<th>Method</th>
<th>CP</th>
<th>Too low</th>
<th>Too high</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.5</td>
<td>Normal</td>
<td>L-stat</td>
<td>0.953</td>
<td>0.022</td>
<td>0.025</td>
<td>0.99</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Normal</td>
<td>BH</td>
<td>0.955</td>
<td>0.021</td>
<td>0.024</td>
<td>1.00</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Normal</td>
<td>Norm</td>
<td>0.942</td>
<td>0.028</td>
<td>0.030</td>
<td>1.02</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Normal</td>
<td>K15</td>
<td>0.971</td>
<td>0.014</td>
<td>0.015</td>
<td>1.19</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Normal</td>
<td>BStsym</td>
<td>0.942</td>
<td>0.028</td>
<td>0.030</td>
<td>1.13</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Uniform</td>
<td>L-stat</td>
<td>0.953</td>
<td>0.022</td>
<td>0.025</td>
<td>0.37</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Uniform</td>
<td>BH</td>
<td>0.954</td>
<td>0.021</td>
<td>0.025</td>
<td>0.37</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Uniform</td>
<td>Norm</td>
<td>0.908</td>
<td>0.046</td>
<td>0.046</td>
<td>0.35</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Uniform</td>
<td>K15</td>
<td>0.963</td>
<td>0.018</td>
<td>0.020</td>
<td>0.44</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Uniform</td>
<td>BStsym</td>
<td>0.937</td>
<td>0.031</td>
<td>0.032</td>
<td>0.45</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Exponential</td>
<td>L-stat</td>
<td>0.953</td>
<td>0.024</td>
<td>0.023</td>
<td>0.79</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Exponential</td>
<td>BH</td>
<td>0.954</td>
<td>0.024</td>
<td>0.022</td>
<td>0.80</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Exponential</td>
<td>Norm</td>
<td>0.924</td>
<td>0.056</td>
<td>0.020</td>
<td>0.75</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Exponential</td>
<td>K15</td>
<td>0.968</td>
<td>0.022</td>
<td>0.010</td>
<td>0.96</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>Exponential</td>
<td>BStsym</td>
<td>0.941</td>
<td>0.039</td>
<td>0.020</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Table 3: Coverage probability and median CI length, as in Table 2

<table>
<thead>
<tr>
<th>$n$</th>
<th>$p$</th>
<th>$F$</th>
<th>Method</th>
<th>CP</th>
<th>Too low</th>
<th>Too high</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>99</td>
<td>0.037</td>
<td>Normal</td>
<td>L-stat</td>
<td>0.951</td>
<td>0.023</td>
<td>0.026</td>
<td>1.02</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Normal</td>
<td>BH</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Normal</td>
<td>Norm</td>
<td>0.925</td>
<td>0.016</td>
<td>0.059</td>
<td>0.83</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Normal</td>
<td>K15</td>
<td>0.970</td>
<td>0.009</td>
<td>0.021</td>
<td>1.55</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Normal</td>
<td>BStsym</td>
<td>0.950</td>
<td>0.020</td>
<td>0.030</td>
<td>1.20</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Cauchy</td>
<td>L-stat</td>
<td>0.950</td>
<td>0.022</td>
<td>0.028</td>
<td>39.37</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Cauchy</td>
<td>BH</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Cauchy</td>
<td>Norm</td>
<td>0.784</td>
<td>0.082</td>
<td>0.134</td>
<td>18.90</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Cauchy</td>
<td>K15</td>
<td>0.957</td>
<td>0.002</td>
<td>0.041</td>
<td>36.55</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Cauchy</td>
<td>BStsym</td>
<td>0.961</td>
<td>0.002</td>
<td>0.037</td>
<td>48.77</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Uniform</td>
<td>L-stat</td>
<td>0.951</td>
<td>0.024</td>
<td>0.026</td>
<td>0.07</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Uniform</td>
<td>BH</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Uniform</td>
<td>Norm</td>
<td>0.990</td>
<td>0.000</td>
<td>0.010</td>
<td>0.12</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Uniform</td>
<td>K15</td>
<td>0.963</td>
<td>0.028</td>
<td>0.009</td>
<td>0.11</td>
</tr>
<tr>
<td>99</td>
<td>0.037</td>
<td>Uniform</td>
<td>BStsym</td>
<td>0.924</td>
<td>0.053</td>
<td>0.022</td>
<td>0.08</td>
</tr>
</tbody>
</table>
closest to being equal-tailed. The normal CI under-covers for the first two $F$ and is almost twice as long as the L-stat CI for the third $F$. BStsym has less under-coverage, and K15 none, but both are generally longer than L-stat. Again, results for additional $F$ are in the Supplemental Appendix, with similar patterns, as well as results for $n = 99$ and $p = 0.05$ or $p = 0.95$, where again BH cannot be computed.

Table 4: Coverage probability and median CI length, as in Table 2

<table>
<thead>
<tr>
<th>$n$</th>
<th>$p$</th>
<th>$F$</th>
<th>Method</th>
<th>CP</th>
<th>Too low</th>
<th>Too high</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.2</td>
<td>Normal</td>
<td>L-stat</td>
<td>0.956</td>
<td>0.023</td>
<td>0.021</td>
<td>1.18</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Normal</td>
<td>BH</td>
<td>0.967</td>
<td>0.028</td>
<td>0.006</td>
<td>1.33</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Normal</td>
<td>Norm</td>
<td>0.944</td>
<td>0.011</td>
<td>0.045</td>
<td>1.13</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Normal</td>
<td>K15</td>
<td>0.952</td>
<td>0.006</td>
<td>0.042</td>
<td>1.42</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Normal</td>
<td>BStsym</td>
<td>0.944</td>
<td>0.022</td>
<td>0.034</td>
<td>1.34</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Cauchy</td>
<td>L-stat</td>
<td>0.960</td>
<td>0.024</td>
<td>0.016</td>
<td>5.05</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Cauchy</td>
<td>BH</td>
<td>0.965</td>
<td>0.028</td>
<td>0.006</td>
<td>7.87</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Cauchy</td>
<td>Norm</td>
<td>0.909</td>
<td>0.014</td>
<td>0.077</td>
<td>2.43</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Cauchy</td>
<td>K15</td>
<td>0.946</td>
<td>0.001</td>
<td>0.053</td>
<td>4.00</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Cauchy</td>
<td>BStsym</td>
<td>0.959</td>
<td>0.005</td>
<td>0.036</td>
<td>5.18</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Uniform</td>
<td>L-stat</td>
<td>0.952</td>
<td>0.022</td>
<td>0.026</td>
<td>0.29</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Uniform</td>
<td>BH</td>
<td>0.963</td>
<td>0.026</td>
<td>0.010</td>
<td>0.30</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Uniform</td>
<td>Norm</td>
<td>0.960</td>
<td>0.006</td>
<td>0.034</td>
<td>0.33</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Uniform</td>
<td>K15</td>
<td>0.953</td>
<td>0.015</td>
<td>0.032</td>
<td>0.39</td>
</tr>
<tr>
<td>25</td>
<td>0.2</td>
<td>Uniform</td>
<td>BStsym</td>
<td>0.930</td>
<td>0.042</td>
<td>0.028</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Table 4 shows a case with $p = 0.2$ (away from the median) where BH can be computed. The patterns among the normal, K15, and BStsym CIs are similar to before. L-stat and BH both attain 95% CP in each case, but L-stat is much closer to equal-tailed, and L-stat is shorter or the same length. The Supplemental Appendix contains additional, similar results.

Table 5 compares the original two-sided CI in equation (10) with the calibrated (“Calib”) version in (11). Even with small $n$, the true differences are small, so we use a larger number of replications ($10^5$) to reduce simulation error. By construction, the calibrated $\alpha$ is always larger than the original $\alpha$, which is why the Calib CI is always shorter and has smaller CP. Similarly, while the dominant $n^{-1}$ term in the L-stat CPE is always positive (i.e., over-
Table 5: Coverage probability and median CI length, as in Table 2; $10^5$ replications. “L-stat” uses equation (10); “Calib” uses (11).

<table>
<thead>
<tr>
<th>$n$</th>
<th>$p$</th>
<th>$F$</th>
<th>Method</th>
<th>CP</th>
<th>Too low</th>
<th>Too high</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.35</td>
<td>Normal</td>
<td>L-stat</td>
<td>0.959</td>
<td>0.023</td>
<td>0.018</td>
<td>1.73</td>
</tr>
<tr>
<td>10</td>
<td>0.35</td>
<td>Normal</td>
<td>Calib</td>
<td>0.952</td>
<td>0.025</td>
<td>0.023</td>
<td>1.64</td>
</tr>
<tr>
<td>10</td>
<td>0.40</td>
<td>Normal</td>
<td>L-stat</td>
<td>0.965</td>
<td>0.020</td>
<td>0.015</td>
<td>1.69</td>
</tr>
<tr>
<td>10</td>
<td>0.40</td>
<td>Normal</td>
<td>Calib</td>
<td>0.943</td>
<td>0.030</td>
<td>0.027</td>
<td>1.50</td>
</tr>
<tr>
<td>10</td>
<td>0.45</td>
<td>Normal</td>
<td>L-stat</td>
<td>0.953</td>
<td>0.023</td>
<td>0.024</td>
<td>1.62</td>
</tr>
<tr>
<td>10</td>
<td>0.45</td>
<td>Normal</td>
<td>Calib</td>
<td>0.948</td>
<td>0.027</td>
<td>0.025</td>
<td>1.59</td>
</tr>
<tr>
<td>10</td>
<td>0.50</td>
<td>Normal</td>
<td>L-stat</td>
<td>0.962</td>
<td>0.019</td>
<td>0.019</td>
<td>1.64</td>
</tr>
<tr>
<td>10</td>
<td>0.50</td>
<td>Normal</td>
<td>Calib</td>
<td>0.942</td>
<td>0.029</td>
<td>0.029</td>
<td>1.47</td>
</tr>
</tbody>
</table>

coverage), the dominant term in the Calib CPE may be positive or negative, although the worst under-coverage in Table 5 is still 94.3%. Since the calibration is separate for the upper and lower endpoints, it also makes the Calib CI (slightly) more equal-tailed.

We also compare L-stat with an unconditional version of the order statistic method in Example 2.1 of Fan and Liu (2015). This isolates the effect of using the beta reference distribution rather than the normal, as well as the effect of interpolation. One method uses a normal approximation to determine $u^b$ and $u^l$ but still interpolates (“Normal”), while a second uses a normal approximation with no interpolation (“Norm/floor”) as in equations (5) and (6) of Fan and Liu (2015). We let $n = 19$, which can be thought of as the local sample size in the conditional setting. We use $Y_i \overset{iid}{\sim} N(0, 1)$, $\alpha = 0.1$, and 1,000 simulation replications, for various $p$.

Table 6: Coverage probability and median CI length, $n = 19$, $Y_i \overset{iid}{\sim} N(0, 1)$, $1 - \alpha = 0.90$, 1,000 replications, various $p$. L-stat: beta reference, interpolation; Normal: normal reference, interpolation; Norm/floor: normal reference, no interpolation.

<table>
<thead>
<tr>
<th>Method</th>
<th>two-sided CP</th>
<th>median length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p = 0.15$</td>
<td>$p = 0.25$</td>
</tr>
<tr>
<td>L-stat</td>
<td>0.905</td>
<td>0.901</td>
</tr>
<tr>
<td>Normal</td>
<td>NA</td>
<td>0.926</td>
</tr>
<tr>
<td>Norm/floor</td>
<td>NA</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td>$p = 0.15$</td>
<td>$p = 0.25$</td>
</tr>
<tr>
<td>L-stat</td>
<td>1.20</td>
<td>1.03</td>
</tr>
<tr>
<td>Normal</td>
<td>NA</td>
<td>1.22</td>
</tr>
<tr>
<td>Norm/floor</td>
<td>NA</td>
<td>1.47</td>
</tr>
</tbody>
</table>
Table 7: Probabilities of CI being too low (i.e., below true value) or too high; parameters same as Table 6.

<table>
<thead>
<tr>
<th>Method</th>
<th>too low</th>
<th></th>
<th>too high</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p = 0.15$</td>
<td>$p = 0.25$</td>
<td>$p = 0.5$</td>
<td>$p = 0.15$</td>
</tr>
<tr>
<td>L-stat</td>
<td>0.048</td>
<td>0.050</td>
<td>0.052</td>
<td>0.047</td>
</tr>
<tr>
<td>Normal</td>
<td>NA</td>
<td>0.062</td>
<td>0.045</td>
<td>NA</td>
</tr>
<tr>
<td>Norm/floor</td>
<td>NA</td>
<td>0.083</td>
<td>0.087</td>
<td>NA</td>
</tr>
</tbody>
</table>

For the 0.2-quantile and below, the normal reference chooses the zeroth order statistic, which does not exist (hence the “NA” entries). In contrast, the beta reference can be used for inference even below the 0.15-quantile. The L-stat CP for $p = 0.15$ is 0.905 (Table 6), and it is still equal-tailed.

The effect of the normal approximation in Table 6 is to make the CI needlessly longer. L-stat already has CP within a few tenths of a percentage point of $1 - \alpha$. Compared with the interpolated normal method, not interpolating makes the CI longer for $p = 0.25$, but shorter for $p = 0.5$, which leads to under-coverage.

Table 7 shows that the normal reference loses the equal-tailed property of the L-stat CIs. When interpolating, the Normal CI is equal-tailed at $p = 0.5$, but not at $p = 0.25$, where the CI is too low 6.2% of the time and too high 1.2% of the time, whereas the L-stat CI is too low 5.0% of the time and too high 4.9% of the time. Not interpolating loses the equal-tailed property even at the median, and the $p = 0.25$ becomes essentially a one-sided CI.

6.2 Conditional simulations

For clarity, we now explicitly write $x_0$ as the point of interest, instead of taking $x_0 = 0$; we also take $d = 1$, $b = 2$, and focus on two-sided inference, both pointwise and joint.

For estimating $f_X(x_0)$ in order to calculate the plug-in bandwidth, any kernel density estimator will suffice. We use \texttt{kde} from package \texttt{ks} \cite{Duong2012}, with the Gaussian-based bandwidth. From the same package, \texttt{kdde} estimates $f'_X(x_0)$. Both functions work for up to
\( d = 6 \)-dimensional \( X \). The objects \( F_{Y \mid X}^{(0,1)}(\xi; x_0) \) and \( F_{Y \mid X}^{(0,2)}(\xi; x_0) \) are estimated using local cubic (mean) regression of \( 1\{Y_i \leq \xi\} \) on \( X \).

We now present conditional quantile simulations comparing our \( L \)-statistic method ("L-stat") with a variety of others. The first ("rqss") is directly available in the popular quantreg package in R (Koenker, 2012). The regression quantile smoothing spline function \texttt{rqss} is used as on page 10 of its vignette, with the Schwarz (1978) information criterion (SIC) for model selection; pointwise CIs come from \texttt{predict.rqss}, and uniform bands from a Hotelling (1939) tube approach coded in \texttt{plot.rqss}.

The second method is a local polynomial following Chaudhuri (1991) but with bootstrapped standard errors ("boot")\(^8\). For this method, we adjust our method’s plug-in bandwidth by a factor of \( n^{1/12} \) to yield the local cubic CPE-optimal bandwidth rate. The function \texttt{rq} performs the local cubic point estimation, and \texttt{summary.rq} generates the bootstrap standard errors; we use 299 replications.

The third method uses the asymptotic normality of a local linear estimator, using results and ideas from Qu and Yoon (2015), although they are concerned more with uniform (over quantiles) than pointwise inference. They suggest using the MSE-optimal bandwidth (Corollary 1) and a conservative type of bias correction for the CIs (Remark 7) that increases the length of the CI. We tried a uniform kernel, but show only results for the version with a Gaussian kernel ("QYg") since it was usually better.

The fourth method is from Section 3.1 in Fan and Liu (2015), based on a symmetrized \( k \)-NN estimator using a bisquare kernel ("FLb"). We use the same code from their simulations (graciously provided to us)\(^9\). Interestingly, although they are in principle just undersmoothing the MSE-optimal bandwidth, the result is very close to the CPE-optimal bandwidth for the sample sizes considered.

Joint CIs for all methods are computed using the Bonferroni approach. Uniform bands are

\(^8\)For inference on quantile marginal effects, Kaplan (2014) found bootstrap to be more accurate than estimating analytic standard errors, and cubic more accurate than linear or quadratic.

\(^9\)This differs somewhat from the description in the text, most notably by an additional factor of 0.4 in the bandwidth.
also examined, with L-stat, QY, and the local cubic bootstrap relying on the adjusted critical value from the Hotelling (1939) tube computations in \texttt{plot.rqss}. Additional implementation details may be seen in the available simulation code.

Our first conditional simulations use Model 1 from Fan and Liu (2015), with $n = 500$ and $p = 0.5$. Their “Direct” method in Table 1 is the same as our FLb, and the pointwise CPs match what they report, as seen in the top left graph of Figure 5. No method has more than slight under-coverage; rqss and QYg are often near 100%. The top right of Figure 5 shows pointwise power, where L-stat is always highest.

The bottom left graph in Figure 5 shows power curves of the hypothesis tests corresponding to the joint (over $x_0 \in \{0, 0.75, 1.5\}$) CIs. We consider tests against deviations of the same magnitude at each $x_0$. This magnitude is shown on the graph’s horizontal axis; the rejection probability at zero is the type I error rate. All methods have good type I error rates: L-stat’s is 6.2%, and other methods’ are below the nominal 5%. L-stat has significantly better power, an advantage of 20–40% at the larger deviations. The bottom right graph in Figure 5 is similar, but based on uniform confidence bands evaluated at 231 values of $x_0$. L-stat is the only method with close to exact size and good power.

Our next simulations use the simulation setup of the rqss vignette in Koenker (2012), which in turn was taken in part from Ruppert et al. (2003 §17.5.1). The main parameters are $n = 400$, $p = 0.5$, $d = 1$, and $\alpha = 0.05$, and

\begin{align}
X_i \overset{iid}{\sim} \text{Unif}(0, 1), \\
Y_i = \sqrt{X_i(1 - X_i)} \sin \left( \frac{2\pi(1 + 2^{-7/5})}{X_i + 2^{-7/5}} \right) + \sigma(X_i)U_i, 
\end{align}

where the $U_i$ are iid Gaussian, $t_3$, Cauchy, or centered $\chi^2_3$, and $\sigma(X) = 0.2$ or $\sigma(X) = 0.2(1 + X)$. The conditional median function is shown in Figure 6. Although the function as a whole is not a common shape in economics, it provides insight into different types of functions at different points. In particular, it has larger higher-order derivatives when $x_0$ is closer to zero, while it becomes relatively smooth as $x_0$ nears one. We can also see
Figure 5: Results from simulation DGP in Fan and Liu (2015), Model 1, $n = 500$, $p = 0.5$. Top left: pointwise CP at $x_0 \in \{0, 0.75, 1.5\}$, interpolated for visual ease. Top right: pointwise power at the same $x_0$ against deviations of $\pm 0.1$. Bottom left: joint power curves. Bottom right: uniform power curves.
performance at local minima, local maxima, and in between. For pointwise and joint CIs, we consider 47 equispaced points, \( x_0 = 0.04, 0.06, \ldots, 0.96 \); uniform confidence bands are evaluated at 231 equispaced values of \( x_0 \). Each simulation has 1,000 replications unless otherwise noted.

Across all eight DGPs (four error distributions, homoskedastic or heteroskedastic), L-stat has consistently accurate pointwise coverage, as shown in the first two columns of Figure 7. At the most challenging points (smallest \( x_0 \)), L-stat can under-cover by around five percentage points. Otherwise, CP is very close to \( 1 - \alpha \) for all \( x_0 \), all distributions, and in the presence or absence of heteroskedasticity.

In contrast, with the exception of boot, the other methods are subject to significant under-coverage in certain cases. As seen in the first two columns of Figure 7, the rqss under-coverage (as low as 50–60% CP depending on the error distribution) is quite significant for \( x_0 \) closer to zero. QYg has under-coverage with the non-normal distributions, especially the Cauchy. FLb has good CP except with the Cauchy, where CP can dip below 70%.

Joint power curves over the 47 values of \( x_0 \) are given in the third column of Figure 7. The horizontal axis of the graphs indicates the deviation of the null hypothesis from the true
Figure 7: Pointwise coverage probabilities by $X$ (first two columns) and joint power curves (third column), for conditional median 95% confidence intervals, $n = 400$, $p = 0.5$, DGP from (19). Distributions of $U_i$ are, top row to bottom row: $N(0, 1)$, $t_3$, Cauchy, and centered $\chi^2_3$. Columns 1 & 3: $\sigma(x) = 0.2$. Column 2: $\sigma(x) = (0.2)(1 + x)$. 38
curve; for example, $-0.1$ refers to a test against a curve lying $0.1$ below the true curve (at all $X$), and zero means the null hypothesis is true. Heteroskedastic versions were similar (and thus not shown). Our method’s size is close to nominal $\alpha = 5\%$ under all four error distributions ($5.7\%, 5.8\%, 7.3\%, 6.3\%)$. In contrast, other methods are size distorted for the Cauchy and/or $\chi^2_3$ conditional distributions; among them, boot is closest to controlling size. The rqss size is next-best; size distortion for FLb and QYg is more serious.

Column 3 of Figure 7 shows that L-stat has the steepest joint power curves. Although boot is the best of the rest, in the top graph (normal $U_i$), L-stat has better size control ($5.7\%$ vs. $6.8\%$) and better power against most deviations, due to its steeper power curve. Even with the Cauchy (third row), where FLb and QYg are quite size distorted, L-stat has better power than them against some alternatives.

Importantly, our method’s joint power curves in Column 3 of Figure 7 are also the most robust to the underlying distribution. L-stat’s size is near $5\%$ for all four distributions. In contrast, boot’s size ranges from only $1.2\%$ for the Cauchy, leading to worse power, up to $10.3\%$ for the $\chi^2_3$ in the fourth graph. Also, unlike L-stat, the local cubic test shows bias in the fourth graph: the probability of rejecting the truth is greater than the probability of rejecting a $-0.05$ deviation.

We also compared uniform confidence bands, where L-stat, boot, and QYg are ad hoc bands based on the rqss uniform critical values. The results are quite similar to the joint power curves (and thus relegated to the supplemental appendix), but with slightly higher rejection rates all around. As with the joint power curves, the L-stat uniform confidence bands are more robust to different distributions of $U_i$, usually maintaining size closer to $\alpha$ and steeper, unbiased power curves.

Figure 8 shows pointwise power, by $X$, against points differing from the true conditional median by $\pm 0.1$. This is calculated as the percentage of simulation replications where a method’s CI excluded the point $0.1$ below the true conditional median, averaged with the value for the point $0.1$ above. L-stat generally has the best power among methods with
Figure 8: Pointwise power by $X$, against deviations of magnitude 0.1 (negative or positive w.p. 1/2 each), for conditional median 95% confidence intervals, $n = 400$, $p = 0.5$, DGP from (19), $\sigma(x) = 0.2$. The $U_i$ are $N(0,1)$ (top left), $t_3$ (top right), Cauchy (bottom left), and centered $\chi^2_3$ (bottom right).
correct coverage.

Figure 9: Same as Figures 7 and 8 but with $p = 0.25$ and homoskedastic normal $U_i$.

L-stat continues to perform well even with $p = 0.25$ instead of $p = 0.5$ (still with $n = 400$, $\alpha = 0.05$). We show this in Figure 9 with normal $U_i$ centered to have $P(U_i < 0) = p$. Pointwise CP (left panel) is similar to $p = 0.5$, except FLb is somewhat worse. At $x_0$ where all methods have good coverage, pointwise power (middle panel) is generally highest for L-stat, although QYg is highest at points (but near zero at others). The L-stat joint CIs are the only ones to achieve correct coverage, as seen in the right panel; next-closest is boot with 86.2% CP. The joint L-stat test is also least biased. Due to its equal-tailed nature, the L-stat power curve is less steep for negative deviations (where the data are sparser) but still roughly parallel to QYg and boot, and it is more steep for positive deviations.

Figure 10 shows $p = 0.25$ and (recentered) Cauchy errors. Only boot maintains correct pointwise and joint CP. L-stat has some under-coverage, but its CP only falls below 93% at 10 of 47 points, and the lowest CP is still above 88%. In contrast, rqss, FLb, and QYg have severe under-coverage at many $x_0$. The cost of boot’s conservative coverage is pointwise power (middle panel) near 5% everywhere, compared with around 20% for L-stat. For joint testing (right panel), L-stat and boot share some slight (below 10% rejection rates) size distortion and bias, while other methods are quite size distorted and/or biased. In terms of power, L-stat is better than boot for positive deviations and vice-versa for negative. However,
the difference is more dramatic for positive deviations. Although L-stat has only 38% power against the −0.2 deviation where boot has 80% power, boot has almost trivial 8.5% power against the +0.15 deviation where L-stat has 95.6% power.

Table 8: Computation time in seconds (rounded), including bandwidth selection. $X_i$ and $Y_i$ both iid Unif(0, 1), $X_i \perp \perp Y_i$, $p = 0.5$; $n$ and number of $x_0$ (spaced between lower and upper quartiles of $X$) shown in table. Run on 3.2GHz Intel i5 processor with 8GB RAM.

<table>
<thead>
<tr>
<th>Method</th>
<th>$#x_0$</th>
<th>400</th>
<th>1000</th>
<th>4000</th>
<th>$10^4$</th>
<th>$10^5$</th>
<th>$10^6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-stat</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Local cubic</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>276</td>
<td></td>
</tr>
<tr>
<td>rqss</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>191</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>L-stat</td>
<td>100</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>Local cubic</td>
<td>100</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>141</td>
<td>4501</td>
</tr>
<tr>
<td>rqss</td>
<td>100</td>
<td>1</td>
<td>5</td>
<td>189</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>L-stat</td>
<td>1000</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>14</td>
<td>86</td>
<td>969</td>
</tr>
<tr>
<td>Local cubic</td>
<td>1000</td>
<td>11</td>
<td>15</td>
<td>37</td>
<td>80</td>
<td>1169</td>
<td>31994</td>
</tr>
<tr>
<td>rqss</td>
<td>1000</td>
<td>1</td>
<td>5</td>
<td>191</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Our method also has a computational advantage. While rqss is the fastest with $n = 400$ and scales best with the number of $x_0$, it slows to taking a few minutes with $n = 4000$. On the machine used, there was not enough memory to compute rqss for $n \geq 10^4$ (tried to allocate 9.9Gb, but only 8 available). The local cubic bootstrap scales worse with the number of $x_0$ than rqss, but scales better with $n$. By only considering local samples of size
it also scales much better than bootstrapping a global estimator using all \( n \) observations. Still, the reliance on resampling makes it slower than L-stat. For example, to examine 100 values of \( x_0 \) when \( n = 10^5 \), the local cubic bootstrap takes over two minutes, whereas L-stat takes under 10 seconds. See Table 8 for other examples.

Overall, the simulation results show the new L-stat method to be fast and accurate. Beside L-stat, the only method to avoid serious under-coverage is the local cubic with bootstrapped standard errors, perhaps due to its reliance on our newly proposed CPE-optimal bandwidth. However, L-stat consistently has better power, more robustness to different conditional distribution shapes, and less bias of its joint hypothesis tests.

7 Conclusion

We derive a uniform \( O(n^{-1}) \) difference between the linearly interpolated and ideal fractional order statistic distributions. We generalize this to \( L \)-statistics to help justify inference procedures based thereon. In particular, this translates to \( O(n^{-1}) \) CPE for the quantile confidence intervals proposed in Hutson (1999), and we provide analytic calibration to \( O(n^{-3/2} \log(n)) \) CPE. We extend these results to a nonparametric conditional quantile model, with both theoretical and Monte Carlo success. The derivation of an optimal bandwidth value (not just rate) and a fast approximation thereof are important practical advantages.


Censored and/or missing data may be accounted for, too. For example, if \( Y \) is top-coded, but the upper CI endpoint does not rely on a top-coded \( Y_i \), then the CI is valid as-is. If some
$Y_i$ are missing and the support of $Y_i$ is $[Y_L, Y_U]$, the most robust CI considers the extremes where all missing $Y_i$ are either $Y_L$ or $Y_U$. A valid two-sided $1 - \alpha$ CI takes its lower endpoint from a one-sided $(1 - \alpha_1)$ CI in the first case and its upper endpoint from a one-sided $(1 - \alpha_1)$ CI in the second case, where $\alpha_1 \in [\alpha/2, \alpha]$ must be determined in a manner similar to that for normal-based inference under partial identification [Imbens and Manski, 2004; Stoye, 2009].

Future work giving more careful consideration to the possibility of smoothing over discrete conditioning variables, as advocated by Li and Racine (2007), for example, may further improve finite-sample performance. So may consideration of the norm used to define $C_h$ when $d > 1$. It’s possible that other methods (like those in the simulations) may benefit in finite samples from a CPE-optimal bandwidth like we implement for our method.

References


Pearson, K. (1933). On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika*, 25:379–410.


A Mathematical proofs and additional lemmas

Additional intermediate steps and explanations are in the supplemental appendix.

Proof of Theorem 1

Proof. For any \( u \), let \( k = \lfloor (n + 1)u \rfloor \) and \( \epsilon = (n + 1)u - k \in [0, 1) \). If \( \epsilon = 0 \), then the objects \( \hat{Q}_X^L(u) \), \( \hat{Q}_X^U(u) \), and \( F^{-1}(\hat{Q}_X^U(u)) \) are identical and equal to \( X_{n:k} \). Otherwise, each lies in between \( X_{n:k} \) and \( X_{n:k+1} \) due to monotonicity of the quantile function and \( k/(n + 1) \leq u < (k + 1)/(n + 1) \):

\[
X_{n:k} = \hat{Q}_X^L(k/(n + 1)) \leq \hat{Q}_X^L(u) \leq \hat{Q}_X^L((k + 1)/(n + 1)) = X_{n:k+1},
\]

\[
X_{n:k} \leq \hat{Q}_X^L(u) = (1 - \epsilon)X_{n:k} + \epsilon X_{n:k+1} \leq X_{n:k+1},
\]

\[
X_{n:k} = F^{-1}(\hat{Q}_X^L(k/(n + 1))) \leq F^{-1}(\hat{Q}_X^L(u)) \leq F^{-1}(\hat{Q}_X^L((k + 1)/(n + 1))) = X_{n:k+1}.
\]

Thus, \( |\hat{Q}_X^L(u) - \hat{Q}_X^L(u)| \leq X_{n:k+1} - X_{n:k} \) and \( |\hat{Q}_X^L(u) - F^{-1}(\hat{Q}_X^L(u))| \leq X_{n:k+1} - X_{n:k} \), so

\[
\sup_{u \in [1/(n + 1), n/(n + 1)]} |\hat{Q}_X^L(u) - \hat{Q}_X^L(u)| \leq \max_{k \in \{1, \ldots, n-1\}} X_{n:k+1} - X_{n:k}, \quad (20)
\]

and similarly for \( |\hat{Q}_X^L(u) - F^{-1}(\hat{Q}_X^L(u))| \).

Taking a mean value expansion,

\[
X_{n:k+1} - X_{n:k} = F^{-1}(U_{n:k+1}) - F^{-1}(U_{n:k}) = (U_{n:k+1} - U_{n:k})/f(F^{-1}(\tilde{u}_k)), \quad (21)
\]

where \( \tilde{u}_k \in [U_{n:k}, U_{n:k+1}] \).

From (3.7) in Bickel (1967), for arbitrarily small \( \eta > 0 \),

\[
P\left[ \max_{k=1, \ldots, n} |U_{n:k} - k/(n + 1)| \geq \eta \right] \to 0, \quad (22)
\]
and since \((k+1)/(n+1) - k/(n+1) = 1/(n+1)\),
\[
P\left[ \sup_{u \in [1/(n+1),n/(n+1)]} \left| u - \tilde{u}_{[(n+1)u]} \right| \geq \eta \right] \to 0.
\]

Combined with the definition of \(U^δ\) in the statement of Theorem \(^\text{10}^\text{10}\), this implies that
\[
P\left( \max_{k=1,\ldots,n-1} f(F^{-1}(\tilde{u}_k)) \leq \delta \right) \to 1.
\]

Given \(\max_{k=1,\ldots,n-1} f(F^{-1}(\tilde{u}_k)) \leq \delta\), we can factor out \(\delta\). Using \[^20^\text{20}^\text{20}\) and \(^\text{21}^\text{21}\), and since \(U^δ_n \subseteq [1/(n+1), n/(n+1)]\),
\[
\sup_{\tilde{u} \in U^δ_n} \left| \tilde{Q}_X^L(u) - \tilde{Q}_X^L(u) \right| \leq \max_{k \in \{1,\ldots,n-1\}} X_{n:k+1} - X_{n:k} \leq \delta^{-1} \max_{k \in \{1,\ldots,n-1\}} (U_{n:k+1} - U_{n:k}).
\]

The marginal distribution of each spacing is \((U_{n:k+1} - U_{n:k}) \sim \beta(1, n)\), which is equivalent to a Kumaraswamy \((1980)\) distribution with the same parameters and CDF \(F(x) = 1 - (1 - x)^n\). With \(a_n\) such that \(\lim_{n \to \infty} a_n/(n^{-1} \log n) = \infty\),
\[
P\left( \sup_{\tilde{u} \in U^δ_n} \left| \tilde{Q}_X^L(u) - \tilde{Q}_X^L(u) \right| > a_n \right) \leq nP(\beta(1, n) > \delta a_n) = n \max\{(1 - \delta a_n)^n, 0\}
\]
\[
\leq \exp\{\log(n) - n\delta a_n\} \to 0.
\]

**Lemma for PDF approximation**

**Lemma 7.** Let \(\Delta k\) be a positive \((J + 1)\)-vector of natural numbers such that \(\sum_{j=1}^{J+1} \Delta k_j = n' \equiv n + 1\) and \(\min_j \{\Delta k_j\} \to \infty\) and define \(k_j \equiv \sum_{i=1}^{j} \Delta k_i\). Let \(X \equiv (X_1, \ldots, X_J)^t\) be the random \(J\)-vector such that
\[
\Delta X \equiv (X_1, X_2 - X_1, \ldots, 1 - X_J)^t \sim \text{Dirichlet}(\Delta k).
\]

Take any sequence \(a_n\) that satisfies conditions a) \(a_n \to \infty\), b) \(a_n^{-1} [\max \{\Delta k_j\}]^{1/2} \to 0\), and c) \(a_n [\min \{\Delta k_j\}]^{-1/2} \to 0\). Define Condition \(\ast(a_n)\) as satisfied by any vector \(\mathbf{x}\) iff
\[
\sup_{j} \left\{ n\Delta k_j^{-1/2} |\Delta x_j - \Delta k_j/n'| \right\} \leq a_n.
\]

Note that the implied rates are the same when centering at the mode of \(\Delta X_j\); \((\Delta k_j - 1)/(n - J)\), instead of at its mean, \(\Delta k_j/n'\).

\(^{10}\)As written, \(U^δ\) cannot contain the whole unit interval \((0, 1)\) if \(X\) has an unbounded distribution. However, the following equation could be modified to allow \(f(F^{-1}(u)) \to \infty\) at some rate as \(u \to 0\), and slowing \(U^δ_n\) to a subset of \([n^{-a}, 1 - n^{-a}]\) for \(a < 1\), at the expense of a larger right-hand side in the statement of the theorem. We omit this since it is not helpful for our goal of high-order accurate confidence intervals.
(i) At any point of evaluation \( \Delta x \) satisfying Condition \( \star(a_n) \), the log Dirichlet PDF of \( \Delta X \) may be uniformly approximated as

\[
\log f_{\Delta x}(\Delta x) = K - \frac{(n - J)^2}{2} \sum_{j=1}^{J+1} \left( \frac{\Delta x_j - \frac{\Delta k_j - 1}{n-j}}{\Delta k_j - 1} \right)^2 + \rho,
\]

where the constant

\[
K \equiv \frac{J}{2} \log(n/2\pi) + \frac{1}{2} \sum_{j=1}^{J+1} \log \left( \frac{n}{\Delta k_j - 1} \right),
\]

\( E(\rho^2) = O(\|\Delta k^{-1}\|_\infty) \), and \( \rho = O(a_n^2\|\Delta k^{-1/2}\|_\infty) \) uniformly. We also have the uniform approximations

\[
\frac{\partial \log[f_{\Delta x}(\Delta x)]}{\partial \Delta x_j} = (n - J) - \frac{(n - J)^2}{2} \left( \frac{\Delta x_j - \frac{\Delta k_j - 1}{n-j}}{\Delta k_j - 1} \right) + O(a_n^2\|\Delta k^{-1}\|_\infty),
\]

\[
\frac{\partial \log[f_{\Delta x}(\Delta x)]}{\partial \Delta k_j/n} = - \frac{\partial \log[f_{\Delta x}(\Delta x)]}{\partial \Delta x_j} + O(\|\Delta k^{-1}\|_\infty),
\]

where the superscript in \( X^k \) emphasizes the dependence of the density on the underlying vector of Dirichlet parameters.

(ii) For all \( x \in \mathbb{R}^J \) satisfying Condition \( \star(a_n) \), uniformly:

\[
\log[f_X(x)] = K + \frac{1}{2}(x - k/n')'H(x - k/n') + O(a_n\|\Delta k^{-1/2}\|_\infty),
\]

\[
\frac{\partial \log[f_X(x)]}{\partial x} = H(x - k/n') + O(a_n^2\|\Delta k^{-1}\|_\infty),
\]

\[
\frac{\partial \log[f_X(x)]}{\partial k/n'} = -H(x - k/n') + O(a_n^2\|\Delta k^{-1}\|_\infty),
\]

where the constant \( K \) is the same as in part (i), and the \( J \times J \) matrix \( H \) has non-zero elements only on the diagonal \( H_{ij} = -n^2(\Delta k_j^{-1} + \Delta k_{j+1}^{-1}) \) and one off the diagonal \( H_{i,j+1} = n^2\Delta k_{j+1}^{-1} \).

(iii) For the Dirichlet-distributed \( \Delta X \), \( 1 - P[\star(a_n)] = o(\exp\{-a_n^2/2\}) \).

(iv) If instead there are asymptotically fixed components of the parameter vector, the largest of which is \( \Delta k_j = M < \infty \), then \( 1 - P[\star(a_n)] = O(\exp\{-a_nM^{-1/2}\}) \).

It can be checked that \( V \equiv H^{-1} \) is such that

\[
V_{i,j} = \min(k_i, k_j)(n'-\max(k_i, k_j)), \tag{24}
\]

connecting the above result and conventional asymptotic normality results for quantiles.
Proof of Lemma 7(i)

Proof. Since $\Delta x \sim \text{Dirichlet}(\Delta k)$, for any $\Delta x$ that sums to one,

$$
\log(f_{\Delta x}(\Delta x)) = \log(\Gamma(n+1)) + \sum_{j=1}^{J+1} \left[ (\Delta k_j - 1) \log(\Delta x_j) - \log(\Gamma(\Delta k_j)) \right].
$$

Applying the Stirling-type bounds in Robbins (1955) to the gamma functions,

$$
\log(f_{\Delta x}(\Delta x)) = \frac{J}{2} \log(n/2\pi) + \frac{1}{2} \sum_{j=1}^{J+1} \log \left( \frac{n}{\Delta k_j - 1} \right) + \sum_{j=1}^{J+1} (\Delta k_j - 1) \log \left( \frac{n \Delta x_j}{\Delta k_j - 1} \right) - J
$$

$$
+ O(||\Delta k^{-1}||_\infty),
$$

where $K$ is the same constant as in the statement of the lemma. The remainder does not depend on $\Delta x$.

We then expand $h(\cdot)$ around the Dirichlet mode, $\Delta x_0$. The cross partials are zero, the first derivative term zeroes out, and the fourth derivative is smaller-order, yielding

$$
h(\Delta x) = h(\Delta x_0) + \sum_{j=1}^{J+1} h_j(\Delta x_0)(\Delta x_j - \Delta x_{0j}) + \frac{1}{2} \sum_{j=1}^{J+1} h_{j,j}(\Delta x_0)(\Delta x_j - \Delta x_{0j})^2
$$

$$
+ \frac{1}{6} \sum_{j=1}^{J+1} h_{j,j,j}(\Delta x_0)(\Delta x_j - \Delta x_{0j})^3 + \frac{1}{24} \sum_{j=1}^{J+1} h_{j,j,j,j}(\Delta \bar{x})(\Delta x_j - \Delta x_{0j})^4
$$

$$
= O(n^{-1}) + 0 - \frac{(n - J)^2}{2} \sum_{j=1}^{J+1} \left( \frac{\Delta x_j - \Delta k_j - 1}{n - J} \right)^2 + \frac{(n - J)^3}{3} \sum_{j=1}^{J+1} \frac{(\Delta x_j - \Delta k_j - 1)^3}{(\Delta k_j - 1)^2}
$$

$$
+ O(a^4_n||\Delta k^{-1}||_\infty),
$$

where Condition $\ast(a_n)$ bounds the cubic term at $O(a_n^3||\Delta k^{-1/2}||_\infty)$ and bounds the quartic term as shown.

When we square the cubic term to derive the mean square bound, we can drop a few smaller-order adjustments, leaving

$$
n^6 \sum_{j=1}^{J+1} \Delta k_j^{-2} \Delta x_j - E(\Delta x_j))^3(\Delta x_i - E(\Delta x_i))^3.
$$

The order of the expectation then depends on the sixth central moment and squared third central moment of $\Delta x_j \sim \beta(\Delta k_j, n + 1 - \Delta k_j)$. The third central moment is $O(\Delta k_j n^{-3})$, so (27) is $O(||\Delta k^{-2}||_\infty)$, even smaller than the $O(||\Delta k^{-1}||_\infty)$ required. The sixth central moment is $O(n^{-11} \Delta k_j^3 n^5) = O(\Delta k_j^6 n^{-6})$. The sum of the fixed number of terms of order $O(n^6 \Delta k_j^{-4} \Delta k_j^2 n^{-6}) = O(\Delta k_j^{-1})$ is $O(||\Delta k^{-1}||_\infty)$ as desired. This completes the proof for the density.
The final two results are derivatives of \((26)\), which are simplified using Condition \(*(a_n)\) and the Dirichlet property \(\|\Delta x - \Delta x_0\|_1 = 0\).

**Proof of Lemma 7(ii)**

Proof. The density for \(X\) may be obtained from the Lemma 7(i) density for \(\Delta x\) simply by plugging in \(\Delta x_j = x_j - x_{j-1}\) since the transformation is unimodular. Centering at the mean instead of the mode introduces only smaller-order error, as does replacing \(n - J\) with \(n\) and \(\Delta k_j - 1\) with \(\Delta k_j\), and the constant \(K\) does not change. The other results are derivatives of this density for \(X\), accounting for the effect of \(x_j\) on both \(\Delta x_j\) and \(\Delta x_{j+1}\), and similarly for \(k_j\).

**Proof of Lemma 7(iii)**

Proof. Define \((l, u) \equiv (\lceil \Delta k_j - a_n\Delta k_j^{1/2} \rceil / n, \lfloor \Delta k_j + a_n\Delta k_j^{1/2} \rfloor / n\). Condition \(*(a_n)\) is violated if \(\Delta X_j \notin (l, u)\) for some \(j\). The marginal distribution of \(\Delta X_j\) is \((\Delta X_j, 1 - \Delta X_j) \sim \text{Dirichlet}(\Delta k_j, n + 1 - \Delta k_j)\). Using the log PDF from Lemma 7(i),

\[
P(\Delta X_j > u) = \int_u^1 f_{\Delta X_j}(t, 1 - t)dt = e^K \int_u^1 \exp\{h(t, 1 - t)\}dt
\]

\[
< e^K \int_u^1 \exp\{h(u, 1 - u) + h'(u, 1 - u)(t - u)\}dt
\]

\[
= e^{K + h(u, 1 - u)} \int_u^1 \exp\{h'(u, 1 - u)(t - u)\}dt
\]

because \(h''(\cdot) < 0\) as seen in the supplemental appendix. At the upper endpoint \(u\), using \((26)\),

\[
h(u, 1 - u) = -(1/2)n^2[\Delta k_j^{-1}(u - \Delta k_j/n)^2 + (n - \Delta k_j)^{-1}(1 - u - (n - \Delta k_j)/n)^2] + O(a_n^3\Delta k_j^{-1/2})
\]

\[
< -a_n^2/2 + O(a_n^3\Delta k_j^{-1/2}),
\]

\[
\frac{\partial h(u, 1 - u)}{\partial u} = \frac{-a_n n^2}{\Delta k_j^{1/2}(n - \Delta k_j)} + O(a_n^2 n \Delta k_j^{-1}),
\]

the dominant term of which is negative. Also,

\[
K - \log[-h'(u, 1 - u)]
\]

\[
= \log(n)/2 - \log(2\pi)/2 + \log(n) - \log(\Delta k_j - 1)/2 - \log(n - \Delta k_j)/2
\]

\[
- \log(a_n) - 2 \log(n) + \log(\Delta k_j)/2 + \log(n - \Delta k_j)
\]

\[
< -\log(2\pi)/2 - \log(a_n) + O(\Delta k_j^{-1}).
\]
Thus,
\[
e^{K + h(u, 1 - u)} \int_u^1 \exp \{ h'(u, 1 - u)(t - u) \} dt < e^{K + h(u, 1 - u)} \int_u^\infty \exp \{ h'(u, 1 - u)(t - u) \} dt < \exp \{ - \log(2\pi) - \log(a_n) + O(\Delta k_j^{-1}) \} \exp \{ -a_n^2/2 + O(a_n^3\Delta k_j^{-1/2}) \}
\]
\[
\simeq \frac{1}{2\pi} a_n^{-1} e^{-a_n^2/2} = o(\exp \{ -a_n^2/2 \}).
\]
Calculations for the lower endpoint are similar and yield the same rate. There exist \( J + 1 \) spacings that can generate violations of Condition \( \ast(a_n) \); summing over them does not change the rate since \( J < \infty \) does not increase with \( n \).

Proof of Lemma 7(iv)

Proof. Since \( \Delta k_j = M < \infty \) is a fixed natural number, we can write \( \Delta X_j = \sum_{i=1}^M \delta_i \), where each \( \delta_i \) is a spacing between consecutive uniform order statistics. The marginal distribution of each \( \delta_i \) is \( \beta(1, n) \), with closed form CDF (Kumaraswamy 1980)
\[
F_{\delta_i}(d) = 1 - (1 - d)^n \text{ for } 0 \leq d \leq 1.
\]
Using this,
\[
P(\Delta X_j > a_n n^{-1} M^{1/2}) \leq P \left( \exists i : \delta_i > \frac{a_n n^{-1} M^{1/2}}{M} \right) \leq M P(\delta_i > a_n n^{-1} M^{-1/2}) = M \left( 1 - a_n n^{-1} M^{-1/2} \right)^n \leq M \exp \{ -a_n M^{-1/2} \}.
\]
If there are multiple fixed spacings in the parameter vector, this argument can be repeated and only the largest (fixed) spacing will enter the asymptotic rate.

Lemma for proving Theorem 2

First, we introduce notation. From earlier, \( u_0 \equiv 0 \) and \( u_{J+1} \equiv 1 \). For all \( j \in \{1, 2, \ldots, J\} \),
\[
k_j \equiv [(n + 1) u_j] \quad \epsilon_j \equiv (n + 1) u_j - k_j,
\]
where the \( \epsilon_j \in (0, 1) \) are interpolation weights as in \( \square \). Also let \( \Delta k \) denote the \( (J+1) \)-vector such that \( \Delta k_j = k_j - k_{j-1} \), and
\[
Y_j \equiv U_{n: k_j} \sim \beta(k_j, n + 1 - k_j) \quad \Delta Y \equiv (Y_1, Y_2 - Y_1, \ldots, 1 - Y_J) \sim \text{Dirichlet}(\Delta k) \quad (28)
\]
\[
\Lambda_j \equiv U_{n: k_{j+1}} - U_{n: k_j} \sim \beta(1, n) \quad Z_j \equiv \sqrt{n}(Y_j - u_j) \quad V_j \equiv \sqrt{n}[F^{-1}(Y_j) - F^{-1}(u_j)]
\]
52
$X \equiv \sum_{j=1}^{J} \psi_j F^{-1}(Y_j)$  
$X_0 \equiv \sum_{j=1}^{J} \psi_j F^{-1}(u_j)$

$\mathbb{W} \equiv \sqrt{n}(X - X_0) = \psi'V$  
$\mathbb{W}_{\epsilon, \Lambda} \equiv \mathbb{W} + n^{1/2} \sum_{j=1}^{J} \epsilon_j \psi_j \Lambda_j [Q'(u_j) + Q''(u_j)(Y_j - u_j)]$

$\hat{u} \equiv \left\{ u_j^{(H)}(\hat{\gamma}(\cdot)) \right\}_{j=1}^{J}$  
$u_0 \equiv \left\{ u_j^{(H)}(\gamma(0)) \right\}_{j=1}^{J}$

where vectors are indicated by bold and matrices (later) by underline. Adding a $\hat{\cdot}$ superscript to any of these objects, like $\mathbb{W}^{\hat{u}}$, indicates having $\hat{\cdot}$ as the vector of quantiles of interest (rather than the original $\cdot$). The values and distributions of $k, Y, \Lambda, V, Z, X,$ and $\mathbb{W}$ are all understood to vary with $n$. We may use a superscript to emphasize that these quantities depend on the chosen vector of quantiles (e.g., $\mathbb{W}^{u_0}$ vs. $\mathbb{W}^{\hat{u}}$).

**Lemma 8.** Let Assumption A2 hold at $\hat{\cdot}$, and let each element of $Y$ and $\Lambda$ satisfy Condition $(2 \log(n))$ as defined in Lemma 7. The following results hold uniformly over any $u = \bar{u} + o(1)$.

(i) The following bounds hold:

$$
|L^L - (X_0 + n^{-1/2}\mathbb{W}_{\epsilon, \Lambda})| = O(n^{-2} \log(n)^3),
$$

$$
|L^I - (X_0 + n^{-1/2}\mathbb{W}_{C, \Lambda})| = O(n^{-2} \log(n)^3),
$$

where $C$ is a $J$-vector of random interpolation coefficients as defined in Jones (2002). Each $C_j \sim \beta(\epsilon_j, 1 - \epsilon_j)$, and they are mutually independent.

(ii) Define $Y$ and $A$ as the $J \times J$ matrices such that $V_{i,j} = \min\{u_i, u_j\}(1 - \max\{u_i, u_j\})$ and $A = \text{diag}\{f(F^{-1}(u_i))\}$. Also define $V_\psi \equiv \psi'(A^{-1}YA^{-1})\psi \in \mathbb{R}$. For any vector $\lambda$ satisfying Condition $(2 \log(n))$,

$$
\sup_w \left| f_{\mathbb{W}_{\epsilon, \Lambda}}(w \mid \lambda) - \phi_{\mathbb{W}_{\lambda}}(w) \right| = O(n^{-1/2} \log(n)),
$$

$$
\sup_{w \in \mathbb{R}} \left\| \frac{\partial f_{\mathbb{W}_{\epsilon, \Lambda}}(w \mid \lambda)}{\partial w} - \frac{\partial \phi_{\mathbb{W}_{\lambda}}(w)}{\partial w} \right\|_\infty = O(n^{-1/2} \log(n)).
$$

For any value $\bar{\epsilon} \in [0, 1]^J$,

$$
\frac{\partial^2 F_{\mathbb{W}_{\epsilon, \Lambda}}(K \mid \lambda)}{\partial \epsilon_j^2} \bigg|_{\epsilon = \bar{\epsilon}} = n \psi_j^2 Q'(u_j)^2 \lambda_j^2 \left[ \frac{\partial \phi_{\mathbb{W}_{\lambda}}(w)}{\partial w} \right] + O(n^{-3/2} \log(n)^3).
$$

**Proof of Lemma 8(i)**

Proof. Since $J$ is fixed and finite, there exists $n_0$ such that for all $n > n_0$, all $u_j$ lie inside the neighborhoods in Assumption A2. Since the third derivative of the quantile function is $Q^{(3)}(u) = [3f'(Q(u))^2 - f(Q(u))f''(Q(u))]f(Q(u))^5$, having $f(Q(u_j))$ uniformly bounded away from zero and $f''(Q(u_j))$ uniformly bounded ensures that the quantile function derivatives do not affect the asymptotic rates of the remainder terms below.
With a Taylor expansion, the object \( L \) may be rewritten as

\[
L = X_0 + n^{-1/2} W_{\epsilon, \Lambda} + \sum_{j=1}^{J} \psi_j \epsilon_j [\nu_{j,1}^L + \nu_{j,2}^L],
\]

\[
\nu_{j,1}^L \equiv \frac{Q^{(3)}(\bar{u}_j)}{2} [Y_j - u_j]^2 \Lambda_j, \quad \nu_{j,2}^L \equiv \frac{Q''(\bar{y}_j)}{2} \Lambda_j^2,
\]

where \( \forall j, \bar{y}_j \in (Y_j, Y_j + \Lambda_j) \) and \( \bar{u}_j \) is between \( u_j \) and \( Y_j \). Applying \( A2 \) and Condition \( \star (2 \log(n)) \) gives the desired rate on the \( \nu^L \) terms. The argument is similar for \( L^I \) but also depends on the random, mutually independent interpolation weights \( C_j \sim \beta(\epsilon_j, 1 - \epsilon_j) \) (Jones, 2002) that replace \( \epsilon_j \):

\[
L^I = X_0 + n^{-1/2} W_{C, \Lambda} + \sum_{j=1}^{J} \psi_j [\nu_{j,1}^I + \nu_{j,2}^I],
\]

\[
\nu_{j,1}^I \equiv \frac{Q^{(3)}(\bar{u}_j)}{2} [Y_j - u_j]^2 C_j \Lambda_j, \quad \nu_{j,2}^I \equiv \frac{Q''(\bar{y}_j)}{2} (C_j \Lambda_j)^2.
\]

### Proof of Lemma 8(ii)

**Proof.** Since \( \Lambda \) contains finite spacings, we cannot apply Lemma 7(ii). Instead, we use

\[
\Lambda_j \sim \beta(1, n), \quad \frac{\Lambda_j}{1 - \sum_{k=1}^{j-1} \Lambda_k} \left( \Lambda_1, \ldots, \Lambda_{j-1} \right) \sim \beta(1, n - j + 1),
\]

\[
\log f_{\Lambda_j}(\lambda_j) = \log(n) - n \lambda_j + O(n^{-1} \log(n)),
\]

\[
\log f_{\Lambda_1(\Lambda_1, \ldots, \Lambda_{j-1})} = \log(n) - n \lambda_j + O(n^{-1} \log(n)),
\]

where the above rate follows from imposing Condition \( \star (2 \log(n)) \) on all values of \( \Lambda \). Taking appropriate products of conditional densities leads to

\[
\log f(\lambda) = J \log(n) - n \sum_{j=1}^{J} \lambda_j + O(n^{-1} \log(n)). \tag{29}
\]

For the joint density of \( \{Y, \Lambda\} \), define the linear transformation \( \tau : \{Y, \Lambda\} \to Y \Lambda \equiv \{\Delta Y_1, \Lambda_1, \Delta Y_2 - \Lambda_1, \ldots, \Lambda_J, \Delta Y_{J+1} - \Lambda_J\}' \). Row operations show this transformation to be unimodular. Now

\[
Y \Lambda \sim \text{Dirichlet}(\Delta k_1, 1, \Delta k_2 - 1, \ldots, 1, \Delta k_{J+1} - 1).
\]

Unfortunately, we still may not apply Lemma 7(ii). However,

\[
\log(f_{\{Y, \Lambda\}}(y, \lambda)) = \log(f_{Y \Lambda}(\tau(y, \lambda))) \cdot |\tau'|^{-1}
\]

\[
= \log(\Gamma(n + 1)) + (\Delta k_1 - 1) \log(\Delta y_1) - \log(\Gamma[\Delta k_1]) \tag{30}
\]

54
Combining (29) and (30),

\[ + \sum_{j=2}^{J+1} \left( (\Delta k_j - 2) \log(\Delta y_j - \lambda_{j-1}) - \log(\Gamma[\Delta k_j - 1]) \right). \]

Combining (29) and (30),

\[ \log f_{Y|\Lambda} = \log(f_{Y,\Lambda}(y, \lambda)) - \log f_\Lambda(\lambda) \]
\[ = \log f_Y(y) + \sum_{j=2}^{J+1} \left\{ (\Delta k_j - 2)[\log(\Delta y_j - \lambda_{j-1}) - \log(\Delta y_j)] + n\lambda_{j-1} \right\} \]
\[ + \left\{ [\log(\Gamma[\Delta k_j]) - \log(\Gamma[\Delta k_j - 1])] - \log n - \log \Delta y_j \right\} \]
\[ = \log \phi_{Y/n}(y - k/(n + 1)) + O(n^{-1/2}\log(n)), \]

where the final result follows from applying algebra and \(*(2\log(n))\) to the analysis of each braced term and applying Lemma 7(ii) directly to the density of \(Y\). Differentiating (31) and applying Lemma 7(iii) to the density derivative of \(Y\) yields

\[ \frac{\partial \log f_{Y|\Lambda}(y; \lambda)}{\partial y} = \frac{\partial \log \phi_{Y/n}(y - k/(n + 1))}{\partial y} + O(\log n). \]

For convenience, we define and calculate

\[ H_{Y_j} \equiv K - n^{1/2} \sum_{j=1}^{J} \epsilon_j \Lambda_j [Q'(u_j) + Q''(u_j)(Y_j - u_j) = K - O(n^{-1/2}\log(n)), \]
\[ \frac{\partial H_{Y_j}}{\partial \epsilon_j} = -n^{1/2} \Lambda_j [Q'(u_j) + Q''(u_j)(Y_j - u_j)] = -n^{1/2} \Lambda_j Q'(u_j) + O(n^{-1}\log(n)), \]
\[ \frac{\partial^2 H_{Y_j}}{\partial \epsilon_j^2} = 0. \]

Since the dependence of \(H\) on \(Y_j\) is smaller-order, we drop the subscript, and

\[ F_{W|\Lambda}(K|\lambda) = \int_{-\infty}^{H} f_{W|\Lambda}(w|\lambda)dw, \]
\[ \frac{\partial F_{W|\Lambda}(K|\lambda)}{\partial \epsilon_j} = f_{W|\Lambda}(H|\lambda) \frac{\partial H}{\partial \epsilon_j}, \]
\[ \frac{\partial^2 F_{W|\Lambda}(K|\lambda)}{\partial \epsilon_j \partial \epsilon_k} = f_{W|\Lambda}(H|\lambda) \left( \frac{\partial H}{\partial \epsilon_j} \right) \left( \frac{\partial H}{\partial \epsilon_k} \right) + f'_{W|\Lambda}(H|\lambda) \left( \frac{\partial H}{\partial \epsilon_j} \right) \left( \frac{\partial H}{\partial \epsilon_k} \right), \]
\[ \frac{\partial^2 F_{W|\Lambda}(K|\lambda)}{\partial \epsilon_j^2} = f'_{W|\Lambda}(H|\lambda) \left( \frac{\partial H}{\partial \epsilon_j} \right)^2 \]
\[ = \left[ f'_{W|\Lambda}(K|\lambda) + O(n^{-1/2}\log(n)) \right] \]
\[ \times \left[ n\psi_j^2 \lambda^2 Q'(u_j)^2 + O(n^{-1/2}\log(n)) \lambda_j Q'(u_j) + O(n^{-2}\log(n)^2) \right]. \]
The final remainder order comes from $Q'(u_j)$ being bounded (from Assumption A2) and $\lambda_j = O(n^{-1} \log(n))$ due to Condition $\star(2 \log(n))$. 

**Proof of Theorem 2(i)**

The following is brief intuition behind the proof. We start by restricting attention to cases where the largest spacing between relevant uniform order statistics, $U_{n:[(n+1)u_j]+1} - U_{n:[(n+1)u_j]}$, and the largest difference between the $U_{n:[(n+1)u_j]}$ and $u_j$ satisfy Condition $\star(2 \log(n))$ as in Lemma 7. By Lemma 7(iii,iv), the error from this restriction is smaller-order. We then use the representation of ideal uniform fractional order statistics from Jones [2002], which is equal in distribution to the linearly interpolated form but with random interpolation weights $C_j \sim \beta(\epsilon_j, 1 - \epsilon_j)$. The leading term in the error is due to $\text{Var}(C_j)$, and by plugging in other calculations from Lemma 8, we see that it is uniformly $O(n^{-1})$ and can be calculated analytically.

**Proof.** We assume that the realized values of $Y$ and $\lambda$ all satisfy Condition $\star(2 \log(n))$. By application of Lemma 7(iii,iv), this induces at most $O(n^{-2})$ error. This is asymptotically negligible, so we ignore it going forward.

Applying Lemma 8[i] to the CDF of $L^L$, 

$$P(L^L < X_0 + n^{-1/2}K) = P(\mathbb{W}_{\epsilon, \lambda} < K) + O(n^{-3/2} \log(n)^3),$$

where the error can be pulled out of the probability statement because $\mathbb{W}_{\epsilon, \lambda}$ has been shown to have a bounded PDF in Lemma 8[i]. The CDF of $L^L$ is then

$$= \int_{[0, 2n^{-1} \log(n)], J} P(\mathbb{W}_{\epsilon, \lambda} < K \mid \lambda) dF_{\lambda}(\lambda) + O(n^{-3/2} \log(n)^3).$$

By a similar series of manipulations,

$$P(L^I < X_0 + n^{-1/2}K) = \int_{[0, 2n^{-1} \log(n)], J} \int_{[0, 1], J} P(\mathbb{W}_{\epsilon, \lambda} < K \mid \lambda) dF_C(c) dF_{\lambda}(\lambda)$$

$$+ O(n^{-3/2} \log(n)^3).$$

The CDF difference between the two distributions is

$$P(L^I < X_0 + n^{-1/2}K) - P(L^L < X_0 + n^{-1/2}K)$$

$$= \int_{[0, 2n^{-1} \log(n)], J} \int_{[0, 1], J} [F_{W_{\epsilon, \lambda} \mid \lambda}(K \mid \lambda) - F_{W_{\epsilon, \lambda} \mid \lambda}(K \mid \lambda)] dF_C(c) dF_{\lambda}(\lambda) + O(n^{-3/2} \log(n)^3)$$

$$= \int_{[0, 2n^{-1} \log(n)], J} \int_{[0, 1], J} \left[ (c - \epsilon)^J dF_{W_{\epsilon, \lambda} \mid \lambda}(K \mid \lambda) \right] dc + \frac{1}{2} (c - \epsilon)^J dF_{W_{\epsilon, \lambda} \mid \lambda}(K \mid \lambda) dF_{\lambda}(\lambda) + O(n^{-3/2} \log(n)^3).$$
Since \( E(C) = \epsilon \), the first term zeroes out. Since the elements of \( C \) are mutually independent, the off-diagonal elements of the Hessian in the quadratic term also zero out. Then we apply Lemma 8(ii) to the Hessian and use \( \text{Var}(C_j) = \epsilon_j(1 - \epsilon_j)/2 \), leaving

\[
= \frac{n}{2} \left( \frac{d[\varphi_{\psi}(w)]}{dw} \right) \sum_{j=1}^{J} \left[ \psi_j'Q'(u_j) \right]^2 \frac{\epsilon_j(1 - \epsilon_j)}{2} \int_{[0,2n^{-1}\log(n)]^J} \lambda_j^2 dF_\Lambda(\lambda) + O(n^{-3/2} \log(n)^3) .
\]

The proof is completed by noting that second raw moments of \( \Lambda \) are such that \( \forall j, E(\Lambda_j^2) = 2/[2(n + 1)(n + 2)] = 2n^{-2} + O(n^{-3}) \) \cite{Kumaraswamy1980}.

\[\Box\]

**Proof of Theorem 2(ii)**

Proof. We again assume that the realized values of \( Y \) and \( \lambda \) satisfy Condition \(*2(2\log(n))\). By application of Lemma 8(iii), this induces at most \( O(n^{-2}) \) error in our calculations, which is asymptotically negligible.

The first result of this part is obtained by solving a first order condition and evaluating the expression obtained in part [i] at the solution \( K = \sqrt{\psi} \). The second result comes directly from the uniform bound on densities obtained in Lemma 8(iii). \[\Box\]

**Proof of Lemma 3**

Proof. For notation here, let \( p \) be the quantile of interest, and let \( u = u'(\alpha) \). From (5), \( u \) solves \( P(B > p) = \alpha \) for \( B \sim \beta((n+1),(1-u)(n+1)) \); equivalently, \( F_B(p) = 1 - \alpha \).

We use a Cornish–Fisher-type expansion from (5.1) in \cite{Pratt1968}. In his equation, \( \gamma = \frac{[p - (1 - p)]/\sigma = (2p - 1)/\sqrt{np(1-p)} = O(n^{-1/2})} \), so

\[
z = z_0 - \frac{\gamma}{6} (z_0^2 - 1) + O(n^{-1}),
\]

where \( z_0 \) and \( \gamma \) depend on \( n, p, \) and \( u_\ell \), and \( \Phi(z) = F_B(p) = 1 - \alpha \) so that \( z = z_{1-\alpha} \). Let

\[
p - u = n^{-1/2}c_1 + n^{-1}c_2 + O(n^{-3/2}).
\]

Using (2.2) from \cite{Pratt1968} and Table A from \cite{PeizerPratt1968},

\[
\begin{align*}
z_0 &= \frac{\sqrt{n(p-u)}}{\sqrt{p(1-p)}} - (1/2)n^{-1/2} \frac{2p-1}{\sqrt{p(1-p)}} + O(n^{-1}), \quad z_0^2 = \frac{n(p-u)^2}{p(1-p)} + O(n^{-1/2}), \\
z_{1-\alpha} &= \frac{\sqrt{n(p-u)}}{\sqrt{p(1-p)}} - (1/2)n^{-1/2} \frac{2p-1}{\sqrt{p(1-p)}} + O(n^{-1}) \\
&\quad - \frac{2p-1}{6\sqrt{np(1-p)}} \left( \frac{n(p-u)^2}{p(1-p)} + O(n^{-1/2}) - 1 \right) + O(n^{-1}) \\
&= \frac{c_1}{\sqrt{p(1-p)}} + n^{-1/2} \frac{c_2}{\sqrt{p(1-p)}} - n^{-1/2} \frac{2p-1}{6\sqrt{p(1-p)}} \left( \frac{c_1^2}{p(1-p)} + 2 \right) + O(n^{-1}).
\end{align*}
\]

57
To solve for $c_1$, we can set equal the first-order terms on the left- and right-hand sides:

$$z_{1-\alpha} = \frac{c_1}{\sqrt{p(1-p)}} \Rightarrow c_1 = z_{1-\alpha}\sqrt{p(1-p)}.$$ 

Plugging in for $c_1$, we can solve for $c_2$ to zero the $n^{-1/2}$ terms:

$$0 = n^{-1/2} \left[ \frac{c_2}{\sqrt{p(1-p)}} - \frac{2p-1}{6\sqrt{p(1-p)}} \left( \frac{z_{1-\alpha}^2 p(1-p)}{p(1-p)} + 2 \right) \right] \Rightarrow c_2 = \frac{2p-1}{6}(z_{1-\alpha}^2 + 2).$$ 

Altogether,

$$u_\ell = p - \frac{c_1}{\sqrt{n}} - \frac{c_2}{n} + O(n^{-3/2}) = p - n^{-1/2}z_{1-\alpha}\sqrt{p(1-p)} - \frac{2p-1}{6n}(z_{1-\alpha}^2 + 2) + O(n^{-3/2}).$$

For $u_h$, we instead have $z_\alpha$, and using $-z_\alpha = z_{1-\alpha}$ yields the result.

\[\square\]

**Proof of Theorem 4**

**Coverage probability**

For the application of Theorem 3 shown in the text,

$$u^h(\alpha) = p + O(n^{-1/2}), \quad u^l(\alpha) = p - O(n^{-1/2}),$$

$$J = 1, \quad \epsilon_h = (n+1)u^h(\alpha) - [(n+1)u^h(\alpha)], \quad \epsilon_\ell = (n+1)u^l(\alpha) - [(n+1)u^l(\alpha)],$$

$$\Psi = \frac{u^h(\alpha)(1-u^h(\alpha))}{f(F^{-1}(p))^2}, \quad \Psi_0 = F^{-1}(u^h(\alpha)), \quad \text{and}$$

$$K = n^{1/2}[F^{-1}(p) - F^{-1}(u^h(\alpha))] = -\frac{z_{1-\alpha}\sqrt{p(1-p)}}{f(F^{-1}(p))} + O(n^{-1/2}),$$

where the first and last lines use Lemma 3, and the last line uses Assumption A2.

**Power**

Consider one-sided power against $D_n \equiv Q(p) + Kn^{-1/2}$ with $K > 0$. Type II error occurs if the lower one-sided CI contains $D_n$. Using Theorem 2 and Lemma 3, power is

$$\mathcal{P}_n^l(D_n) = P \left( D_n \notin (-\infty, \hat{Q}_X^L(u^h(\alpha))) \right)$$

$$= P \left( F^{-1}(\hat{Q}_U^L(u^h(\alpha))) - F^{-1}(u^h(\alpha)) < Kn^{-1/2} + F^{-1}(p) - F^{-1}(u^h(\alpha)) \right) + O(n^{-1})$$

$$= P \left( \frac{\hat{Q}_U^L(u^h(\alpha)) - u^h(\alpha)}{f(F^{-1}(u^h(\alpha)))} < n^{-1/2}K + \frac{n^{-1/2}z_\alpha\sqrt{p(1-p)}}{f(F^{-1}(p))} \right) + O(n^{-1/2})$$

$$= P \left( \frac{\sqrt{n} \left( \hat{Q}_U^B(u^h(\alpha)) - u^h(\alpha) \right)}{\sqrt{p(1-p)}} < \frac{Kf(F^{-1}(p))}{\sqrt{p(1-p)} + z_\alpha} \right) + O(n^{-1/2}\log(n))$$
\[ \Phi \left( z_\alpha + \frac{K f(F^{-1}(p))}{\sqrt{p(1-p)}} \right). \]

Analogous reasoning gives similar results for upper one-sided and two-sided power.

**Proof of Lemma 5**

Since the result is similar to other kernel bias results, and since the special case of \( d = 1 \) and \( b = 2 \) is already given in [Bhattacharya and Gangopadhyay (1990)](1990), we leave the proof to the supplemental appendix and provide a sketch here. The approach is to start from the definitions of \( Q_{Y|X}(p; C_h) \) and \( Q_{Y|X}(p; x) \),

\[
p = \int_{C_h} \left\{ \int_{-\infty}^{Q_{Y|X}(p; C_h)} f_{Y|X}(y; x) \, dy \right\} f_{X|C_h}(x) \, dx,
\]

\[
p = \int_{-\infty}^{Q_{Y|X}(p; x)} f_{Y|X}(y; x) \, dy,
\]

leading to

\[ 0 = \int_{C_h} \left\{ \int_{Q_{Y|X}(p; C_h)}^{Q_{Y|X}(p; x)} f_{Y|X}(y; x) \, dy \right\} f_{X|C_h}(x) \, dx. \]

After a change of variables to \( w = x/h \), a Taylor expansion around \( w = 0 \) can be taken. When \( b = 2 \), a second-order expansion is justified; generally, the smoothness determines the order of both the expansion and the remainder. There are many terms since \( x \) (now \( w \)) appears in multiple places, but the expansion is standard calculus. Finally, the conditional quantile function derivatives are translated to conditional CDF and PDF derivatives by identities following from derivatives of

\[ p = F_{Y|X}(Q_{Y|X}(p; x); x). \]

**Proof of Theorem 6**

To be concrete, we take the unconditional method to be the uncalibrated [Hutson (1999)](1999) method. This does not affect the proof structure, only the order of the term CPE\(_U\) below. As in [Chaudhuri (1991)](1991), we consider a deterministic bandwidth sequence, leaving treatment of a random (data-dependent) bandwidth to future work.

Theorem 4 states the CPE for Hutson’s (1999) CIs in terms of \( n \). In the conditional setting, we instead have the effective sample size \( N_n \), which is random. Using the same argument as [Chaudhuri (1991)](1991) proof of Thm. 3.1, p. 769), invoking Bernstein’s inequality and the Borel–Cantelli Lemma, \( N_n \) is almost surely of order \( nh^d \). This result might hold under dependent sampling (for \( X \)) by replacing Bernstein’s inequality for independent data with that in [Fan et al. (2012)](2012), but such is beyond our scope. Another difference with the unconditional case is that the local sample’s distribution, \( F_{Y|C_h} \), changes with \( n \) (through
In the lower one-sided case, let \( CPE \equiv \exp(\alpha) \) where the order of the approximation error from switching to \( \tilde{Q}_{Y|X}^{L}(u_h) \) comes from Theorem 2, and the other remainder is the subsequent magnitude \( N \). The \( CPE \) of \( \tilde{Q}_{Y|X}^{L}(u_h) \) follows a (collapsing) beta distribution, which converges to a (collapsing) normal PDF at a \( \sqrt{N} \) rate (Lemma 7). The CP is

\[
P\left(Q_{Y|X}(p; 0) < \tilde{Q}_{Y|X}^{L}(u_h)\right)
= P\left(Q_{Y|X}(p; C_h) < \tilde{Q}_{Y|X}^{L}(u_h)\right)
+ \left[ P\left(Q_{Y|X}(p; 0) < \tilde{Q}_{Y|X}^{L}(u_h)\right) - P\left(Q_{Y|X}(p; C_h) < \tilde{Q}_{Y|X}^{L}(u_h)\right)\right]
= 1 - \alpha + \text{CPE}_{U} + \text{CPE}_{Bias},
\]

where \( \text{CPE}_{U} \) is CPE due to the unconditional method and \( \text{CPE}_{Bias} \) comes from the bias.

As before, define \( B_h \equiv Q_{Y|X}(p; C_h) - Q_{Y|X}(p; 0) \). From Lemma 5, \( B_h = O(h) \) with \( b \equiv \min\{s_Q, s_X + 1, 2\} \). Let \( F_{Q_{Y|X}^{L}(u_h)}(\cdot) \) and \( f_{Q_{Y|X}^{L}(u_h)}(\cdot) \) be the CDF and PDF of \( \tilde{Q}_{Y|X}^{L}(u_h) \). We can approximate

\[
\text{CPE}_{Bias} = P\left(Q_{Y|X}(p; 0) < \tilde{Q}_{Y|X}^{L}(u_h)\right) - P\left(Q_{Y|X}(p; C_h) < \tilde{Q}_{Y|X}^{L}(u_h)\right)
= P\left(\tilde{Q}_{Y|X}^{L}(u_h) < Q_{Y|X}(p; C_h)\right) - P\left(\tilde{Q}_{Y|X}^{L}(u_h) < Q_{Y|X}(p; 0)\right)
= P\left(\tilde{Q}_{Y|X}^{L}(u_h) < Q_{Y|X}(p; C_h)\right) - P\left(\tilde{Q}_{Y|X}^{L}(u_h) < Q_{Y|X}(p; C_h)\right) + O(B_h N^{-1/2})
= F_{Q_{Y|X}^{L}(u_h)}(Q_{Y|X}(p; C_h)) - F_{Q_{Y|X}^{L}(u_h)}(Q_{Y|X}(p; 0)) + O(B_h N^{-1/2})
= B_h f_{Q_{Y|X}^{L}(u_h)}(Q_{Y|X}(p; C_h)) + O(B_h N^{-1/2} + B^2_h N). \tag{33}
\]

where the order of the approximation error from switching to \( \tilde{Q}_{Y|X}^{L}(u_h) \) from \( \tilde{Q}_{Y|X}^{L}(u_h) \) comes from Theorem 2 and the other remainder is the subsequent \( B^2_h \) term in the Taylor expansion that would be multiplied by an \( O(N) \) PDF derivative as in (35). From the aforementioned PDF convergence of \( F_{Y|X}^{L}(\tilde{Q}_{Y|X}^{L}(u_h)) \) to a normal, \( f_{Q_{Y|X}^{L}(u_h)}(Q_{Y|X}(p; 0)) \) \( \asymp \) \( N^{-1/2} \). Since \( B_h = O(h) \) from Lemma 5, the dominant term of \( \text{CPE}_{Bias} \) is \( O(N^{-1/2} h^b) \). The expression in (33) holds for \( B_h > 0 \) (leading to over-coverage) or \( B_h < 0 \) (under-coverage).

The dominant terms of \( \text{CPE}_{U} \) and \( \text{CPE}_{Bias} \) are thus respectively \( O(N^{-1}) \) and \( O(N^{-1/2} h^b) \). The term \( \text{CPE}_{U} \) in Theorem 4 is always positive. The optimal \( h \) sets equal the orders of magnitude \( N^{-1} \asymp N^{-1/2} h^b \), leading to \( h \asymp n^{-3/(2b+3d)} \) and CPE of \( O(n^{-2b/(2b+4d)}) \).

If the calibrated method from Theorem 4 is used, then \( \text{CPE}_{U} = O\left(N^{-3/2} \log(N_n)\right) \). Ignoring the log for simplicity, the optimal bandwidth rate is \( h \asymp n^{-2/(b+2d)} \). This leads to CPE of \( O(n^{-3b/(2b+4d)} \log(n)) \).
In the upper one-sided case, similarly,

$$\text{CPE}_{\text{Bias}} = P\left( Q_{Y|X}(p; 0) < \hat{Q}_{Y|C_h}^L(u_{\ell}) \right) - P\left( Q_{Y|X}(p; C_h) > \hat{Q}_{Y|C_h}^L(u_{\ell}) \right)$$

$$= F_{\hat{Q}_{Y|C_h}^L}(Q_{Y|X}(p; 0)) - F_{\hat{Q}_{Y|C_h}^L}(Q_{Y|X}(p; C_h)) + O(B_h N_n^{-1/2})$$

$$= -B_h f_{\hat{Q}_{Y|C_h}^L}(Q_{Y|X}(p; C_h)) + O(B_h N_n^{-1/2} + B_h^2 N_n).$$  \hspace{1cm} (34)

Opposite before, $B_h > 0$ now contributes under-coverage and $B_h < 0$ over-coverage, but the order of $\text{CPE}_{\text{Bias}}$ is the same.

**Two-sided case**

With two-sided inference, the lower and upper endpoints have opposite bias effects. For the median, the dominant terms of these effects cancel completely. For other quantiles, there is a partial, order-reducing cancellation. Two-sided CP is similar to (32) and (33) but with an additional Taylor expansion term:

$$P\left( \hat{Q}_{Y|C_h}^L(u_{\ell}) < Q_{Y|X}(p; 0) < \hat{Q}_{Y|C_h}^L(u_h) \right)$$

$$= 1 - \alpha + \text{CPE}_U + \left[ 1 - F_{\hat{Q}_{Y|C_h}^L}(Q_{Y|X}(p; C_h)) \right] - \left[ 1 - F_{\hat{Q}_{Y|C_h}^L}(Q_{Y|X}(p; 0)) \right]$$

$$+ F_{\hat{Q}_{Y|C_h}^L}(Q_{Y|X}(p; C_h)) - F_{\hat{Q}_{Y|C_h}^L}(Q_{Y|X}(p; 0)) + O(B_h N_n^{-1/2})$$

$$= 1 - \alpha + \text{CPE}_U + B_h \left[ f_{\hat{Q}_{Y|C_h}^L}(Q_{Y|X}(p; C_h)) - f_{\hat{Q}_{Y|C_h}^L}(Q_{Y|X}(p; C_h)) \right]$$

$$+ (1/2)B_h^2 \left[ f'_{\hat{Q}_{Y|C_h}^L}(Q_{Y|X}(p; C_h)) - f'_{\hat{Q}_{Y|C_h}^L}(Q_{Y|X}(p; C_h)) \right]$$

$$+ O\left\{ B_h^3 f''_{\hat{Q}}(Q_{Y|X}(p; C_h)) + B_h N_n^{-1/2} \right\}. \hspace{1cm} (35)$$

For the special case of the median, the $B_h$ term zeroes out. Consequently, the $B_h^2$ term dominates, decreasing overall CPE. The optimal rate of $h$ equates $(nh^d)^{-1} \approx h^{2b}(nh^d)$, so $h^* \approx n^{-1/(b+d)}$ and CPE is $O(n^{-b/(b+d)})$. With the calibrated method, the rates are instead $h^* \approx n^{-5/(4b+5d)}$ and CPE $= O(n^{-6b/(4b+5d) \log(n)}).

Even with $p \neq 1/2$, the same rates hold. As seen in (46), the PDF difference multiplying $B_h$ is only $O(1)$, smaller than the $O(N_n^{1/2})$ PDF value multiplying $B_h$ in the one-sided expression (33). This makes the $B_h$ and $B_h^2$ terms the same order, as discussed further in Section 3.1.
B Implementation of methods

B.1 Plug-in bandwidth

Plug-in bandwidth: overview and setup

The dominant terms of $CPE_U$ and $CPE_{Bias}$ have known expressions, not simply rates. Theorem 4 gives an exact expression for the $O(N_n^{-1})$ $CPE_U$ term. Instead of using this for calibration, which achieves smaller theoretical $CPE$, we use it to pin down the optimal bandwidth value since the bandwidth is a crucial determinant of finite-sample $CPE$. $CPE$ calibration, which achieves smaller theoretical $CPE$, we use it to pin down the optimal bandwidth sets the dominant terms of $CPE_U$ and $CPE_{Bias}$ to sum to zero if possible, or else minimizes their sum.

The plug-in bandwidth is derived for $d = 1$ and $b = 2$. Let $\phi(\cdot)$ be the standard normal PDF, $z_{1-\alpha}$ be the $(1 - \alpha)$-quantile of the standard normal distribution, $u_h > p$ (or $u_\ell < p$) be the quantile determining the high (low) endpoint of a lower (upper) one-sided CI, $\epsilon_h \equiv (N_n + 1)u_h - [(N_n + 1)u_\ell]$, and $\epsilon_\ell \equiv (N_n + 1)u_\ell - [(N_n + 1)u_h]$. Let $I_H$ denote a 100$(1 - \alpha)$% CI constructed using Hutson’s (1999) method on a univariate data sample of size $N_n$. The coverage probability of a lower one-sided $I_H$ (with the upper one-sided result substituting $\epsilon_\ell$ for $\epsilon_h$) is

$$P\{F^{-1}(p) \in I_H\} = 1 - \alpha + N_n^{-1} z_{1-\alpha} \frac{\epsilon_h (1 - \epsilon_h)}{p(1 - p)} \phi(z_{1-\alpha}) + O(N_n^{-3/2} \log(N_n)),$$  
(36)

or for a two-sided CI,

$$P\{F^{-1}(p) \in I_H\} = 1 - \alpha + N_n^{-1} z_{1-\alpha/2} \frac{\epsilon_h (1 - \epsilon_h) + \epsilon_\ell (1 - \epsilon_\ell)}{p(1 - p)} \phi(z_{1-\alpha/2}) + O(N_n^{-3/2} \log(N_n)).$$  
(37)

In either case there is $O(N_n^{-1})$ over-coverage.

The $CPE_{Bias}$ expressions depend on the PDF and PDF derivatives of the CI endpoints. For $u \in (0, 1),$

$$F_{Y|C_h}(\hat{Q}_u) \sim \beta[(N_n + 1)u, (N_n + 1)(1 - u)],$$

and let $f_\beta(\cdot)$ and $F_\beta(\cdot)$ be the corresponding PDF the CDF. For the lower one-sided Hutson (1999) CI, upper endpoint quantile $u = u_h$ satisfies $F_\beta(p) = \alpha$. Applying the chain rule and the identity $F_{Y|C_h}(Q_{Y|X}(p; C_h)) = p,$

$$f_{\hat{Q}_u}(x) = \frac{\partial}{\partial x} F_{\hat{Q}_u}(x) = \frac{\partial}{\partial x} F_{\beta}(F_{Y|C_h}(x)) = f_\beta(F_{Y|C_h}(x)) f_{Y|C_h}(x),$$

$$f'_{\hat{Q}_u}(x) = \frac{\partial}{\partial x} f_{\hat{Q}_u}(x) = f_\beta(F_{Y|C_h}(x)) f_{Y|C_h}(x) f_{Y|C_h}(x) + f_\beta(F_{Y|C_h}(x)) f'_{Y|C_h}(x),$$

$$f_{\hat{Q}_u}(Q_{Y|X}(p; C_h)) = f_\beta(F_{Y|C_h}(Q_{Y|X}(p; C_h))) f_{Y|C_h}(Q_{Y|X}(p; C_h)) = f_\beta(p; u) f_{Y|C_h}(F_{Y|C_h}^{-1}(p))$$

$$= f_\beta(p; u) f_{Y|X}(\xi_p; 0),$$

$$f'_{\hat{Q}_u}(Q_{Y|X}(p; C_h)) - f'_{\hat{Q}_u}(Q_{Y|X}(p; C_h))$$

62
Using (40) and Lemma 3, for

\[ f_\beta(p; u) \equiv \frac{\Gamma(N_n + 1)}{\Gamma(0)\Gamma((N_n + 1)(1 - u))}p^{(N_n+1)u-1}(1-p)^{(N_n+1)(1-u)-1} \]

\[ = \frac{\sqrt{N_n}}{\sqrt{u(1-u)}} \left[ \phi \left( \frac{[p - u]/\sqrt{u(1-u)/N_n}}{O(N_n^{-1/2})} \right) + O(N_n^{-1/2}) \right], \tag{40} \]

where the beta PDF approximation is from Lemma 7. To avoid iteration, let \( \epsilon_h = \epsilon_\ell = 0.2 \) as a rule of thumb. (More precisely, \( \epsilon = 0.5 - (1/2)\sqrt{1 - 4\sqrt{2}/9} \approx 0.20 \) makes the constants cancel.) Since \( f_\beta'(\cdot) \) is uniformly bounded in a neighborhood of zero, then

\[ P_C = \int_{C_h} f_X(x) \, dx = \int_{C_h} [f_X(0) + x f_X'(0) + (1/2)x^2 f_X''(\tilde{x})] \, dx = 2hf_X(0) + O(h^2), \]

\[ N_n \equiv nP_C = 2nhf_X(0) + O(h^3). \tag{41} \]

**Plug-in bandwidth: one-sided**

For a one-sided CI, CPE\(_U\) is in (36), and CPE\(_\text{Bias}\) is found by substituting (38) into (33) or (34). When CPE\(_\text{Bias}\) < 0, with \( \epsilon = \epsilon_h \) or \( \epsilon = \epsilon_\ell \), the optimal \( h \) equates

\[ N_n^{-1}z_{1-\alpha} \frac{\epsilon(1-\epsilon)}{p(1-p)} \phi(z_{1-\alpha}) = h^2 \frac{f_X(0)F_X^{(0,2)}(\xi_p; 0) + 2f_X'(0)F_X^{(0,1)}(\xi_p; 0)}{6f_X(0)f_X(\xi_p; 0)} f_\beta(p; u) f_Y|X(\xi_p; 0). \tag{42} \]

Using (40) and Lemma 3 for \( u \in \{u_\ell, u_h\} \),

\[ f_\beta(p; u) = N_n^{1/2}[u(1-u)]^{-1/2} \left[ \phi \left( \frac{\pm z_{1-\alpha}\sqrt{u(1-u)/N_n}}{\sqrt{u(1-u)/N_n}} \right) + O(N_n^{-1/2}) \right] \]

\[ \pm N_n^{1/2}[u(1-u)]^{-1/2}\phi(z_{1-\alpha}). \]

Plugging in \( n \equiv 2nhf_X(0) \) and \( u = p + O(N_n^{-1/2}) \), the optimal \( h \) is now an explicit function of known values and estimable objects. Let \( \hat{h}_{++} \) be the plug-in bandwidth if both CPE\(_U\) > 0 and CPE\(_\text{Bias}\) > 0, and \( \hat{h}_{+-} \) if CPE\(_\text{Bias}\) < 0. Up to smaller-order terms,

\[ [2n\hat{h}_{+-} f_X(0)]^{-1}z_{1-\alpha} \frac{\epsilon(1-\epsilon)}{p(1-p)} \phi(z_{1-\alpha}) = \hat{h}_{+-}^2 \frac{f_X(0)F_X^{(0,2)}(\xi_p; 0) + 2f_X'(0)F_X^{(0,1)}(\xi_p; 0)}{6f_X(0)f_X(\xi_p; 0)} \]

\[ \times [2n\hat{h}_{+-} f_X(0)]^{1/2}[p(1-p)]^{-1/2}\phi(z_{1-\alpha})f_Y|X(\xi_p; 0), \]
and solving for $\hat{h}$ (and plugging in estimators of the unknown objects) yields the expression in the text. The extra coefficient on $\hat{h}^{++}$ is due to taking the first-order condition of the CPE minimization problem rather than simply setting the two types of CPE to sum to zero. These $\hat{h}$ hold for both lower and upper one-sided inference.

**Plug-in bandwidth: two-sided, $p = 1/2$**

For two-sided inference with $p = 1/2$, the $B_h$ term becomes zero, as shown in (44). Then $\text{CPE}_{\text{Bias}} < 0$ since $B_h^2 > 0$, $f'_{Q_{Y|X}^{CPE}}(p) < 0$, and $f'_{Q_{Y|X}^{CPE}}(p) > 0$. The optimal $h$ sets $\text{CPE}_{\text{Bias}} + \text{CPE}_U = 0$. Substituting (37) and (39) into (35), $\hat{h}$ solves

$$N_n^{-1}z_{1-\alpha/2}\frac{\epsilon_h(1-\epsilon_h) + \epsilon_\ell(1-\epsilon_\ell)}{p(1-p)}\phi(z_{1-\alpha/2})$$

$$= -(1/2)\hat{h}^4 \left(\frac{f_X(0)f_{Y|X}^{(0,0)}(\xi_p;0) + 2f_X(0)f_{Y|X}^{(0,1)}(\xi_p;0)}{6f_X(0)f_{Y|X}(\xi_p;0)}\right)^2$$

$$\times \left\{-z_{1-\alpha/2}N_n\phi(z_{1-\alpha/2})2[p(1-p)]^{-1}[f_{Y|X}(\xi_p;0)]^2\right\},$$

$$[2n\hat{h}f_X(0)]^{-2}\epsilon(1-\epsilon)$$

$$= \frac{\hat{h}^4}{36f_X(0)^2f_{Y|X}(\xi_p;0)}\left(f_X(0)f_{Y|X}^{(0,0)}(\xi_p;0) + 2f_X(0)f_{Y|X}^{(0,1)}(\xi_p;0)\right)^2,$$

$$\hat{h} = n^{-1/3}\left|f_X(0)f_{Y|X}^{(0,0)}(\xi_p;0) + 2f_X(0)f_{Y|X}^{(0,1)}(\xi_p;0)\right|^{-1/3},$$

(43)

again using rule-of-thumb $\epsilon = 0.2$ and rounding down a leading 1.19 to one.

**Plug-in bandwidth: two-sided, $p \neq 1/2$**

For two-sided inference with $p \neq 1/2$, the dominant bias terms do not cancel completely, but the difference is of smaller order than the $O(N_n^{1/2})$ in the one-sided case. From (46),

$$f_\beta(p; u_h) - f_\beta(p; u_\ell) = -(1/3)z_{1-\alpha/2}\phi(z_{1-\alpha/2})\frac{2p - 1}{p(1-p)} + O(N_n^{-1/2}) = O(1).$$

Our two CPE terms are thus of orders $N_n^{-1} \approx n^{-1}h^{-1}$ and $h^2$. This implies that $h^* \approx n^{-1/3}$ and that CPE is $O(n^{-2/3})$. If so, the $B_h^2$ term is of order $h^4N_n = h^5n = n^{-2/3}$, so it must also be included. Although the second term from the product rule derivative in the $B_h^2$ term is not zero this time, it is smaller-order and thus omitted below.

Using (35), (38), (39), and (46),

$$\text{CPE}_{\text{Bias}} \approx B_h\left[f_{Q_{Y|X}^{CPE}}(Q_{Y|X}(p; C_h)) - f_{Q_{Y|X}^{CPE}}(Q_{Y|X}(p; C_h))\right]$$

$$+ (1/2)B_h^2\left[f'_{Q_{Y|X}^{CPE}}(Q_{Y|X}(p; C_h)) - f'_{Q_{Y|X}^{CPE}}(Q_{Y|X}(p; C_h))\right]$$

$$+ (1/8)B_h^3\left[f''_{Q_{Y|X}^{CPE}}(Q_{Y|X}(p; C_h)) - f''_{Q_{Y|X}^{CPE}}(Q_{Y|X}(p; C_h))\right]$$

$$+ (1/24)B_h^4\left[f'''_{Q_{Y|X}^{CPE}}(Q_{Y|X}(p; C_h)) - f'''_{Q_{Y|X}^{CPE}}(Q_{Y|X}(p; C_h))\right].$$
\[ \hat{h} = B_h f_{Y|C_h} \left( F_{Y|C_h}^{-1}(p) \right) \left[ f_{\beta}(p; u_h) - f_{\beta}(p; u_\ell) \right] 
\] 
\[ - (1/2) B_h^2 \hat{h} z_{1-\alpha/2} N_n \phi(z_{1-\alpha/2}) 2[p(1-p)]^{-1} \left[ f_{Y|C_h} \left( F_{Y|C_h}^{-1}(p) \right) \right]^2 
\] 
\[ \frac{f_{Y|X}(\xi_p;0) z_{1-\alpha/2} \phi(z_{1-\alpha/2})}{p(1-p)} \left[ B_h(1 - 2p)/3 - B_h^2 N_n f_{Y|X}(\xi_p;0) \right]. \tag{44} \]

With large enough \( h \), \(^{[44]} \) is always negative, so there always exists an optimal \( h \) that cancels the dominant terms of CPE\(_C\) and CPE\(_{Bias}\). This \( h \) solves

\[ -N_n^{-1} z_{1-\alpha/2} \frac{\epsilon_h(1 - \epsilon_h) + \epsilon_\ell(1 - \epsilon_\ell)}{p(1-p)} \phi(z_{1-\alpha/2}) 
\]

\[ = f_{Y|X}(\xi_p;0) \frac{z_{1-\alpha/2} \phi(z_{1-\alpha/2})}{p(1-p)} \left[ B_h(1 - 2p)/3 - B_h^2 N_n f_{Y|X}(\xi_p;0) \right]. \]

Continuing to solve for \( \hat{h} \) using \(^{[41]} \) and then Lemma \(^{[5]} \)

\[ -[2n \hat{h} f_X(0)]^{-1} 2\epsilon(1 - \epsilon) = f_{Y|X}(\xi_p;0) \left\{ \hat{h}^2 \left( B_h/h^2 \right)(1 - 2p)/3 - \hat{h}^4 (B_h^2/h^4) [2n \hat{h} f_X(0)] f_{Y|X}(\xi_p;0) \right\}, \]

\( \{n^{-1}[f_X(0)]^{-1} \epsilon(1 - \epsilon) \}
\]

\[ = \hat{h}^6 \left\{ 2n[f_{Y|X}(\xi_p;0)]^2 f_X(0) B_h^2/h^4 \right\} - \hat{h}^3 \{ f_{Y|X}(\xi_p;0)[(1 - 2p)/3] B_h/h^2 \}, \]

\[ 0 = (\hat{h}^3)^2 \{na\} - (\hat{h}^3)\{b\} - \{c/n\}, \]

\[ \hat{h}^3 = \frac{b \pm \sqrt{b^2 + 4ac}}{2an}, \]

\[ \hat{h} = n^{-1/3} \left( \frac{B_h/|B_h|}{2 f_X(0) F_{Y|X}^{(0,2)}(\xi_p;0) + 2 f_X(0) F_{Y|X}^{(0,1)}(\xi_p;0)} \right) \left( \frac{(1 - 2p)^2 + 4}{(1 - 2p)^2 + 4} \right) \]

setting \( \epsilon \) to match \(^{[43]} \) when \( p = 1/2 \).

**Plug-in bandwidth: beta PDF difference approximation**

The following approximates the beta PDF difference \( f_{\beta}(p; u_h) - f_{\beta}(p; u_\ell) \). For either \( u = u_h \) or \( u = u_\ell \), from the proof of Lemma \(^{[7]} \), where here \( \Delta x_1 = p, \Delta x_2 = 1 - p, \Delta k_1 = (n + 1)u, \Delta k_2 = (n + 1)(1 - u), J = 1, \) and \( u = p = O(n^{-1/2}) \),

\[ \log(f_{\Delta x}(\Delta x)) = K + h(\Delta x) + O(n^{-1}), \]

\[ K = (J/2) \log(n/2\pi) + \frac{1}{2} \sum_{j=1}^{J+1} \log(n/|\Delta k_j - 1|), \]

\[ e^K = [1 + O(n^{-1})] \sqrt{\frac{n}{2\pi}} \sqrt{\frac{1}{u(1-u)}}, \]

65
\[ h(\Delta x) = -(1/2)(n - J)^2 \sum_{j=1}^{J+1} \frac{(\Delta x_j - [\Delta k_j - 1]/[n - J])^2}{\Delta k_j - 1} \]
\[ + (1/3)(n - J)^3 \sum_{j=1}^{J+1} \frac{(\Delta x_j - [\Delta k_j - 1]/[n - J])^3}{(\Delta k_j - 1)^2} + O(n^{-1}) \]
\[ = -(1/2)n[(u - p)^2/[u(1 - u)] + 2(u - p)(2u - 1)/[nu(1 - u)]] \]
\[ + (1/3)n(p - u)^3[u^2 - (1 - u)^2] + O(n^{-1}). \]

First, note that \( e^K \) will put the \( \sqrt{n} \) coefficient in front, so the \( O(n^{-1}) \) remainder in \( h(\Delta x) \) is required in order to yield \( O(n^{-1/2}) \) overall. Even within \( e^K \), we must keep track of \( u = p \pm n^{-1/2}z\sqrt{p(1 - p)} + O(n^{-1}) \), not just \( u \approx p \), where for brevity \( z \equiv z_{1-\alpha/2} \). So

\[ [u_h(1 - u_h)]^{-1/2} = [p(1 - p)]^{-1/2} - (1/2)n^{-1/2}z \frac{1 - 2p}{p(1 - p)} + O(n^{-1}), \]
\[ e^K(u_h) = \frac{n^{1/2}}{\sqrt{2\pi p(1 - p)}} \left[ 1 + (1/2)n^{-1/2}z \frac{2p - 1}{\sqrt{p(1 - p)}} \right] + O(n^{-1/2}), \]
\[ e^K(u_\ell) = \frac{n^{1/2}}{\sqrt{2\pi p(1 - p)}} \left[ 1 - (1/2)n^{-1/2}z \frac{2p - 1}{\sqrt{p(1 - p)}} \right] + O(n^{-1/2}). \]

Second, for \( u_h \), let \( u_h = p + n^{-1/2}c_1 + n^{-1}c_2 + O(n^{-3/2}) \) from Lemma 3. Then, \((u - p)^2 = n^{-1}c_1^2 + 2n^{-3/2}c_1c_2 + O(n^{-2}) \) and \((u - p)^3 = n^{-3/2}c_2^3 + O(n^{-2}) \). Plugging into (45) and using the small \( x \) approximations \( e^x = 1 + x + O(x^2) \) and \( 1/(1 + x) = 1 - x + O(x^2) \),

\[ h(\Delta x) = -\frac{1}{2p(1 - p)} - n^{-1/2} \left( \frac{c_1(c_2 + 2p - 1)}{p(1 - p)} - \frac{(2p - 1)c_1^2}{3p^2(1 - p)^2} + \frac{c_1^2z(2p - 1)}{2[p(1 - p)]^{3/2}} \right) + O(n^{-1}) \]
\[ = -z^2/2 - 2n^{-1/2}z \frac{2p - 1}{3\sqrt{p(1 - p)}} + O(n^{-1}), \]
\[ \exp\{h(\Delta x)\} = e^{-z^2/2} \left[ 1 - (2/3)n^{-1/2}z \frac{2p - 1}{\sqrt{p(1 - p)}} \right] + O(n^{-1}), \]
while for \( u_\ell \),

\[ \exp\{h(\Delta x)\} = e^{-z^2/2} \left[ 1 + (2/3)n^{-1/2}z \frac{2p - 1}{\sqrt{p(1 - p)}} \right] + O(n^{-1}). \]

Altogether,

\[ f_\beta(p; u_h) = \frac{n^{1/2}}{\sqrt{p(1 - p)}} \phi(z) \left[ 1 - (1/6)n^{-1/2}z \frac{2p - 1}{\sqrt{p(1 - p)}} \right] + O(n^{-1/2}), \]
\[ f_\beta(p; u_\ell) = \frac{n^{1/2}}{\sqrt{p(1 - p)}} \phi(z) \left[ 1 + (1/6)n^{-1/2}z \frac{2p - 1}{\sqrt{p(1 - p)}} \right] + O(n^{-1/2}), \]
\[ f_\beta(p; u_h) - f_\beta(p; u_\ell) = -(1/3)z\phi(z) \frac{2p - 1}{p(1 - p)} + O(n^{-1/2}). \]  
(46)