

IMPROVED QUANTILE INFERENCE VIA FIXED-SMOOTHING ASYMPTOTICS AND EDGEWORTH EXPANSION

DAVID M. KAPLAN

ABSTRACT. Estimation of a sample quantile's variance requires estimation of the probability density at the quantile. The common quantile spacing method involves smoothing parameter m . When $m, n \rightarrow \infty$, the corresponding Studentized test statistic asymptotically follows a standard normal distribution. Holding m fixed asymptotically yields a nonstandard distribution dependent on m that contains the Edgeworth expansion term capturing the variance of the quantile spacing. Consequently, the fixed- m distribution is more accurate than the standard normal under both asymptotic frameworks. For the fixed- m test, I propose an m to maximize power subject to size control, as calculated via Edgeworth expansion. Compared with similar methods, the new method controls size better and maintains good or better power in simulations. Results for two-sample quantile treatment effect inference are given in parallel.

KEYWORDS: Edgeworth expansion; Fixed-smoothing asymptotics; Inference; Quantile; Studentize; Testing-optimal.

1. INTRODUCTION

This paper considers inference on population quantiles, specifically via the Studentized test statistic from Siddiqui (1960) and Bloch & Gastwirth (1968) (jointly SBG hereafter), for one- and two-sample setups. Median inference is a special case. Non-parametric inference on conditional quantiles is an immediate extension if the conditioning variables are all discrete¹; for continuous conditioning variables, the approach

Date: July 5, 2013; first version available online January 14, 2011.

Email: KaplanDM@Missouri.edu. Thanks to Yixiao Sun for much patient and enthusiastic guidance and discussion. Thanks also to Karen Messer, Dimitris Politis, Andres Santos, Hal White, Brendan Beare, Ivana Komunjer, Joel Sobel, Marjorie Flavin, Kelly Paulson, Matt Goldman, Linchun Chen, and other attentive and interactive listeners from UCSD audiences. Mail: Department of Economics, University of Missouri–Columbia, 909 University Ave, 118 Professional Bldg, Columbia, MO 65211-6040, USA..

¹In addition to common variables like race and gender, variables like age measured in years may be treated as discrete as long as enough observations exist for each value of interest.

in Kaplan (2012) may be used with an adjusted bandwidth rate. Like a t -statistic for the mean, the SBG statistic is normalized using a consistent estimate of the variance of the quantile estimator, and it is asymptotically standard normal.

I develop a new asymptotic theory that is more accurate than the standard normal reference distribution traditionally used with the SBG statistic. With more accuracy comes improved inference properties,² i.e. controlling size where previous methods were size-distorted without sacrificing much power. A plug-in procedure to choose the testing-optimal smoothing parameter m translates the theory into practice.

This work builds partly on Goh (2004), who suggests using fixed- m asymptotics to improve inference on the Studentized quantile. Goh (2004) uses simulated critical values to examine the performance of existing m suggestions, which are tailored to standard normal critical values. Here, a simple and accurate analytic approximation is given for the critical values. Using this approximation, the new procedure for selecting m is tailored to the fixed- m distribution. I also provide the theoretical justification for the accuracy of the fixed- m distribution, which complements the simulations in Goh (2004).

The two key results here are a nonstandard fixed- m asymptotic distribution and a higher-order Edgeworth expansion. For a scalar location model, Siddiqui (1960) gives the fixed- m result, and Hall & Sheather (1988, hereafter cited as “HS88”) give a special case of the Edgeworth expansion in Theorem 1 below. The Edgeworth expansion is more accurate than a standard normal since it contains higher-order terms that otherwise end up in the remainder, so its remainder becomes of smaller order. There are intuitive reasons why the fixed- m approximation is also more accurate than the standard normal, but we show this by rigorous theoretical arguments.

In the standard first-order asymptotics, both $m \rightarrow \infty$ and $n \rightarrow \infty$ (and $m/n \rightarrow 0$); since $m \rightarrow \infty$, this may be called “large- m ” asymptotics. In contrast, fixed- m asymptotics only approximates $n \rightarrow \infty$ while fixing m at its actual finite sample value. Fixed- m is an instance of “fixed-smoothing” asymptotics in the sense that this variance does not go to zero in the limit as it does in “increasing-smoothing.” It turns out that the fixed- m asymptotics includes the high-order Edgeworth term capturing the variance of the quantile spacing. Consequently, the fixed- m distribution is

²Inverting the level- α tests proposed here yields level- α confidence intervals. I use hypothesis testing language throughout, but size distortion is analogous to coverage probability error, and higher power corresponds to shorter interval lengths.

higher-order accurate under the conventional large- m asymptotics, while the standard normal distribution is only first-order accurate. Under the fixed- m asymptotics, the standard normal is not even first-order accurate. In other words, the fixed- m distribution is more accurate than the standard normal irrespective of the asymptotic framework. Similar results on the accuracy of fixed-smoothing asymptotics have been found in time series inference (e.g., Sun, Phillips & Jin, 2008); this result suggests fixed-smoothing asymptotics may be valuable more generally.

Not only is accuracy gained with the Edgeworth and fixed- m distributions, but they are sensitive to m . By reflecting the effect of m , they allow us to determine the best choice of m . With the fixed- m and Edgeworth results, I construct a test dependent on m using the fixed- m critical values, and then evaluate the type I and type II error of the test using the more accurate Edgeworth expansion. Then I can optimally select m to minimize type II error subject to control of type I error. Approximating (instead of simulating) the fixed- m critical values reveals the Edgeworth/fixed- m connection above, and it also makes the test easier to implement in practice.

The critical value correction here provides size control robust to incorrectly chosen m , whereas in HS88 the choice of m is critical to controlling size (or not). In simulations, the new method has correct size even where the HS88 method is size-distorted. Power is still good because m is explicitly chosen to maximize it, using the Edgeworth expansion; in simulations, it can be significantly better than HS88. HS88 do not provide a separate result for the two-sample case. Monte Carlo simulations show that the new method controls size better than the common Gaussian plug-in version of HS88 and various bootstrap methods, while maintaining competitive power. The method of Hutson (1999), following a fractional order statistic approach, has better properties but is not always computable. A two-sample analog with similarly good performance is given by Goldman & Kaplan (2012). For two-sample inference, bootstrap can perform well, though the new method is good and open to a particular refinement that would increase power; more research in both approaches is needed.

Section 2 presents the model and hypothesis testing problem. Section 3 gives a non-standard asymptotic result and corresponding corrected critical value, while Section 4 gives Edgeworth expansions of the standard asymptotic limiting distribution. Using these, a method for selecting smoothing parameter m is given in Section 5, which is followed by a simulation study and conclusion. Results for the two-sample extension

are provided in parallel. More details and discussion are available in the working paper version; full technical proofs and calculations are in the supplemental appendices; all are available on the author's website.

2. QUANTILE ESTIMATION AND HYPOTHESIS TESTING

For simplicity and intuition, consider an iid sample of continuous random variable X , whose p -quantile is ξ_p . The estimator $\hat{\xi}_p$ used in SBG is an order statistic. The r th order statistic for a sample of n values is defined as the r th smallest value in the sample and written as $X_{n,r}$, such that $X_{n,1} < X_{n,2} < \dots < X_{n,n}$. The SBG estimator is

$$\hat{\xi}_p = X_{n,r}, \quad r = \lfloor np \rfloor + 1,$$

where $\lfloor np \rfloor$ is the floor function (greatest integer not exceeding np). Writing the cumulative distribution function (cdf) of X as $F(x)$ and the probability density function (pdf) as $f(x) \equiv F'(x)$, I make the usual assumption that $f(x)$ is positive and continuous in a neighborhood of the point ξ_p . Consequently, ξ_p is the unique p -quantile such that $F(\xi_p) = p$.

The standard asymptotic result (for the vector form, see Mosteller, 1946; Siddiqui, 1960, also gives the following scalar result) for a central³ quantile estimator is

$$\sqrt{n}(X_{n,r} - \xi_p) \xrightarrow{d} N[0, p(1-p)\{f(\xi_p)\}^{-2}].$$

A consistent estimator of $1/f(\xi_p)$ that is asymptotically independent of $X_{n,r}$ leads to the Studentized sample quantile, which has the pivotal asymptotic distribution

$$(1) \quad \frac{\sqrt{n}(X_{n,r} - \xi_p)}{\sqrt{p(1-p)}\left\{\widehat{1/f(\xi_p)}\right\}} \xrightarrow{d} N(0, 1).$$

Siddiqui (1960) and Bloch & Gastwirth (1968) propose and show consistency of

$$(2) \quad \widehat{1/f(\xi_p)} = S_{m,n} \equiv \frac{n}{2m}(X_{n,r+m} - X_{n,r-m})$$

when $m \rightarrow \infty$ and $m/n \rightarrow 0$ as $n \rightarrow \infty$.

For two-sided inference on ξ_p , I consider the parameterization $\xi_p = \beta - \gamma/\sqrt{n}$. The null and alternative hypotheses are $H_0 : \xi_p = \beta$ and $H_1 : \xi_p \neq \beta$, respectively. When

³“Central” means that in the limit, $r/n \rightarrow p \in (0, 1)$ as $n \rightarrow \infty$; i.e., $r \rightarrow \infty$ and is some fraction of the sample size n . In contrast, “intermediate” would take $r \rightarrow \infty$ but $r/n \rightarrow 0$ (or $r/n \rightarrow 1$, $n - r \rightarrow \infty$); “extreme” would fix $r < \infty$ or $n - r < \infty$.

$\gamma = 0$ the null is true. The test statistic examined in this paper is

$$(3) \quad T_{m,n} \equiv \frac{\sqrt{n}(X_{n,r} - \beta)}{S_{m,n} \sqrt{p(1-p)}}$$

and will be called the SBG test statistic due to its use of (2). From (1), $T_{m,n}$ is asymptotically standard normal when $\gamma = 0$. A corresponding hypothesis test would then compare $T_{m,n}$ to critical values from a standard normal distribution.

For the two-sample case, assume that there are independent samples of X and Y , with n_x and n_y observations, respectively. For simplicity, let $n = n_x = n_y$. For instance, if $2n$ individuals are separated into balanced treatment and control groups, one might want to test if the treatment effect at quantile p has significance level α . The marginal pdfs are $f_X(\cdot)$ and $f_Y(\cdot)$. Interest is in testing if $\xi_{p,x} = \xi_{p,y}$. Under the null hypothesis $H_0 : \xi_{px} = \xi_{py} = \xi_p$, the first-order asymptotic result is

$$\sqrt{n}(X_{n,r} - \xi_p) - \sqrt{n}(Y_{n,r} - \xi_p) \xrightarrow{d} N(0, p(1-p) [\{f_X(\xi_p)\}^{-2} + \{f_Y(\xi_p)\}^{-2}]),$$

using the fact that the variance of the sum (or difference) of two independent normals is the sum of the variances. The pivot for the two-sample case is then

$$\frac{\sqrt{n}(X_{n,r} - Y_{n,r})}{\sqrt{\{f_X(\xi_p)\}^{-2} + \{f_Y(\xi_p)\}^{-2}} \sqrt{p(1-p)}} \xrightarrow{d} N(0, 1).$$

The Studentized version uses the same quantile spacing estimators as above:

$$\tilde{T}_{m,n} \equiv \frac{\sqrt{n}(X_{n,r} - Y_{n,r})}{\sqrt{\{n/(2m)\}^2(X_{n,r+m} - X_{n,r-m})^2 + \{n/(2m)\}^2(Y_{n,r+m} - Y_{n,r-m})^2} \sqrt{p(1-p)}}.$$

The same m is used for X and Y in anticipation of a Gaussian plug-in approach that would yield the same m for X and Y regardless. This two-sample setup is easily extended to the case of unequal sample sizes $n_x \neq n_y$, starting with $\sqrt{n_x}(X_{n,r} - \xi_p) - \sqrt{n_x/n_y} \sqrt{n_y}(Y_{n,r} - \xi_p)$ and assuming n_x/n_y is constant asymptotically, but this is omitted for clarity.

3. FIXED- m ASYMPTOTICS AND CORRECTED CRITICAL VALUE

The standard asymptotic result with $m \rightarrow \infty$ as $n \rightarrow \infty$ is a standard normal distribution for the SBG test statistic $T_{m,n}$ under the null hypothesis when $\gamma = 0$. Since in reality $m < \infty$ for any given finite sample test, holding m fixed as $n \rightarrow \infty$ may give a more accurate asymptotic approximation of the true finite sample test statistic distribution. With $m \rightarrow \infty$, there is increasing smoothing and $\text{Var}(S_{m,n}) =$

$O(1/m) \rightarrow 0$; with m fixed, we have fixed-smoothing asymptotics since that variance does not disappear. The fixed- m distribution is given below, along with a simple formula for critical values that doesn't require simulation.

3.1. Fixed- m asymptotics. Siddiqui (1960, eqn. 5.4) provides the fixed- m asymptotic distribution; Goh (2004, Appendix D.1) provides a nice alternative proof for the median, which readily extends to any quantile.⁴ If $\gamma = 0$,

$$(4) \quad T_{m,n} \xrightarrow{d} \mathcal{Z}/\mathcal{V}_{4m} \equiv T_{m,\infty} \quad \text{as } n \rightarrow \infty, m \text{ fixed,}$$

with $\mathcal{Z} \sim N(0, 1)$, $\mathcal{V}_{4m} \sim \chi_{4m}^2/(4m)$, $\mathcal{Z} \perp \mathcal{V}_{4m}$, and $S_{m,n}$ and $T_{m,n}$ as in (2) and (3).

The above distribution is conceptually similar to the Student's t -distribution. A t -statistic from normally distributed iid data has a standard normal distribution if either the variance is known (so denominator is constant) or the sample size approaches infinity. The distribution $T_{m,\infty}$ is also standard normal if either the variance is known (denominator in $T_{m,n}$ is constant) or $m \rightarrow \infty$. When using an estimated variance in a finite sample t -statistic, the more accurate t -approximation has fatter tails than a standard normal, and it is given by $\mathcal{Z}/\sqrt{\mathcal{V}_v}$, where $\mathcal{Z} \sim N(0, 1)$, $\mathcal{V}_v \sim \chi_v^2/v$ with v the degrees of freedom, and $\mathcal{Z} \perp \mathcal{V}_v$. Similarly, when an estimated variance is used in the SBG test statistic (the random $S_{m,n}$ instead of constant $1/f(\xi_p)$ in the denominator), the result is the distribution in (4) above with asymptotically independent numerator and denominator. The fixed- m distribution reflects this uncertainty in the variance estimator that is lost under the standard asymptotics.

For the two-sample test statistic under the null, as $n \rightarrow \infty$ with m fixed,

$$\begin{aligned} \tilde{T}_{m,n} &\xrightarrow{d} \frac{\mathcal{Z}}{\mathcal{U}} \equiv \tilde{T}_{m,\infty}, \quad \text{where} \\ \mathcal{U} &\sim (1 + \epsilon)^{1/2}, \quad \epsilon \equiv \frac{(\mathcal{V}_{4m,1}^2 - 1) + (\mathcal{V}_{4m,2}^2 - 1)\{f_X(\xi_p)/f_Y(\xi_p)\}^2}{1 + \{f_X(\xi_p)/f_Y(\xi_p)\}^2}, \end{aligned}$$

and \mathcal{Z} , $\mathcal{V}_{4m,1}$, and $\mathcal{V}_{4m,2}$ are mutually independent. The derivation is in the full appendix online. The strategy is the same, using results from Siddiqui (1960) for asymptotic independence and distributions for each component.

Unlike the one-sample case, $\tilde{T}_{m,\infty}$ is not pivotal. We can either estimate the density derivative $f_X(\xi_p)/f_Y(\xi_p)$ or consider the upper bound for critical values. This is the

⁴The result stated for a general quantile in Theorem 2 in Section 3.5 appears to have a typo and not follow from a generalization of the proof given, nor does it agree with Siddiqui (1960).

same nuisance parameter faced by the two-sample fractional order statistic method of Goldman & Kaplan (2012). Note that ϵ is a weighted average of $\mathcal{V}_{4m,1}^2 - 1$ and $\mathcal{V}_{4m,2}^2 - 1$, where the weights sum to one and $\mathcal{V}_{4m,1} \perp \mathcal{V}_{4m,2}$. To see the effect of the weights, consider any W_1 and W_2 with $\text{Cov}(W_1, W_2) = 0$, $\text{Var}(W_1) = \text{Var}(W_2) = \sigma_W^2$, and $\lambda \in [0, 1]$. Then

$$\begin{aligned} \text{Var}\{\lambda W_1 + (1 - \lambda)W_2\} &= \lambda^2 \text{Var}(W_1) + (1 - \lambda)^2 \text{Var}(W_2) + 2\lambda(1 - \lambda)\text{Cov}(W_1, W_2) \\ &= \sigma_W^2 \{\lambda^2 + (1 - \lambda)^2\}. \end{aligned}$$

This is maximized by $\lambda = 1$ or $1 - \lambda = 1$, giving σ_W^2 . The minimum has first order condition $0 = 2\lambda + 2(1 - \lambda)(-1)$, yielding $\lambda = 1/2 = 1 - \lambda$ and $\text{Var}\{\lambda W_1 + (1 - \lambda)W_2\} = \sigma_W^2/2$. This means the variance of ϵ (and \mathcal{U} and $\tilde{T}_{m,\infty}$) is smallest when the weights are each $1/2$, which is when $f_X(\xi_p) = f_Y(\xi_p)$. When the weights are zero and one, which is when the variance of ϵ is largest, we get the special case of testing against a constant and $\tilde{T}_{m,\infty} = T_{m,\infty}$. Thus, critical values from the one-sample case provide conservative inference in the two-sample case.

3.2. Corrected critical value. An approximation of the fixed- m cdf around the standard normal cdf will lead to critical values based on the standard normal distribution. The standard normal cdf is $\Phi(\cdot)$, the standard normal pdf is $\phi(\cdot) = \Phi'(\cdot)$, and the first derivative $\phi'(z) = -z\phi(z)$. The first three central moments of the χ_{4m}^2 distribution are $4m$, $8m$, and $32m$, respectively. The general approach here is to use the independence of \mathcal{Z} and \mathcal{V}_{4m} to rewrite the cdf in terms of Φ , and then to expand around $E(\mathcal{V}_{4m}) = 1$. Using (4), for a critical value z ,

$$\begin{aligned} P(T_{m,\infty} < z) &= P(\mathcal{Z}/\mathcal{V}_{4m} < z) = E\{\Phi(z\mathcal{V}_{4m})\} = E[\Phi\{z + z(\mathcal{V}_{4m} - 1)\}] \\ &= E[\Phi(z) + \Phi'(z)z(\mathcal{V}_{4m} - 1) + (1/2)\Phi''(z)\{z(\mathcal{V}_{4m} - 1)\}^2 + O(m^{-2})] \\ &= \Phi(z) - \frac{z^3\phi(z)}{2} E\left\{\left(\frac{\chi_{4m}^2 - 4m}{4m}\right)^2\right\} + O(m^{-2}) \\ (5) \quad &= \Phi(z) - \frac{z^3\phi(z)}{4m} + O(m^{-2}). \end{aligned}$$

Note that (5) is a distribution function itself: its derivative in z is $\{\phi(z)/(4m)\}(z^4 - 3z^2 + 4m)$, which is positive for all z when $m > 9/16$ (as it always is); the limits at $-\infty$ and ∞ are zero and one; and it is càdlàg since it is differentiable everywhere. The approximation error $O(m^{-2})$ does not change if m is fixed, but it goes to zero as $m \rightarrow \infty$, which the selected m does as $n \rightarrow \infty$ (see Section 5.3). Uniform convergence

of $\Phi(z) - z^3\phi(z)/(4m)$ to $P(T_{m,\infty} < z)$ over $z \in \mathbb{R}$ as $m \rightarrow \infty$ then obtains via Pólya's Theorem (e.g., DasGupta, 2008, Thm. 1.3(b)). Appendix A shows the accuracy of (5) for $m > 2$; for $m \leq 2$, simulated critical values can be used, as in the provided code.

To find the value z that makes the nonstandard cdf above take probability $1 - \alpha$ (for rejection probability α under the null for an upper one-sided test), I set $\Phi(z) - z^3\phi(z)/(4m) = 1 - \alpha$ and solve for z . If z is some deviation of order m^{-1} around the value $z_{1-\alpha}$ such that $\Phi(z_{1-\alpha}) = 1 - \alpha$ for the standard normal cdf, then solving for c in $z = z_{1-\alpha} + c/m$ gives

$$(6) \quad c = z_{1-\alpha}^3/4 + O(m^{-1}),$$

and thus the upper one-sided test's corrected critical value is

$$z = z_{1-\alpha} + c/m = z_{1-\alpha} + \frac{z_{1-\alpha}^3}{4m} + O(m^{-2}).$$

For a symmetric two-sided test, since the additional term in (5) is an odd function of z , the critical value is the same but with $z_{1-\alpha/2}$, yielding

$$(7) \quad z = z_{\alpha,m} + O(m^{-2}), \quad z_{\alpha,m} \equiv z_{1-\alpha/2} + \frac{z_{1-\alpha/2}^3}{4m}.$$

Note $z_{\alpha,m} > z_{1-\alpha/2}$ and depends on m ; e.g., $z_{0.05,m} = 1.96 + 1.88/m$. Using the same method with an additional term in the expansion, as compared in Appendix A,

$$z = z_{1-\alpha/2} + \frac{z_{1-\alpha/2}^3}{4m} + \frac{z_{1-\alpha/2}^5 + 8z_{1-\alpha/2}^3}{96m^2} + O(m^{-3}).$$

The rest of this paper uses (7); the results in Section 5 also go through with the above third-order corrected critical value or with a simulated critical value.

In the two-sample case under $H_0 : F_X^{-1}(p) = F_Y^{-1}(p) = \xi_p$, after calculating some moments (see supplemental Two-sample Appendix C), the fixed- m distribution can be approximated by

$$\begin{aligned} P(\tilde{T}_{m,\infty} < z) &= \Phi(z) + \phi(z)zE(\mathcal{U} - 1) + (1/2)\{-z\phi(z)\}z^2E\{(\mathcal{U} - 1)^2\} \\ &\quad + O[E\{(\mathcal{U} - 1)^3\}] \\ &= \Phi(z) - \frac{\phi(z)}{4m} \left[z^3 \frac{\{f_X(\xi_p)\}^{-4} + \{f_Y(\xi_p)\}^{-4}}{\tilde{S}_0^4} - 2z \frac{\{f_X(\xi_p)\}^{-2}\{f_Y(\xi_p)\}^{-2}}{\tilde{S}_0^4} \right] \\ &\quad + O(m^{-2}) \end{aligned}$$

$$\begin{aligned}
 &= \Phi(z) - \frac{\phi(z)}{4m} \left[z^3 \frac{1 + \delta^4}{(1 + \delta^2)^2} - 2z \frac{\delta^2}{(1 + \delta^2)^2} \right] + O(m^{-2}), \\
 (8) \quad \tilde{S}_0 &\equiv \left([f_X\{F_X^{-1}(p)\}]^{-2} + [f_Y\{F_Y^{-1}(p)\}]^{-2} \right)^{1/2}, \quad \delta \equiv f_X(\xi_p)/f_Y(\xi_p).
 \end{aligned}$$

The corresponding critical value is

$$\begin{aligned}
 z &= \tilde{z}_{\alpha,m} + O(m^{-2}) \leq z_{\alpha,m} + O(m^{-2}), \\
 \tilde{z}_{\alpha,m} &= z_{1-\alpha/2} + \frac{z_{1-\alpha/2}^3 [\{f_X(\xi_p)\}^{-4} + \{f_Y(\xi_p)\}^{-4}] - 2z_{1-\alpha/2} \{f_X(\xi_p)\}^{-2} \{f_Y(\xi_p)\}^{-2}}{4m\tilde{S}_0^4} \\
 (9) \quad &= z_{1-\alpha/2} + \frac{z_{1-\alpha/2}^3(1 + \delta^4) - 2z_{1-\alpha/2}\delta^2}{4m(1 + \delta^2)^2}.
 \end{aligned}$$

When the pdf of Y collapses toward a constant, $\{f_Y(\xi_p)\}^{-1} \rightarrow 0$ and so $\delta \rightarrow 0$; $P(\tilde{T}_{m,\infty} < z)$ reduces to (5), and $\tilde{z}_{\alpha,m}$ to $z_{\alpha,m}$. Thus, as an alternative to estimating $\delta = f_X(\xi_p)/f_Y(\xi_p)$, the one-sample critical value $z_{\alpha,m}$ provides conservative two-sample inference, as discussed in Section 3.1. Under the exchangeability assumption used for permutation tests, $\delta = 1$ and $\tilde{z}_{\alpha,m}$ attains its smallest possible value.

4. EDGEWORTH EXPANSION

HS88 give the Edgeworth expansion for the asymptotic distribution of $T_{m,n}$ when $\gamma = 0$. To calculate the type II error of a test using the fixed- m critical values (Section 5.2), the case $\gamma \neq 0$ is needed. Section 5.2 shows how to apply this result.

4.1. One-sample case. The result here includes the result of HS88 as a special case, and indeed the results match⁵ when $\gamma = 0$. The u_i functions are indexed by γ , so $u_{1,0}$ is the same as u_1 from HS88 (page 385), and similarly for $u_{2,0}$ and $u_{3,0}$.

Theorem 1. *Assume $f(\xi_p) > 0$ and that, in a neighborhood of ξ_p , f'' exists and satisfies a Lipschitz condition, i.e. for some $\epsilon > 0$ and all x, y sufficiently close to ξ_p , $|f''(x) - f''(y)| \leq \text{constant}|x - y|^\epsilon$. Suppose $m = m(n) \rightarrow \infty$ as $n \rightarrow \infty$, in such a manner that for some fixed $\delta > 0$ and all sufficiently large n , $n^\delta \leq m(n) \leq n^{1-\delta}$.*

⁵There appears to be a typo in the originally published result; the first part of the first term of u_1 in HS88 was $(1/6)[p(1-p)]^{1/2}$, but it appears to be $(1/6)[p/(1-p)]^{1/2}$ as given above.

Define $C \equiv \gamma f(\xi_p)/\sqrt{p(1-p)}$, and defining functions

$$\begin{aligned} u_{1,\gamma}(z) &\equiv \frac{1}{6} \left(\frac{p}{1-p} \right)^{1/2} \frac{1+p}{p} (z^2 - 1) - C \sqrt{\frac{1-p}{p}} \left[1 - \frac{pf'(\xi_p)}{\{f(\xi_p)\}^2} - \frac{1}{2(1-p)} \right] z \\ &\quad - \frac{1}{2} \left(\frac{p}{1-p} \right)^{1/2} \left[1 + \frac{f'(\xi_p)}{\{f(\xi_p)\}^2} (1-p) \right] z^2 \\ &\quad - \{([np] + 1 - np) - 1 + (1/2)(1-p)\} \{p(1-p)\}^{-1/2}, \\ u_{2,\gamma}(z) &\equiv \frac{1}{4} (2Cz^2 - C^2z - z^3), \text{ and} \\ u_{3,\gamma}(z) &\equiv \frac{3\{f'(\xi_p)\}^2 - f(\xi_p)f''(\xi_p)}{6\{f(\xi_p)\}^4} (z - C), \end{aligned}$$

it follows that

$$\begin{aligned} &\sup_{-\infty < z < \infty} \left| P(T_{m,n} < z) - \{ \Phi(z + C) + n^{-1/2} u_{1,\gamma}(z + C) \phi(z + C) \right. \\ &\quad \left. + m^{-1} u_{2,\gamma}(z + C) \phi(z + C) + (m/n)^2 u_{3,\gamma}(z + C) \phi(z + C) \right| \\ (10) \quad &= o\{m^{-1} + (m/n)^2\}. \end{aligned}$$

A sketch of the proof is given in Appendix D. The assumptions are the same as in HS88.

For $\gamma = 0$, the term $m^{-1} u_{2,0}(z) \phi(z) = -z^3 \phi(z)/(4m)$ is identical to the term in the fixed- m distribution in (5). Thus under the null, the fixed- m distribution captures the high-order Edgeworth term associated with the variance of $S_{m,n}$. In other words, the fixed- m distribution is high-order accurate under the conventional large- m asymptotics, while the standard normal distribution is only first-order accurate. Since the fixed- m distribution is also more accurate under fixed- m asymptotics, where the standard normal is not even first-order accurate, theory strongly indicates that fixed- m critical values are more accurate, which is born out in simulations here and in Goh (2004).

Let $\gamma \neq 0$ so that the null hypothesis is false, where as before $H_0 : \xi_p = \beta$ with $\xi_p = \beta - \gamma/\sqrt{n}$. Letting $S_0 \equiv 1/f(\xi_p)$,

$$\begin{aligned} T_{m,n} &= \frac{\sqrt{n}(X_{n,r} - \xi_p) - \gamma}{S_{m,n} \sqrt{p(1-p)}} = \frac{\sqrt{n}(X_{n,r} - \xi_p)}{S_{m,n} \sqrt{p(1-p)}} - \frac{\gamma}{\sqrt{p(1-p)} S_0} \left(\frac{S_0}{S_{m,n}} + 1 - 1 \right), \\ P(T_{m,n} < z) &= P \left\{ \frac{\sqrt{n}(X_{n,r} - \xi_p)}{S_{m,n} \sqrt{p(1-p)}} - \frac{\gamma}{\sqrt{p(1-p)} S_0} \left(\frac{S_0}{S_{m,n}} + 1 - 1 \right) < z \right\} \end{aligned}$$

$$\begin{aligned}
 &= P \left\{ \frac{\sqrt{n}(X_{n,r} - \xi_p)}{S_{m,n}\sqrt{p(1-p)}} - \frac{\gamma}{\sqrt{p(1-p)}S_0} \left(\frac{S_0}{S_{m,n}} - 1 \right) < z + C \right\} \\
 (11) \quad &= P \left\{ \frac{\sqrt{n}(X_{n,r} - \xi_p) + \gamma(S_{m,n}/S_0 - 1)}{S_{m,n}\sqrt{p(1-p)}} < z + C \right\}.
 \end{aligned}$$

If the true S_0 were known and used in $T_{m,n}$ instead of its estimator $S_{m,n}$, this would be simply the distribution from HS88 with a shift of the critical value by $C \equiv \gamma/\{S_0\sqrt{p(1-p)}\}$, which is γ normalized by the true (hypothetically known) variance. But $S_{m,n}$ is random, so the HS88 expansion is insufficient and Theorem 1 is needed.

4.2. Two-sample case. The strategy and results are similar; the full proof may be found in the supplemental Two-sample Appendix.

Theorem 2. *Let $X_i \stackrel{iid}{\sim} F_X$ and $Y_i \stackrel{iid}{\sim} F_Y$, and $X_i \perp Y_j \forall i, j$. With assumptions in Theorem 1 applied to both F_X and F_Y , and letting $\tilde{C} \equiv \gamma\tilde{S}_0^{-1}/\sqrt{p(1-p)}$, define functions*

$$\begin{aligned}
 \tilde{u}_{1,\gamma}(z) &\equiv \frac{1}{6} \frac{1+p}{\sqrt{p(1-p)}} \left(\frac{g_x^3 - g_y^3}{\tilde{S}_0^3} \right) (z^2 - 1) \\
 &\quad + \{ (a_2g_x^2 - a'_2g_y^2)(1-p) + (a_1g_x^2 - a'_1g_y^2) \} \{ p/(1-p) \}^{1/2} (2p^2\tilde{S}_0^3)^{-1} (z^2) \\
 &\quad - \{ 2(a_2g_x^2 - a'_2g_y^2) + (a_1g_x^2 - a'_1g_y^2)/(1-p) \} (2p\tilde{S}_0^3)^{-1} C \{ (1-p)/p \}^{1/2} (z) \\
 &\quad + \{ (a_2g_x^2 - a'_2g_y^2) - \tilde{S}_0^2(a_2 - a'_2) \} \{ p(1-p) \}^{1/2} (2p^2\tilde{S}_0^3)^{-1} \\
 &\quad - \{ ([np] + 1 - np) - 1 + (1/2)(1-p) \} (g_x - g_y) \left(\tilde{S}_0\sqrt{p(1-p)} \right)^{-1}, \\
 \tilde{u}_{2,\gamma}(z) &\equiv -\frac{1}{4} \left(\frac{g_x^4 + g_y^4}{\tilde{S}_0^4} \right) z^3 + \frac{1}{2} \frac{g_x^2g_y^2}{\tilde{S}_0^4} z - \frac{1}{2} \frac{g_x^2g_y^2}{\tilde{S}_0^4} \tilde{C} + \frac{1}{4} \left(\frac{g_x^4 + g_y^4}{\tilde{S}_0^4} \right) (2\tilde{C}z^2 - \tilde{C}^2z), \\
 \tilde{u}_{3,\gamma}(z) &\equiv \frac{g_x g_x'' + g_y g_y''}{6\tilde{S}_0^2} (z - \tilde{C}),
 \end{aligned}$$

$$\begin{aligned}
 g_X(\cdot) &\equiv 1/f_X(F_X^{-1}(\cdot)), \quad g_x \equiv g_X(p), \quad g_x'' \equiv g_X''(p), \quad H_X(x) \equiv F_X^{-1}(e^{-x}), \quad a_i \equiv H_X^{(i)} \left(\sum_{j=r}^n j^{-1} \right), \\
 g_Y(\cdot) &\equiv 1/f_Y(F_Y^{-1}(\cdot)), \quad g_y \equiv g_Y(p), \quad g_y'' \equiv g_Y''(p), \quad H_Y(y) \equiv F_Y^{-1}(e^{-y}), \quad a'_i \equiv H_Y^{(i)} \left(\sum_{j=r}^n j^{-1} \right),
 \end{aligned}$$

where $H^{(i)}(\cdot)$ is the i th derivative of function $H(\cdot)$.

It follows that

$$\begin{aligned} & \sup_{-\infty < z < \infty} \left| P\left(\tilde{T}_{m,n} < z\right) - \left\{ \Phi(z+C) + n^{-1/2}\tilde{u}_{1,\gamma}(z+C)\phi(z+C) \right. \right. \\ & \quad \left. \left. + m^{-1}\tilde{u}_{2,\gamma}(z+C)\phi(z+C) + (m/n)^2\tilde{u}_{3,\gamma}(z+C)\phi(z+C) \right\} \right| \\ & = o\{m^{-1} + (m/n)^2\}. \end{aligned}$$

For a two-sided test, $\tilde{u}_{1,\gamma}$ will cancel out, so the unknown terms therein may be ignored. Under the null, when $\tilde{C} = 0$, the $\tilde{u}_{2,\gamma}$ term again exactly matches the fixed- m distribution, demonstrating that the fixed- m distribution is more accurate than the standard normal under large- m (as well as fixed- m) asymptotics. A similar comment applies here, too, about the effect of $\tilde{S}_{m,n}$ being random.

5. OPTIMAL SMOOTHING PARAMETER SELECTION

5.1. Type I error. In order to select m to minimize type II error subject to control of type I error, expressions for type I and type II error dependent on m are needed. Comparing with the Edgeworth expansion in (10) when $\gamma = 0$, the approximate type I error of the two-sided symmetric test can be calculated using the corrected critical values from (7). Note that $u_{1,0}(z)\phi(z)$ in (10) is an even function since $\phi(z) = \phi(-z)$ and z only appears as z^2 in $u_{1,0}(z)$, so $u_{1,0}(z) = u_{1,0}(-z)$; thus it will cancel out for a two-sided test. Also note that the functions $u_{2,0}(z)$ and $u_{3,0}(z)$ are odd, so $u_{2,0}(-z) = -u_{2,0}(z)$ and $u_{3,0}(-z) = -u_{3,0}(z)$. Below, the second high-order term will disappear due to the use of the fixed- m critical value, leaving only the third high-order term.

Proposition 3. *If $\gamma = 0$, then $P(|T_{m,n}| > z_{\alpha,m}) = e_I + o\{m^{-1} + (m/n)^2\}$,*

$$(12) \quad e_I = \alpha - 2(m/n)^2 u_{3,0}(z_{1-\alpha/2})\phi(z_{1-\alpha/2}).$$

Proof. Starting with (10) with $\gamma = 0$, the two-sided symmetric test rejection probability under the null hypothesis for critical value z is

$$\begin{aligned} P(|T_{m,n}| > z \mid H_0) &= P(T_{m,n} > z \mid H_0) + P(T_{m,n} < -z \mid H_0) \\ &= 2 - 2\Phi(z) - 2m^{-1}u_{2,0}(z)\phi(z) \\ &\quad - 2(m/n)^2 u_{3,0}(z)\phi(z) + o\{m^{-1} + (m/n)^2\}. \end{aligned}$$

With the corrected critical value $z_{\alpha,m} = z_{1-\alpha/2} + z_{1-\alpha/2}^3/(4m)$,

$$\begin{aligned} P(|T_{m,n}| > z_{\alpha,m} \mid H_0) &= 2 - 2\Phi(z_{\alpha,m}) + 2z_{\alpha,m}^3\phi(z_{\alpha,m})/(4m) \\ &\quad - 2(m/n)^2 u_{3,0}(z_{\alpha,m})\phi(z_{\alpha,m}) + o\{m^{-1} + (m/n)^2\} \\ &= \alpha - 2(m/n)^2 u_{3,0}(z_{1-\alpha/2})\phi(z_{1-\alpha/2}) + o\{m^{-1} + (m/n)^2\}. \end{aligned}$$

□

The dominant part of the type I error, e_I , depends on m , n , and $z_{1-\alpha/2}$, as well as $f(\xi_p)$, $f'(\xi_p)$, and $f''(\xi_p)$ through $u_{3,0}(z_{1-\alpha/2})$. Up to higher-order remainder terms, the type I error does not exceed nominal size α if $u_{3,0}(z_{1-\alpha/2}) \geq 0$, and there will be size distortion if $u_{3,0}(z_{1-\alpha/2}) < 0$.

For common regions of common distributions, $u_{3,0}(z_{1-\alpha/2}) \geq 0$ and so $e_I \leq \alpha$ when using corrected critical values. Since $z_{1-\alpha/2} > 0$, the sign of $u_{3,0}(z_{1-\alpha/2})$ depends on the sign of $3f'(\xi_p)^2 - f(\xi_p)f''(\xi_p)$ or equivalently of the third derivative of the inverse cdf $\frac{\partial^3}{\partial p^3}F^{-1}(p)$. According to HS88, for $p = 0.5$, this is positive for all symmetric unimodal densities and almost all skew unimodal densities, except when f' changes quickly near ξ_p . I found that for any quantile p , the sign is positive for t -, normal, exponential, χ^2 , and Fréchet distributions; see Appendix B. This suggests that simply using the fixed- m corrected critical values alone is enough to reduce e_I to α or below.

For the two-sample case, as shown in the full appendix, the result is similar.

Proposition 4. *If $\gamma = 0$, then*

$$\begin{aligned} P\left(|\tilde{T}_{m,n}| > z_{\alpha,m}\right) &\leq P\left(|\tilde{T}_{m,n}| > \tilde{z}_{\alpha,m}\right) = \tilde{e}_I + o\{m^{-1} + (m/n)^2\}, \\ \tilde{e}_I &= \alpha - 2(m/n)^2 \tilde{u}_{3,0}(z_{1-\alpha/2})\phi(z_{1-\alpha/2}). \end{aligned}$$

Here also, the corrected critical value approximated from the fixed- m distribution leads to the dominant part of type I error being bounded by α for all common distributions, since the sign of \tilde{u}_3 is positive for the same reasons as in the one-sample case. The one-sample critical value $z_{\alpha,m}$ gives conservative inference; the infeasible $\tilde{z}_{\alpha,m}$ results in better power but requires estimating $\delta \equiv f_X(\xi_p)/f_Y(\xi_p)$; the $\tilde{z}_{\alpha,m}$ under exchangeability ($\delta = 1$) leads to the best power but also size distortion if exchangeability is violated.

5.2. Type II Error. As above, I use the Edgeworth expansion in (10) to approximate the type II error of the two-sided symmetric test using critical values from (7). Since a uniformly most powerful test does not exist for general alternative hypotheses, I will follow a common strategy in the optimal testing literature and pick a reasonable alternative hypothesis against which to maximize power. The hope is that this will produce a test near the power envelope at all alternatives, even if it is not strictly the uniformly most powerful.

I choose to maximize power against the alternative where first-order power is 50% for a two-sided test. From above, the type II error is thus $0.5 = P(|T_{m,n}| < z_{1-\alpha/2}) \doteq G_{C^2}(z_{1-\alpha/2}^2)$, where G_{C^2} is the cdf of a noncentral χ^2 distribution with one degree of freedom and noncentrality parameter C^2 , and C was defined in Theorem 1. For $\alpha = 0.05$, this gives $\gamma f(\xi_p)/\sqrt{p(1-p)} \equiv C = \pm 1.96$, or $\gamma = \pm 1.96\sqrt{p(1-p)}/f(\xi_p)$.

Calculation of the following type II error may be found in the working paper.

Proposition 5. *If C^2 solves $0.5 = G_{C^2}(z_{1-\alpha/2})$, and writing f for $f(\xi_p)$ and similarly f' and f'' , then*

$$\begin{aligned}
 P(|T_{m,n}| < z_{\alpha,m}) &= e_{II} + o\{m^{-1} + (m/n)^2\}, \\
 e_{II} &= 0.5 + (1/4)m^{-1}\{\phi(z_{1-\alpha/2} - C) - \phi(z_{1-\alpha/2} + C)\}Cz_{1-\alpha/2}^2 \\
 &\quad + (m/n)^2 \frac{3(f')^2 - ff''}{6f^4} z_{1-\alpha/2} \{\phi(z_{1-\alpha/2} + C) + \phi(z_{1-\alpha/2} - C)\} \\
 (13) \qquad &\quad + O(n^{-1/2}).
 \end{aligned}$$

There is a bias term and a variance term in e_{II} above. Per Bloch & Gastwirth (1968), the variance and bias of $S_{m,n}$ are indeed of order m^{-1} and $(m/n)^2$, respectively, as given in their (2.5) and (2.6):

$$\begin{aligned}
 \text{AsyVar}(S_{m,n}) &= (2mf^2)^{-1} && \text{as } m \rightarrow \infty, m/n \rightarrow 0, \\
 \text{AsyBias}(S_{m,n}) &= (m/n)^2 \frac{3(f')^2 - ff''}{6f^5} && \text{as } m \rightarrow \infty, m/n \rightarrow 0,
 \end{aligned}$$

similar to above aside from the additional $1/f^2$ and $1/f$ factors. As $m \rightarrow \infty$, $\text{Var}(S_{m,n}) \rightarrow 0$ (increasing smoothing); with m fixed, this variance is also fixed (fixed smoothing). As n grows but $S_{m,n}$ only uses a proportion of the n observations approaching zero, the bias will also decrease. The bias decreases to zero in the fixed- m thought experiment, too, and thus is not captured by the fixed- m asymptotic distribution.

Now there are expressions for the dominant components of type I and type II error as functions of m , given in (12) and (13), so m can be chosen to minimize e_{II} subject to control of e_I .

First, I look at e_{II} more closely and try to sign some terms. As discussed before, for common distributions $\frac{3(f')^2 - ff''}{6f^4} \geq 0$, so the entire $(m/n)^2$ expression is positive; type II error from this term is increasing in m , so a smaller value of m would minimize it. The m^{-1} term in (13) is also positive since $\phi(z_{1-\alpha/2} - C) > \phi(z_{1-\alpha/2} + C)$, so it is minimized by a larger m .

It is then possible that minimizing e_{II} gives an “interior” solution m , i.e. $m \in [1, \min(r-1, n-r)]$. For example, for a standard normal distribution where $n = 100$ and $p = 0.5$, m can be between one and 49. As shown in Appendix B, $3(f')^2 - ff'' = [\phi(0)]^2$, and dividing by $6f^4$ yields $1/(6[\phi(0)]^2) = 2\pi/6$. Then, in choosing m to minimize $(1/4)m^{-1}\phi(0)1.96^3 + (m/n)^2\sqrt{2\pi}(1.96)/6 = m^{-1}(0.751) + m^2(0.819)10^{-4}$, the m^{-1} term dominates for small m and the other term for large m , so that e_{II} is minimized by $m = 17$, a feasible choice in practice.

The two-sample case is similar. Calculations are in the supplemental appendix.

Proposition 6. *If C^2 solves $0.5 = G_{C^2}(z_{1-\alpha/2})$,*

$$\theta \equiv \tilde{S}_0^{-4}(g_x^4 + g_y^4) = (1 + \delta_a^4)/(1 + \delta_a^2)^2,$$

$g_x \equiv 1/f_X(F_X^{-1}(p))$ and $g_y \equiv 1/f_Y(F_Y^{-1}(p))$ are as defined in Theorem 2, as are g_x'' and g_y'' , $\delta_a \equiv f_X(F_X^{-1}(p))/f_Y(F_Y^{-1}(p))$ similar to (8), and

$$\tilde{S}_0 \equiv [\{f_X(F_X^{-1}(p))\}^{-2} + \{f_Y(F_Y^{-1}(p))\}^{-2}]^{1/2}$$

as in (8), then

$$\begin{aligned} P(|\tilde{T}_{m,n}| < \tilde{z}_{\alpha,m}) &= \tilde{e}_{II} + o\{m^{-1} + (m/n)^2\}, \\ \tilde{e}_{II} &= 0.5 + (1/4)m^{-1}[\phi(z_{1-\alpha/2} + C)\{-\theta C z_{1-\alpha/2}^2 + (1 - \theta)z_{1-\alpha/2}\} \\ &\quad + \phi(z_{1-\alpha/2} - C)\{\theta C z_{1-\alpha/2}^2 + (1 - \theta)z_{1-\alpha/2}\}] \\ &\quad + (m/n)^2 \frac{g_x g_x'' + g_y g_y''}{6\tilde{S}_0^2} z_{1-\alpha/2} \{\phi(z_{1-\alpha/2} + C) + \phi(z_{1-\alpha/2} - C)\} \\ &\quad + O(n^{-1/2}). \end{aligned}$$

5.3. Choice of m . With the fixed- m corrected critical values, $e_I \leq \alpha$ for all quantiles of common distributions, as discussed. In implementing a method to choose m , below

I use a Gaussian plug-in assumption (as is commonly done with other suggestions for m), and consequently the calculated e_I is indeed always less than nominal α (see Appendix B). Thus, the method reduces to minimizing e_{II} . Since e_{II} has a positive m^{-1} component and a positive m^2 component, it will be U-shaped for positive m . If the first-order condition yields an infeasibly big m , the biggest feasible m is the minimizer over the feasible range.

For a randomized alternative for a symmetric two-sided test, the first-order condition of (13) is

$$\begin{aligned} 0 = \frac{\partial e_{II}}{\partial m} &= -(1/4)m^{-2}[\phi(z_{1-\alpha/2} - C) - \phi(z_{1-\alpha/2} + C)]Cz_{1-\alpha/2}^2 \\ &\quad + (2m/n^2)\frac{3(f')^2 - ff''}{6f^4}z_{1-\alpha/2}[\phi(z_{1-\alpha/2} + C) + \phi(z_{1-\alpha/2} - C)], \\ m &= \left\{ (1/4)\{\phi(z_{1-\alpha/2} - C) - \phi(z_{1-\alpha/2} + C)\}Cz_{1-\alpha/2}^2 \right. \\ &\quad \left. / \left\{ (2/n^2)\frac{3(f')^2 - ff''}{6f^4}z_{1-\alpha/2}[\phi(z_{1-\alpha/2} + C) + \phi(z_{1-\alpha/2} - C)] \right\} \right\}^{1/3}. \end{aligned}$$

Remember that C is chosen ahead (such as $C = 1.96$), ϕ is the standard normal pdf, $z_{1-\alpha/2}$ is determined by α , and n is known for any given sample; but the object $\frac{3(f')^2 - ff''}{6f^4}$ is unknown. Since size control isn't dependent on m , it may be worth estimating that object even though the variance of that estimator will be large; this option has not been explored.

As is common in the kernel bandwidth selection literature, I will plug in⁶ ϕ for f (or rather $\phi(\Phi^{-1}(p))$ for $f(\xi_p)$). In all the simulations below, when np is close to 1 or n , this makes no difference in the m selected; this is also true if F is actually normal, or special cases like the median for log-normal. In some cases the m chosen will be very different, but the effect on power is small; e.g., with $\text{Unif}(0, 1)$, $n = 45$, $p = 0.5$, $\alpha = 0.05$, the Gaussian plug-in yields $m = 9$ instead of $m = 22$, but the power loss is only 4% at the alternative considered in Proposition 5. Still, estimation of f , f' , and f'' may be best for larger n .

The Gaussian plug-in yields

$$\begin{aligned} m_K(n, p, \alpha, C) &= \left\{ (1/4)\{\phi(z_{1-\alpha/2} - C) - \phi(z_{1-\alpha/2} + C)\}Cz_{1-\alpha/2}^2 \right. \\ &\quad \left. / \left\{ (2/n^2)\frac{3(\phi'(\Phi^{-1}(p)))^2 - \phi(\Phi^{-1}(p))\phi''(\Phi^{-1}(p))}{6(\phi(\Phi^{-1}(p)))^4} \right\} \right\}^{1/3}. \end{aligned}$$

⁶Here, using $N(0, 1)$ is mathematically equivalent to $N(\mu, \sigma^2)$ since μ and σ cancel out.

$$\begin{aligned}
 & \times z_{1-\alpha/2} [\phi(z_{1-\alpha/2} + C) + \phi(z_{1-\alpha/2} - C)] \}^{1/3} \\
 & = n^{2/3} (C z_{1-\alpha/2})^{1/3} (3/4)^{1/3} \left(\frac{(\phi(\Phi^{-1}(p)))^2}{2(\Phi^{-1}(p))^2 + 1} \right)^{1/3} \\
 & \quad \times \left(\frac{\phi(z_{1-\alpha/2} - C) - \phi(z_{1-\alpha/2} + C)}{\phi(z_{1-\alpha/2} - C) + \phi(z_{1-\alpha/2} + C)} \right)^{1/3},
 \end{aligned}$$

and for the 50% first-order power C and $\alpha = 0.05$, as calculated below,

$$(14) \quad m_K(n, p, \alpha = 0.05, C = 1.96) = n^{2/3} (1.42) \left(\frac{(\phi(\Phi^{-1}(p)))^2}{2(\Phi^{-1}(p))^2 + 1} \right)^{1/3}.$$

Compare (14) with the suggested m from HS88. Using their Edgeworth expansion under the null, they calculate the level error of a two-sided symmetric confidence interval constructed by inverting $T_{m,n}$ with standard normal critical values. Since the $n^{-1/2}$ term disappears for two-sided tests and confidence intervals, m can be chosen to make the level error zero as long as $u_{3,0}$ is positive⁷. Rewriting their (3.1) to match my notation above,

$$m_{HS} = \left[n^{2/3} z_{1-\alpha/2}^{2/3} [1.5f^4 / (3(f')^2 - ff'')]^{1/3} \right].$$

HS88 proceed with a median-specific, data-dependent method. More commonly, papers like Koenker & Xiao (2002) plug in ϕ for f when computing h_{HS} (HS=Hall and Sheather) in their simulations, as shown explicitly on page 3 of their electronic appendix; Goh & Knight (2009) use the same Gaussian plug-in version for their simulations. Koenker & Xiao (2002) also use a Gaussian plug-in procedure based on Bofinger (1975), whose m is of rate $n^{4/5}$ to minimize MSE, that is also considered in the simulation study below. Their “bandwidth” h is just m/n , so they use the equivalent of

$$(15) \quad m_{HS} = n^{2/3} z_{1-\alpha/2}^{2/3} \left(1.5 \frac{[\phi(\Phi^{-1}(p))]^2}{2(\Phi^{-1}(p))^2 + 1} \right)^{1/3},$$

$$(16) \quad m_B = n^{4/5} \left(4.5 \frac{[\phi(\Phi^{-1}(p))]^4}{[2(\Phi^{-1}(p))^2 + 1]^2} \right)^{1/5}.$$

For $p = 0.5$, $\alpha = 0.05$, these two will be equal if $n = 20.93$, or rather $m_B < m_{HS}$ if $n < 21$ and $m_B > m_{HS}$ if $n \geq 21$. Since they use the same critical values, size and power will be smaller whenever m is bigger.

⁷If $u_{3,0}$ is negative, their first order condition for m is similar but with an additional $(1/2)^{1/3}$ coefficient; in Section 3, they only consider positive $u_{3,0}$.

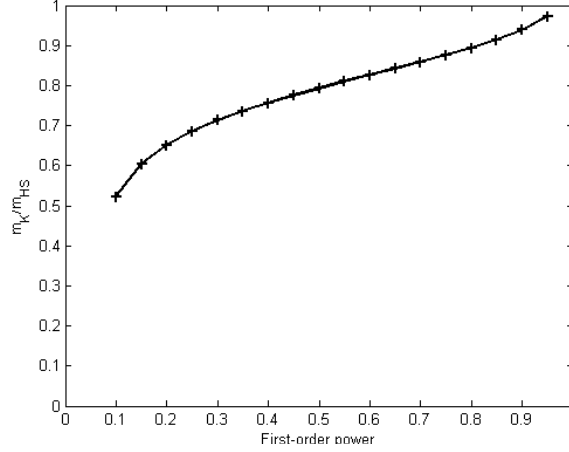


FIGURE 1. Plot of the ratio m_K/m_{HS} against the first-order power used to calculate C ; 0.5 is used throughout this paper.

Since $z_{1-\alpha/2}^{2/3}(1.5)^{1/3} = 1.79$ for $\alpha = 0.05$, $m_K(n, p, \alpha = 0.05, C = 1.96) = 0.79m_{HS}$. This ratio is calculated for other values of C based on other first-order power values, and the plot is shown in Figure 1; the dependence is mild until values close to 0% or 100% are chosen. Remember that m_K is optimized for the test using corrected critical values from the fixed- m distribution, while m_{HS} is chosen for use with standard normal critical values. It then makes sense that $m_K < m_{HS}$ since the fixed- m critical values control size, allowing for smaller m when otherwise e_I might be a binding constraint for a test with standard normal critical values.

For the two-sample case, the strategy is the same. The testing-optimal \tilde{m}_K is found by taking the first-order condition of (6) with respect to m , and then solving for m . Since the two-sample \tilde{e}_{II} has positive m^{-1} and m^2 terms, it is U-shaped, which again justifies picking the largest feasible m if the calculated m is infeasibly large. Depending if $\tilde{z}_{\alpha, m}$ or $z_{\alpha, m}$ is used, solving for m in the FOC yields

$$\begin{aligned} \tilde{z}_{\alpha, m} : \tilde{m}_K &= n^{2/3}(3/4)^{1/3} \left(\frac{\tilde{S}_0^2}{g_x g_x'' + g_y g_y''} \right)^{1/3} \\ &\quad \times \left\{ \frac{2g_x^2 g_y^2}{\tilde{S}_0^4} + \frac{g_x^4 + g_y^4}{\tilde{S}_0^4} C z_{1-\alpha/2} \frac{\phi(z_{1-\alpha/2} - C) - \phi(z_{1-\alpha/2} + C)}{\phi(z_{1-\alpha/2} - C) + \phi(z_{1-\alpha/2} + C)} \right\}^{1/3}, \\ z_{\alpha, m} : \tilde{m}_K &= n^{2/3}(3/4)^{1/3} \left(\frac{S_0^2}{g_x g_x'' + g_y g_y''} \right)^{1/3} \end{aligned}$$

$$\times \left\{ \frac{2g_x^2 g_y^2}{S_0^4} (z_{1-\alpha/2}^2 + 2) + \frac{g_x^4 + g_y^4}{S_0^4} C z_{1-\alpha/2} \frac{\phi(z_{1-\alpha/2} - C) - \phi(z_{1-\alpha/2} + C)}{\phi(z_{1-\alpha/2} - C) + \phi(z_{1-\alpha/2} + C)} \right\}^{1/3},$$

where the latter m is weakly larger, with equality in the limiting special case where either X or Y is a constant. I again chose to proceed with a Gaussian plug-in (using sample variances) for the g and \tilde{S}_0 terms due to the difficulty of estimating g'' , but power can be increased by more refined plug-in methods.

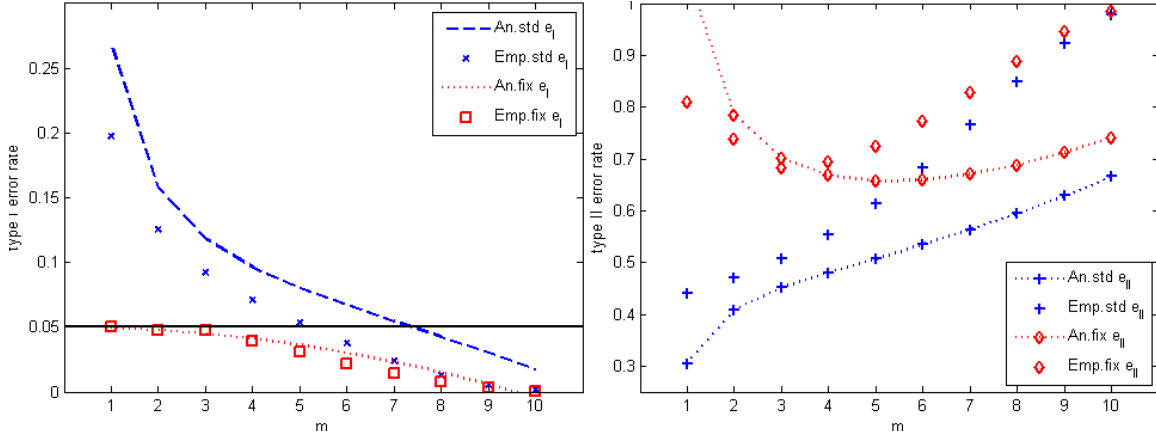


FIGURE 2. Analytic and empirical e_I and e_{II} by m .

Both: $n = 21$, $p = 0.5$, $\alpha = 0.05$, F is log-normal with $\mu = 0$, $\sigma = 3/2$, 5000 replications (for empirical values). Lines are analytic; markers without lines are empirical. Legends: “An” is analytic (Edgeworth-derived), “Emp” is empirical (simulated), “std” is standard normal critical values, and “fix” is fixed- m critical values.

Right (e_{II}): randomized alternative (see text), $\gamma/\sqrt{n} = 0.80$.

Figure 2 compares the Edgeworth approximations to the true (simulated) type I and type II error for scalar data drawn from a log-normal distribution. For non-normal distributions like this, the approximation can be significantly different. With standard normal critical values, type I error is monotonically decreasing with m while type II error is monotonically increasing, so setting $e_I = \alpha$ exactly should minimize e_{II} subject to $e_I \leq \alpha$. With fixed- m critical values, this is not true, so the Edgeworth expansion for general γ in Theorem 1 is necessary. With the fixed- m critical values, type I error is always below α (approximately, and usually truly), and the type II error curve near the selected m is very flat since it is near the minimum where the slope is zero. Consequently, a larger-than-optimal m can result in larger power loss for HS88 than the new method, and a smaller-than-optimal m incurs size distortion for HS88 but not the new method.

6. SIMULATION STUDY⁸

Regarding m , Siddiqui (1960) originally suggested a value of m on the order of $n^{1/2}$. Bloch & Gastwirth (1968) then suggested a rate of $n^{4/5}$, to minimize the asymptotic mean square error of $S_{m,n}$. With this $n^{4/5}$ rate, Bofinger (1975) then suggested the m_B above in (16). HS88, using an Edgeworth expansion under the null hypothesis, find that an m of order $n^{2/3}$ minimizes the level error of two-sided tests, and they provide an infeasible expression for m . For the median, they also give a data-dependent method for selecting m ; for a general quantile, the Gaussian plug-in version of m_{HS} in (15) is more commonly used, as in Koenker & Xiao (2002) and Goh & Knight (2009).

The following simulations compare five methods. The first method is presented in this paper, with fixed- m corrected critical values and m_K chosen to minimize type II error subject to control of type I error; the code is publicly available as Kaplan (2011). For the two-sample $\tilde{z}_{\alpha,m}$, I implemented a simplistic estimator of $\theta \equiv \tilde{S}_0^{-4}(g_x^4 + g_y^4)$ using the quantile spacings; a better estimator of θ would improve performance, as Figure 11 exemplifies. The second method uses standard normal critical values and m_{HS} as in (15). The third method uses standard normal critical values and m_B as in (16). These three methods are referred to as “New method,” “HS,” and “B,” respectively, in the text as well as figures in this section. For the new method, I used the floor function to get an integer value for m_K ; for HS and B, I instead rounded to the nearest integer, to not amplify their size distortion.

The fourth method “BS” is a bootstrap method. The conclusion of much experimentation was that a symmetric percentile- t using bootstrapped variance had the best size control in the simulations presented below. For the number of outer (B_1) and inner (B_2 , for the variance estimator) bootstrap replications, $B_1 = 99$ and $B_2 = 100$ performed best and thus were used; any exceptions are noted below.

Compared to the best-performing bootstrap method, increasing B_1 from 99 to 999 more often *increased* size distortion (Table 3), though not by much. Decreasing B_2 from 100 to 25 was also worse. Empirical size distortion was also reduced by taking the $1 - \alpha$ quantile of the absolute values and comparing the absolute value of the

⁸MATLAB code implementing a hybrid of Hutson’s method and the new method, `quantile_inf.m`, is publicly available through the author’s website or MATLAB File Exchange; R code `quantile_inf.R` is also available available on the author’s website. MATLAB code for simulations is available from the author upon request.

(non-bootstrap) test statistic instead of taking $\alpha/2$ and $1 - \alpha/2$ quantiles from the sample of B_1 Studentized bootstrap test statistics—i.e., the symmetric test was better than the equal-tailed test. The more basic percentile method (without Studentization; either equal-tail or symmetric) had worse size distortion in small samples or less central quantiles, though less distortion with the median in larger samples; and Studentizing with the SBG sparsity estimator or a kernel density estimator worsened the size distortion in some cases. Assuming a normally distributed test statistic and bootstrapping only the variance also had severe size distortion in many cases. For example, for $F = N(0, 1)$, $n = 11$, $p = 0.1$, $\alpha = 5\%$, one-sample empirical size of the symmetric percentile- t with $B_1 = 99$ and $B_2 = 100$ is 9.7%; with $B_1 = 999$ instead, 9.8%; with an equal-tail test instead, 14.6%; with a symmetric percentile with $B_1 = 99$ or 999, 17.3% or 17.5%. Using m -out-of- n bootstrap had the potential to perform better if the oracle m were used, but in many cases there was size distortion even with the best m , and the risked increase in size distortion from picking the wrong m was often large. The optimal choice of smoothing parameter there and with the smoothed bootstrap is difficult but critical; but see Bickel & Sakov (2008) and Ho & Lee (2005) for recent efforts. With respect to the example setup, a percentile method smoothing with $h = 0.92n^{-1/5}$ gives 5.1% for $N(0, 1)$, but 23.1% for slash, or 9.0% and 17.1% for smoothed percentile- t ; compare also the top and bottom halves of Table 2 in Brown, Hall & Young (2001).

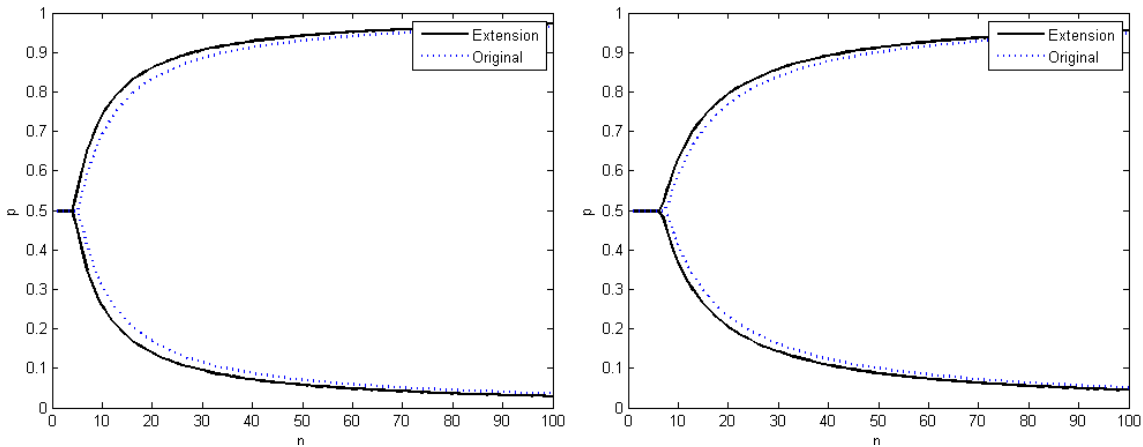


FIGURE 3. Graph showing for which combinations of quantile $p \in (0, 1)$ and sample size n Hutson (1999) is computable, in the middle region. An extension marginally increases the region by not requiring an equal-tail test. Left: $\alpha = 5\%$. Right: $\alpha = 1\%$.

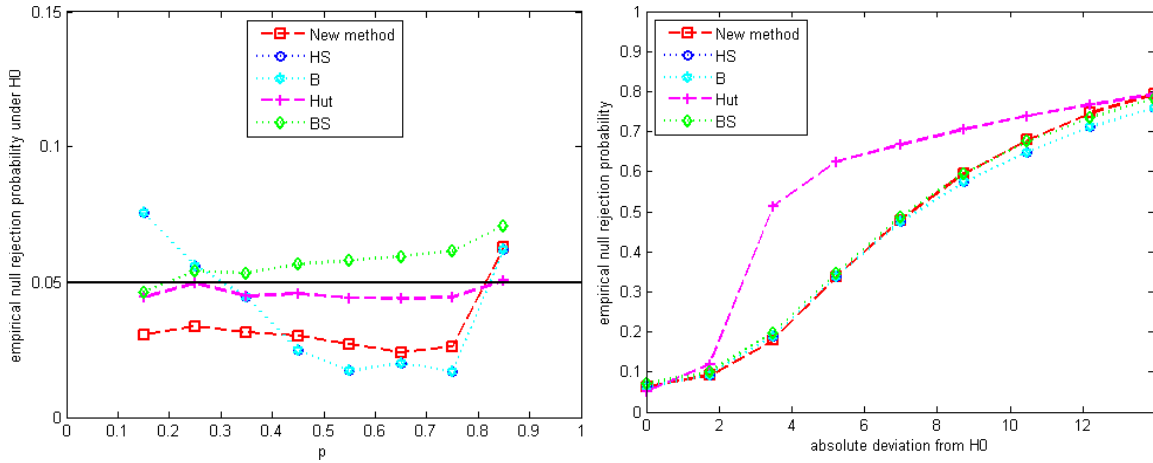


FIGURE 4. Example of superior performance of Hutson (1999): size is always controlled, power is better. In this one-sample case, uncalibrated power is better for almost all alternatives even though all other methods are size distorted. Distribution is log-normal with $\mu = 0$, $\sigma = 3/2$; $\alpha = 5\%$, $n = 21$, and power is for $p = 0.85$.

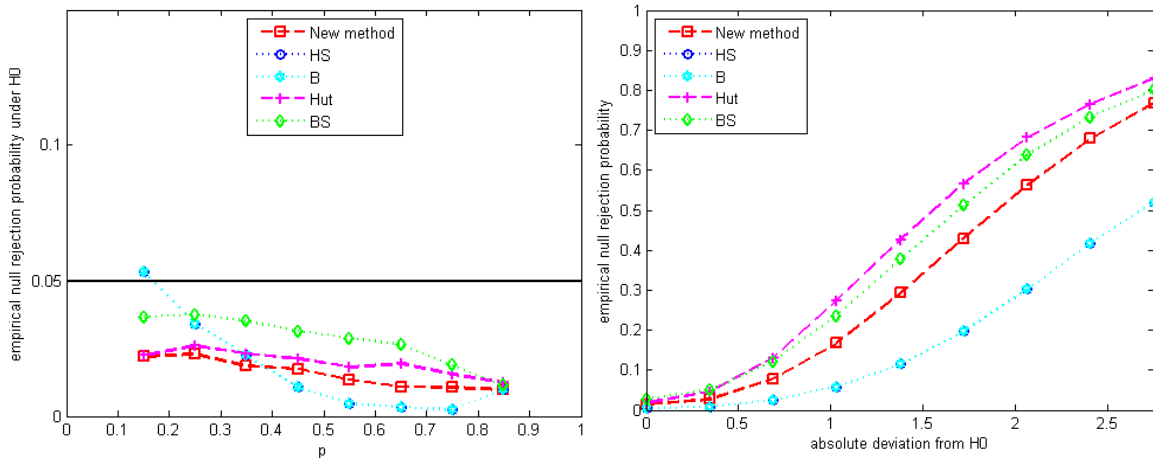


FIGURE 5. Example of superior performance of Hutson (1999), two-sample. Distributions are both log-normal with $\mu = 0$, $\sigma = 3/2$; $\alpha = 5\%$, $n = 21$, and power is for $p = 0.55$.

The fifth method is an exact method from Hutson (1999). For one-sample inference, it performed best in every single simulation where it could be used, so the results here focus on cases when it cannot be computed. An example of its superior size control and power is Figure 4. A graph showing when a simple extension of Hutson (1999), as well as the original, can be computed is in Figure 3. Instead of putting $\alpha/2$ probability in each tail as in Hutson (1999), the extension shifts probability between the tails if

necessary. Simulations show that it performs equally well within the region of Figure 3. For two-sample inference, a simplistic extension of Hutson (1999) outperforms all other methods most of the time, as in Figure 5, but not always. Since it seems like a more refined extension of Hutson (1999) will be best for two-sample inference, the results here focus on cases where that method cannot be used. However, if such a refinement is not found, bootstrap and the new method provide strong competition; HS and B can have significantly worse power, as Figure 5 shows.

Another “exact” method computes $\sum_{i=1}^n 1\{X_i < \beta\}$, which under $H_0 : \xi_p = \beta$ is distributed Binomial(n, p). One downside is that the power curve often flattens out as ξ_p deviates farther from β . A bigger downside is that randomization is needed, particularly for small n and for np near one or n . For instance, with $n = 11$, $p = 0.1$, and $\alpha = 5\%$, there is a 31.4% probability of observing zero X_i less than ξ_p . For a two-sided test, we would randomly reject $0.025/0.314 = 8\%$ of the time we observe zero, and not reject 92% of the time we observe zero. Because of the common aversion to randomized tests, this method is left out of the simulation results. However, the randomized binomial test always has exact size, and its power curve near H_0 is often steeper than the non-exact tests; both are desirable properties.

The hope of this paper was to produce a new method that controls size better than HS and B (and BS) while keeping power competitive. For one-sample inference in cases when Hutson (1999) is not computable, the following simulations show that the new method eliminates or reduces the size distortion of HS, B, and BS. The new method also has good power, sometimes even better than BS. When Hutson (1999) is computable, the new method can have better power than HS and B, too.

The two-sample results are more complicated since performance depends on how similar the two distributions are. The one-sample results represent one extreme of the two-sample case as the ratio $f_X(\xi_p)/f_Y(\xi_p)$ goes to infinity (or zero), when size is hardest to control. The other extreme is when $f_X(\xi_p) = f_Y(\xi_p)$. In that case, HS and B⁹ have less size distortion, but still some. BS, though, seems to control size and have better power than the new method with estimated θ ; with the true θ , BS and the new method perform quite similarly. General two-sample setups would have results in between these extremes of $f_X/f_Y \rightarrow \infty$ (or zero) and $f_X/f_Y = 1$.

Unless otherwise noted, 10,000 simulation replications were used, and $\alpha = 5\%$.

⁹Since HS88 and Bofinger (1975) do not provide values of m for the two-sample case, I used their m values for the univariate case.

TABLE 1. Empirical size as percentage, $n = 3, 4$, $p = 0.5$, one-sample.

| Distribution | $n = 3$ | | | $n = 4$ | | |
|--------------------------------|---------|-----|------|---------|------------------|------|
| | New | BS | HS/B | New | BS | HS/B |
| $N(0, 1)$ | 1.6 | 1.2 | 11.9 | 2.4 | 4.1 | 15.4 |
| $\text{Logn}(0, \sigma = 3/2)$ | 3.3 | 1.3 | 12.8 | 1.8 | 3.6 | 10.0 |
| $\text{Exp}(1)$ | 3.0 | 1.2 | 13.5 | 2.3 | 4.6 | 12.9 |
| $\text{Uniform}(0, 1)$ | 3.2 | 1.5 | 16.0 | 5.0 | 6.8 ^a | 21.3 |
| $\text{GEV}(0, 1, 0)$ | 1.9 | 1.3 | 12.1 | 2.1 | 4.0 | 13.6 |
| χ_1^2 | 4.8 | 1.6 | 14.8 | 2.6 | 4.3 | 11.8 |

^a6.7% with 999 outer replications

TABLE 2. Empirical size as percentage, $n = 3, 4$, $p = 0.5$, two-sample with identical independent distributions.

| Distribution | $n = 3$ | | | $n = 4$ | | |
|--------------------------------|------------------|-----|------|---------|-----|------|
| | New ^a | BS | HS/B | New | BS | HS/B |
| $N(0, 1)$ | 0.8 | 1.3 | 6.3 | 1.1 | 2.5 | 6.5 |
| $\text{Logn}(0, \sigma = 3/2)$ | 0.4 | 0.8 | 4.5 | 0.6 | 1.3 | 3.7 |
| $\text{Exp}(1)$ | 1.1 | 1.4 | 6.4 | 0.9 | 1.9 | 5.9 |
| $\text{Uniform}(0, 1)$ | 1.4 | 1.9 | 8.6 | 1.8 | 3.6 | 8.6 |

^aIf true θ used instead: 1.6, 1.1, 1.6, 2.8.

In the extreme sample size case where $n = 3$ or $n = 4$, HS and B are severely size distorted while the new method and BS control size (except BS for $n = 4$ and uniform distribution), as shown in Table 1. In the other special case for two-sample, where the distributions are identical, all rejection probabilities are much smaller, though some size distortion remains for HS and B; see Figure 2. For one-sample power, BS is very poor for $n = 3$, but better than the new method at some alternatives for $n = 4$; see Figures 6. For two-sample power, BS is equal to the new method for a range of alternatives with $n = 3$, and better for $n = 4$; see Figure 6.

For $n = 11$, all methods can have size distortion, but the new method has the least, while HS and B have the most. This occurs even for a $N(0, 1)$, but is worse for the fat-tailed slash distribution; see Figure 7. Whether or not BS is size distorted, the new method still has competitive power with it, as shown in Figure 8.

For $n = 45$ and $p = 0.05$ or $p = 0.95$, the new method controls size except for empirical size around 5.5% for the slash distribution. In contrast, BS, HS, and B are all size distorted, sometimes significantly; see Table 3. Still, the new method's

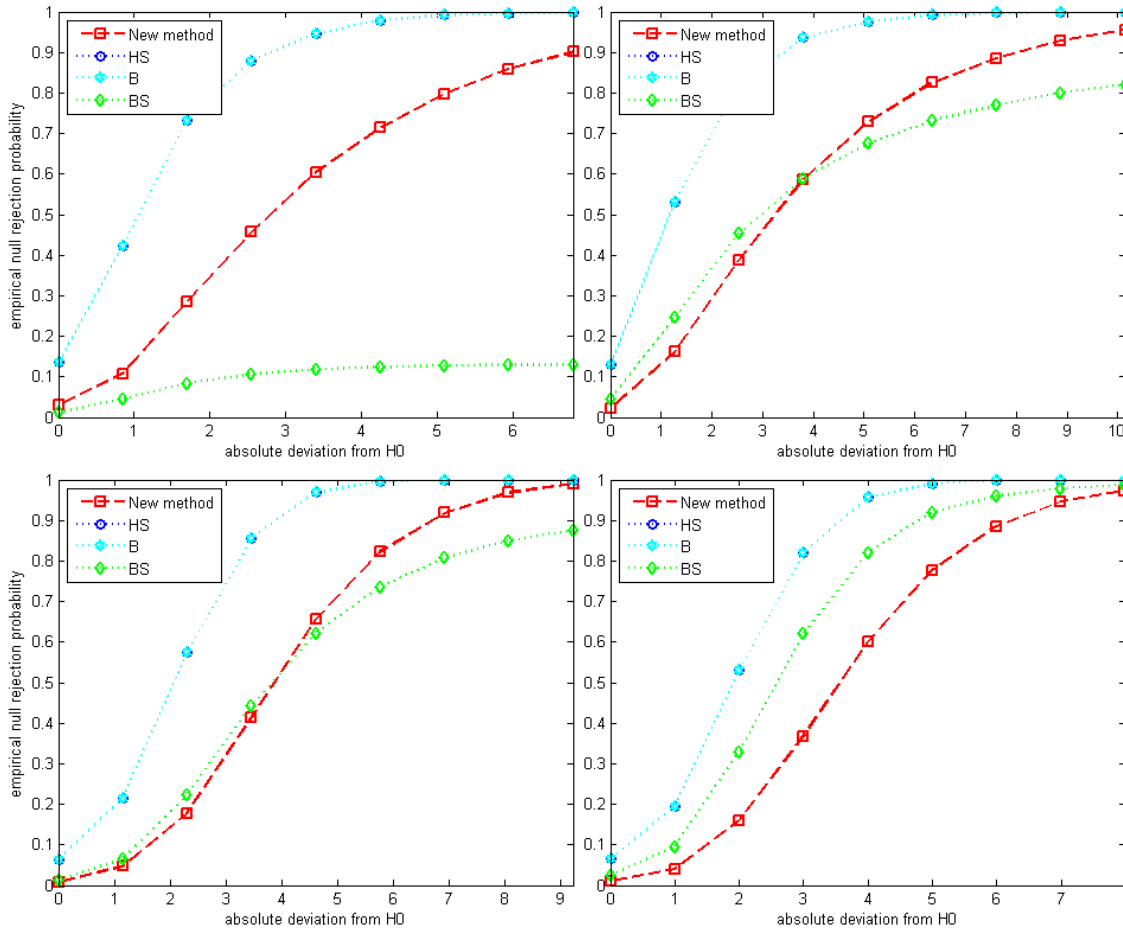


FIGURE 6. Empirical power properties; $p = 0.5$. Top row: $F = \text{Exp}(1)$, one-sample. Bottom: $N(0, 1)$, two-sample. Left column: $n = 3$. Right: $n = 4$.

TABLE 3. Empirical size as percentage, $p = 0.95$, one-sample.

| Distribution | $n = 45$ | | | | $n = 21$ | | | |
|--------------------------------|----------|-------------|-----|------|----------|-------------|------|------|
| | New | BS; $B_1 =$ | | | New | BS; $B_1 =$ | | |
| | | 99 | 999 | HS/B | | 99 | 999 | HS/B |
| $N(0, 1)$ | 3.2 | 6.4 | 6.5 | 10.4 | 6.7 | 11.3 | 11.5 | 25.3 |
| Slash | 5.3 | 7.8 | 8.1 | 10.7 | 12.9 | 17.8 | 18.6 | 30.3 |
| $\text{Logn}(0, \sigma = 3/2)$ | 5.0 | 7.3 | 7.5 | 10.7 | 11.3 | 14.6 | 15.2 | 28.7 |
| $\text{Exp}(1)$ | 4.0 | 6.9 | 7.0 | 10.7 | 8.0 | 12.6 | 13.1 | 26.9 |
| $\text{Uniform}(0, 1)$ | 3.3 | 5.4 | 5.3 | 11.2 | 4.7 | 6.8 | 6.7 | 22.3 |
| $\text{GEV}(0, 1, 0)$ | 4.1 | 7.2 | 7.3 | 11.0 | 7.3 | 12.2 | 12.6 | 25.9 |
| χ_1^2 | 3.6 | 6.5 | 6.4 | 10.2 | 8.7 | 13.5 | 13.9 | 27.5 |

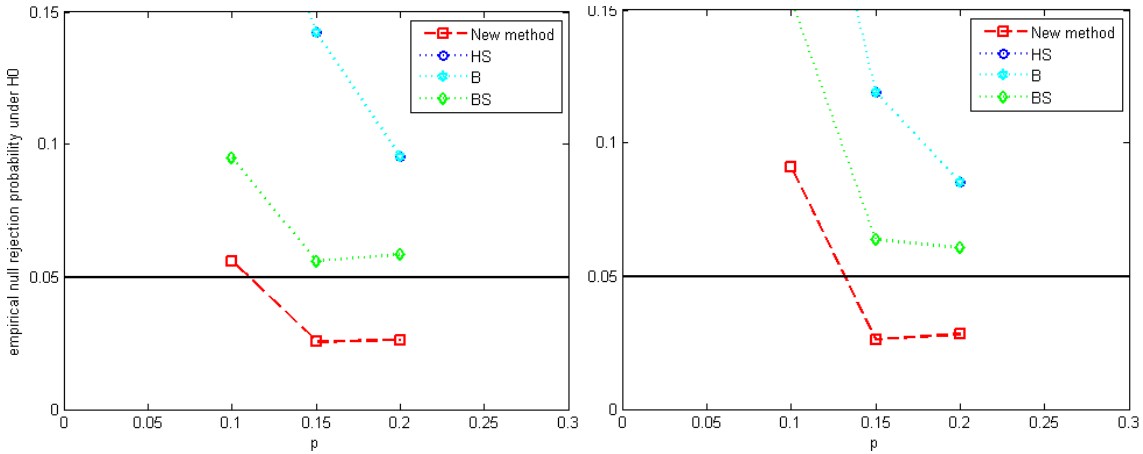


FIGURE 7. Empirical size properties, one-sample, $n = 11$. Left: $N(0, 1)$. Right: slash.

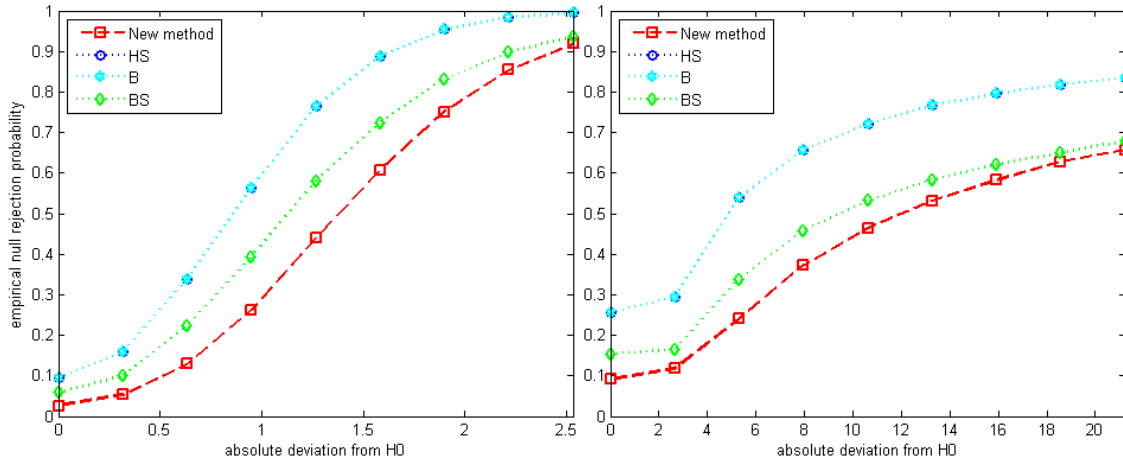


FIGURE 8. Empirical power properties, one-sample, $n = 11$. Left: $N(0, 1)$, $p = 0.2$. Right: slash, $p = 0.1$.

TABLE 4. Empirical size as percentage, $p = 0.95$, two-sample with independent distributions of same shape but differing in variance by factor of 25. Distribution with larger variance is shown.

| Distribution | $n = 21$ | | | $n = 29$ | | | $n = 45$ | | |
|--------------------------------|----------|------|------|----------|-----|------|----------|-----|------|
| | New | BS | HS/B | New | BS | HS/B | New | BS | HS/B |
| $N(0, 1)$ | 6.3 | 9.5 | 21.1 | 3.8 | 6.9 | 13.5 | 3.4 | 5.5 | 8.6 |
| $\text{Logn}(0, \sigma = 3/2)$ | 11.5 | 15.6 | 28.2 | 5.9 | 9.3 | 17.8 | 4.6 | 7.2 | 10.1 |
| $\text{Exp}(1)$ | 7.4 | 10.9 | 21.6 | 4.0 | 7.9 | 14.1 | 3.5 | 6.0 | 8.1 |
| $\text{Uniform}(0, \sqrt{12})$ | 5.6 | 6.6 | 19.8 | 4.3 | 5.7 | 16.0 | 3.9 | 5.2 | 10.7 |

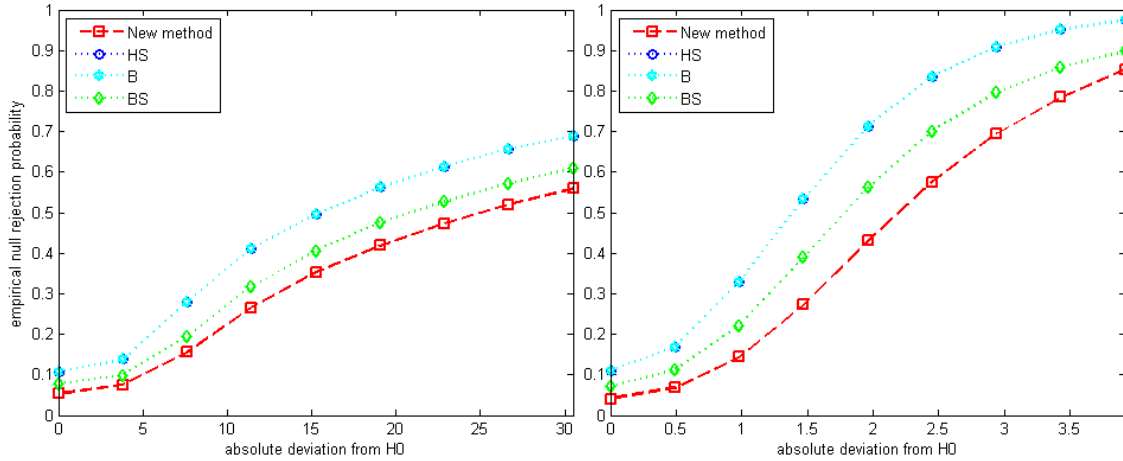


FIGURE 9. Empirical power properties, one-sample. $n = 45$, $p = 0.95$. Left: slash. Right: $GEV(0,1,0)$.

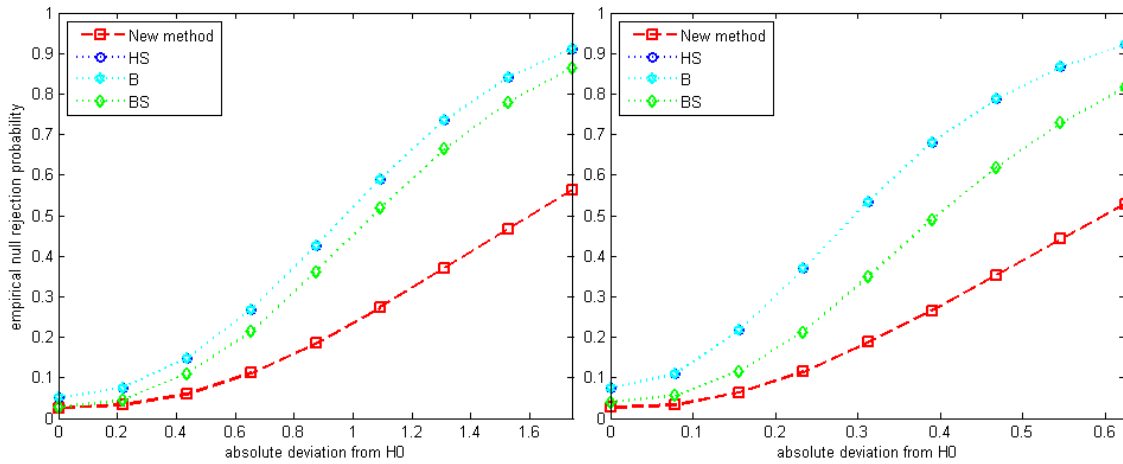


FIGURE 10. Empirical power properties, two-sample. $n = 45$, $p = 0.95$. Left: $N(0,1)$ distributions. Right: $Uniform(0, \sqrt{12})$.

power is competitive with BS even without calibration, as in Figure 9. For the two-sample case with identical distributions, BS always controls size along with the new method. However, moving toward the limiting case where one distribution is a constant, when there is a significant different in variance between the two samples' distributions (factor of 25), BS has significantly worse size distortion than the new method for $n = 21$, and has size distortion while the new method has none for $n = 29$ and $n = 45$; see Table 4. In the case of identical distributions, the new method has power competitive with BS for $n = 21$, but by $n = 45$ BS has significantly better power; see Figures 10 and 11.

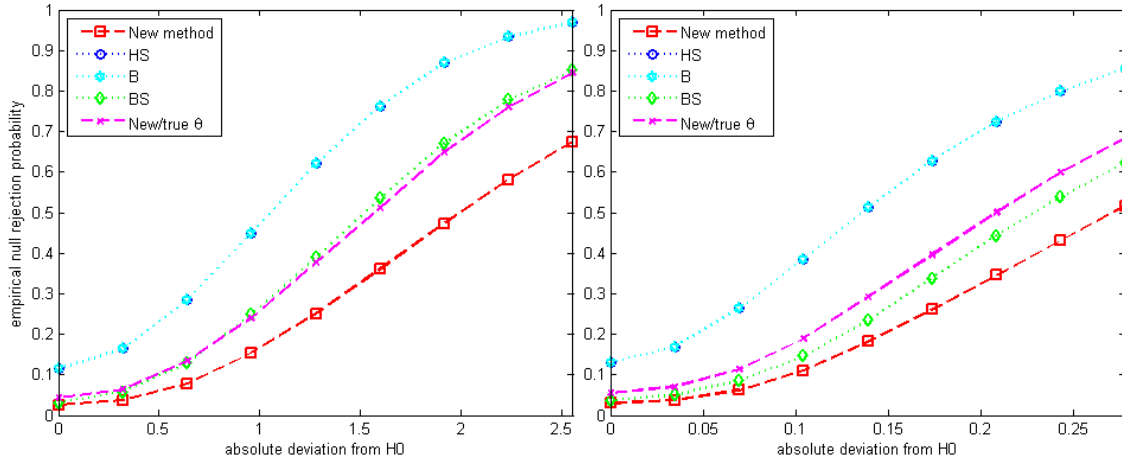


FIGURE 11. Empirical power properties, two-sample. $n = 21$. Left: $N(0, 1)$ distributions, $p = 0.05$. Right: $\text{Exp}(1)$, $p = 0.05$. Better estimators of θ would increase power for the new method.

The new method can't be used as-is when $np < 1$ or $np \geq n - 1$, i.e. when the estimator is the sample minimum or maximum. It would be computationally possible to use a one-sided quantile spacing in those cases, but at that point an extreme quantile framework is more appropriate; see Chernozhukov & Fernández-Val (2011).

For any sample size, it appears that quantiles close enough to zero or one will cause size distortion in HS, B, and BS. For example, even with $n = 250$ and a normal distribution, tests of the $p = 0.005$ quantile for $\alpha = 0.05$ have empirical size 9.6% for BS and 20.7% for HS and B, compared with 5.3% for the new method.

There are two different proximate “causes” of size distortion for HS and B in general, illustrated in Figure 12. When $n = 21$ and $p = 0.2$, $r = \lfloor (21)(0.2) \rfloor + 1 = 5$, which means that m can be at most 4 such that $r - m \geq 1$; and $m_{HS} = m_B = 4$, the biggest m possible. This is the best choice, but still size distorted. For $n = 21$ and any $p < 0.35$, $m_{HS} = m_B = r - 1$, the maximum possible m . However, when $n = 21$ and $p \geq 0.35$, m_{HS} and m_B are strictly less than the maximum possible m . In those cases, HS and B could have chosen a larger m to lessen the size distortion, but didn't. In some cases, both “causes” apply, such as with $F = \text{Exp}(1)$, $n = 71$, $p = 0.2$, and $\alpha = 0.05$, where bigger m_{HS} would have reduced size distortion, but not eliminated it. Focusing on cases when Hutson (1999) can't be computed, the use of standard critical values is always the proximate cause since m_{HS} is always the maximum allowed.

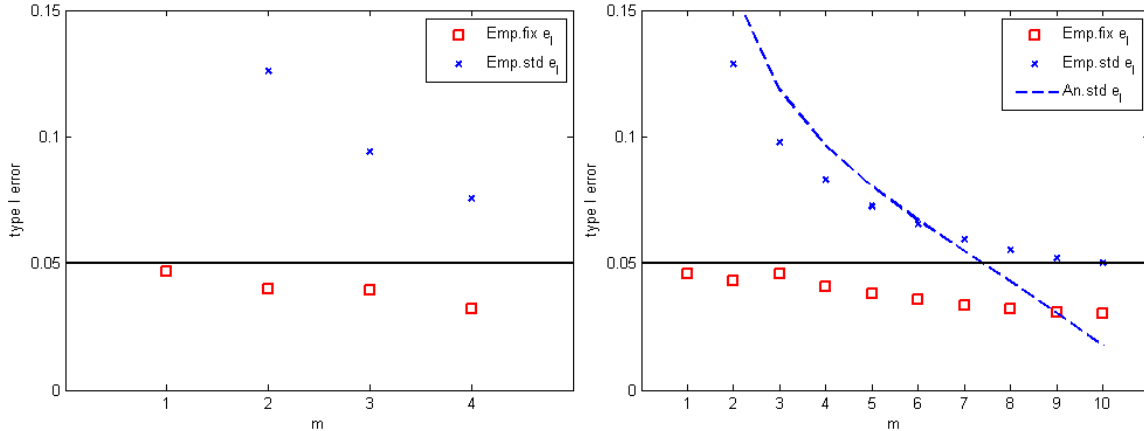


FIGURE 12. Proximate causes of size distortion for HS and B. Empirical type I error (x markers) is plotted against m , for a test with standard normal critical values; missing marker on the left for $m = 1$ is above 0.15. The dashed line in the right plot is the plug-in e_I used to select m_{HS} . Left: no m can control size. Right: approximation error leads to $m_{HS} = 7 < 10$ and size distortion. $F = \text{Uniform}(0, 1)$, $n = 21$, $\alpha = 0.05$, and $p = 0.2$ on the left, $p = 0.5$ on the right; 5000 replications.

Of theoretical (if not directly practical) interest is Figure 13. The new method chooses m to maximize power, and in some cases this produces significantly better power than HS or B. This is true in both the one-sample case (top row) and the two-sample case (bottom row).

Theoretical calculations corroborate the new method’s power advantage over HS in the Figure 13 simulations. Type II error for the new method is given in Proposition 5. As seen in Appendix C, using standard normal critical value $z_{1-\alpha/2}$ instead of $z_{\alpha,m}$ will only alter the m^{-1} term. To compare the new method to HS, only the m^{-1} and $(m/n)^2$ terms are needed since the first-order term and the $n^{-1/2}$ term are identical. The m^{-1} term is always larger for the new method due to the more conservative critical value, while the $(m/n)^2$ term is always larger for HS since $m_{HS} > m_K$.

Specifically, for $\alpha = 0.05$, $C = 1.96$ (equivalent to -1.96 since the Cauchy distribution is symmetric about its median), and $p = 0.5$, the smoothing parameters are $m_K \approx (0.77)n^{2/3}$ and $m_{HS} \approx (0.97)n^{2/3}$, using (14) and (15). As discussed, I rounded m_{HS} to the nearest integer to avoid exacerbating the size distortion already seen in Table 3, for example, while the floor function is used for the new method. For $n = 11$, $m_K = \lfloor 3.8 \rfloor = 3$ and $m_{HS} = \text{round}(4.8) = 5$; for $n = 41$, $m_K = \lfloor 9.2 \rfloor = 9$ and

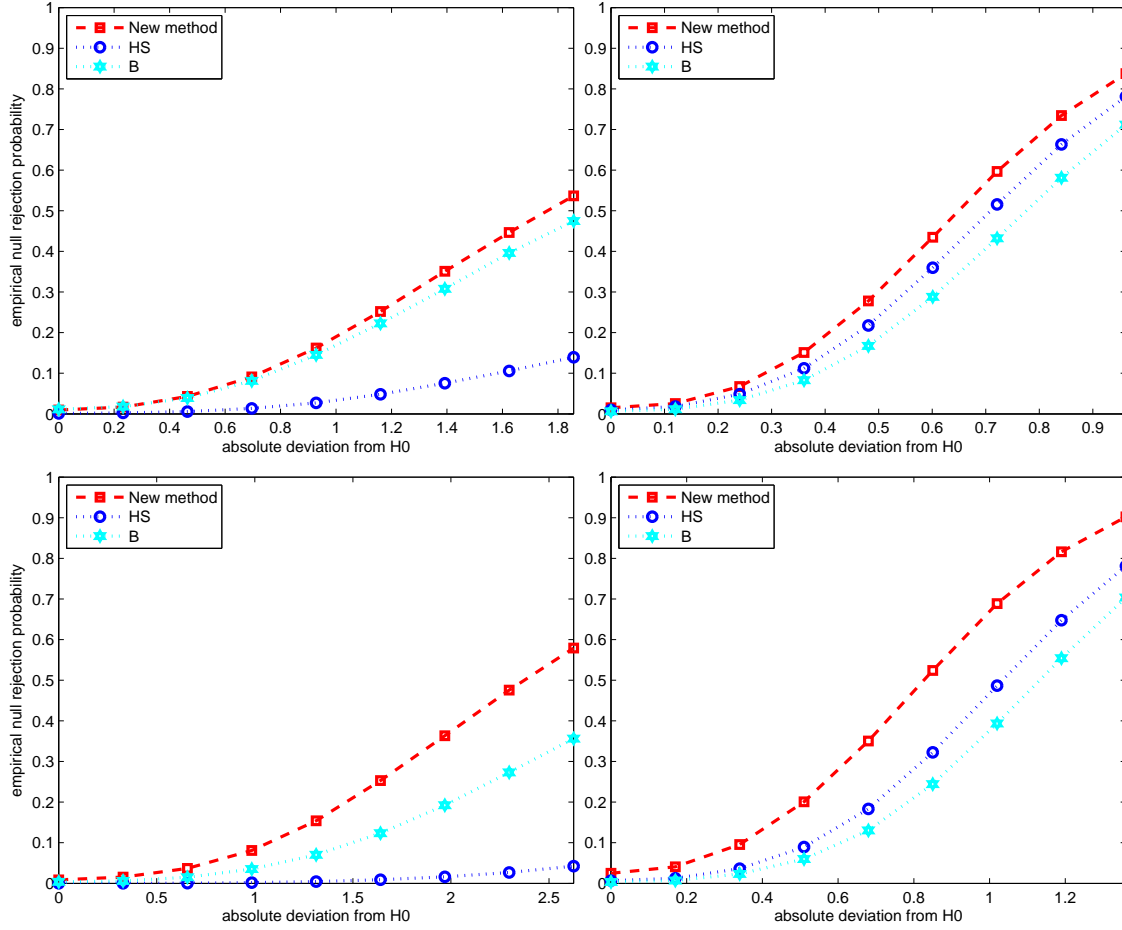


FIGURE 13. Empirical power curves, comparing $n = 11$ (left column) and $n = 41$ (right column). 50,000 replications each, $p = 0.5$, and distribution is Cauchy. Top row: one-sample. Bottom row: two-sample.

$m_{HS} = \text{round}(11.5) = 12$. For the Cauchy distribution, $f(x) = \pi^{-1}(1 + x^2)^{-1}$, $f(0) = \pi^{-1}$, $f'(0) = 0$, and $f''(0) = -2\pi^{-1}$, so

$$\frac{3(f')^2 - ff''}{6f^4} = \pi^2/3.$$

Taking $\phi(1.96 + 1.96) \approx 0$ and $\phi(0) = (2\pi)^{-1/2}$, the new method's type II error has m^{-1} term approximately $(0.751)m_K^{-1}$, while the m^{-1} term for HS is approximately zero. The $(m/n)^2$ term is approximately $(2.57)(m/n)^2$. This means that, up to $o(m^{-1} + m^2/n^2)$ terms, the type II error difference is

$$e_{II}^K - e_{II}^{HS} = m_K^{-1}(0.751) + (m_K^2 - m_{HS}^2)(2.57/n^2).$$

Plugging in $n = 11$, $m_K = 3$, and $m_{HS} = 5$, $e_{II}^K - e_{II}^{HS} = -0.090 < 0$, meaning better power for the new method. With $n = 41$, $m_K = 9$, and $m_{HS} = 12$, $e_{II}^K - e_{II}^{HS} = -0.013 < 0$, again better power for the new method. To compare with Figure 13, note that $\gamma/\sqrt{n} = (C/f(0))\sqrt{p(1-p)/n} \approx 0.9$ for $n = 11$ and $\gamma/\sqrt{n} \approx 0.5$ for $n = 41$. The analytic differences from the m^{-1} and $(m/n)^2$ terms are still smaller than the differences in the top row of Figure 13; with small samples, the $o(m^{-1} + m^2/n^2)$ terms also contribute significantly.

The $\pi^2/3$ coefficient in the $(m/n)^2$ term, based on the Cauchy PDF and its derivatives, is key to having that term outweigh the m^{-1} term. With a standard normal distribution, the coefficient is $\pi/3$, smaller by a factor of π ; for other distributions and/or quantiles, the coefficient may be larger.

In the two-sample case, the results shown and discussed were for independent samples, as the theoretical results assumed. If there is enough negative correlation, size distortion occurs for all methods; if there is positive correlation, power will decrease for all methods. The effect of uncorrelated, dependent samples could go either way. In practice, examining the sample correlation as well as one of the many tests for statistical independence is recommended, if independence is not known a priori.

7. CONCLUSION

This paper proposes new one- and two-sample quantile testing procedures based on the SBG test statistic in (3). Critical values dependent on smoothing parameter m are derived from fixed-smoothing asymptotics. These are more accurate than the conventional standard normal critical values since the fixed- m distribution is shown to be more accurate under both fixed- m and large- m asymptotics. Type I and II errors are approximated using an Edgeworth expansion, and the testing-optimal m_K minimizes type II error subject to control of type I error, up to higher-order terms. Simulations show that, compared with the previous SBG methods from HS88 and Bofinger (1975), the new method greatly reduces size distortion while maintaining good power. The new method also outperforms bootstrap methods for one-sample tests and certain two-sample cases. Consequently, this new method is recommended in one-sample cases when Hutson (1999) is not computable. In two-sample cases when Goldman & Kaplan (2012) is not computable, performance depends on the unknown distributions and warrants further research into both bootstrap methods and the two-sample plug-in method (particularly estimating θ) proposed here. Finally,

theoretical justification for fixed-smoothing asymptotics is provided outside of the time series context; there are likely additional models that may benefit from this perspective.

REFERENCES

- [1] BICKEL, PETER J. & ANAT SAKOV, ‘On the choice of m in the m out of n bootstrap and confidence bounds for extrema.’ *Statistica Sinica*, **18** (3), pp. 967–985, 2008.
- [2] BLOCH, DANIEL A. & JOSEPH L. GASTWIRTH, ‘On a Simple Estimate of the Reciprocal of the Density Function.’ *The Annals of Mathematical Statistics*, **39** (3), pp. 1083–1085, 1968.
- [3] BOFINGER, EVE, ‘Estimation of a density function using order statistics.’ *Australian Journal of Statistics*, **17** (1), pp. 1–7, 1975.
- [4] BROWN, B. M., PETER HALL & G. A. YOUNG, ‘The smoothed median and the bootstrap.’ *Biometrika*, **88** (2), pp. 519–534, 2001.
- [5] CHERNOZHUKOV, VICTOR & IVÁN FERNÁNDEZ-VAL, ‘Inference for Extremal Conditional Quantile Models, with an Application to Market and Birth-weight Risks.’ *The Review of Economic Studies*, **78** (2), pp. 559–589, 2011.
- [6] DASGUPTA, ANIRBAN, *Asymptotic Theory of Statistics and Probability*, New York: Springer, pp.3, 2008.
- [7] DAVID, HERBERT ARON, *Order Statistics*, Wiley series in probability and mathematical statistics: Wiley, 2nd edition, 1981.
- [8] GOH, S. CHUAN & KEITH KNIGHT, ‘Nonstandard Quantile-Regression Inference.’ *Econometric Theory*, **25** (5), pp. 1415–1432, 2009.
- [9] GOH, S. CHUAN, ‘Smoothing choices and distributional approximations for econometric inference.’ Ph.D. dissertation, UC-Berkeley, 2004.
- [10] GOLDMAN, MATT & DAVID M. KAPLAN, ‘IDEAL Quantile Inference via Interpolated Duals of Exact Analytic L -statistics.’ 2012. Working paper.
- [11] HALL, PETER & SIMON J. SHEATHER, ‘On the Distribution of a Studentized Quantile.’ Technical Report 71, Department of Statistics, Pennsylvania State University, 1987.
- [12] ——— ‘On the Distribution of a Studentized Quantile.’ *Journal of the Royal Statistical Society. Series B (Methodological)*, **50** (3), pp. 381–391, 1988.
- [13] HO, YVONNE H. S. & STEPHEN M. S. LEE, ‘Iterated Smoothed Bootstrap Confidence Intervals for Population Quantiles.’ *The Annals of Statistics*, **33** (1), pp. 437–462, 2005.
- [14] HUTSON, ALAN D., ‘Calculating nonparametric confidence intervals for quantiles using fractional order statistics.’ *Journal of Applied Statistics*, **26** (3), pp. 343–353, 1999.
- [15] KAPLAN, DAVID M., ‘Inference on quantiles: confidence intervals, p-values, and testing.’ MATLAB Central File Exchange, <http://www.mathworks.com/matlabcentral/fileexchange/>

TABLE 5. Simulated rejection probabilities (%) for different fixed- m critical value approximations, $\alpha = 5\%$.

| m | rejection probability for critical value | | | |
|----|--|--------------------|--------------------|------|
| | Goh (2004, simulated) | including m^{-1} | including m^{-2} | t |
| 1 | 4.93 | 8.21 | 5.84 | 6.87 |
| 2 | 5.03 | 6.25 | 5.20 | 5.47 |
| 3 | 5.05 | 5.70 | 5.10 | 5.28 |
| 4 | 5.15 | 5.46 | 5.06 | 5.21 |
| 5 | 4.92 | 5.23 | 4.96 | 5.08 |
| 10 | 5.01 | 5.14 | 5.06 | 5.14 |
| 20 | 5.04 | 5.01 | 4.99 | 5.04 |

32410-quantile-inference-testing-univariate-and-bivariate, August, 2011. Retrieved 25 March 2012.

- [16] ——— ‘IDEAL inference on conditional quantiles.’ 2012. Working paper.
- [17] KOENKER, ROGER & ZHIJIE XIAO, ‘Inference on the Quantile Regression Process.’ *Econometrica*, **70** (4), pp. 1583–1612, 2002.
- [18] MOSTELLER, F., ‘On Some Useful “Inefficient” Statistics.’ *Annals of Mathematical Statistics*, **17**, pp. 377–408, 1946.
- [19] SIDDIQUI, MOHAMMED M., ‘Distribution of quantiles in samples from a bivariate population.’ *J. Research of the NBS, B. Math. and Math. Physics*, **64B** (3), pp. 145–150, 1960.
- [20] SUN, YIXIAO, PETER C. B. PHILLIPS & SAINAN JIN, ‘Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing.’ *Econometrica*, **76** (1), pp. 175–194, 2008.

APPENDIX A. ACCURACY OF APPROXIMATION OF FIXED- m CRITICAL VALUES

The approximate fixed- m critical value in (7) is quite accurate for all but $m = 1, 2$, as Table 5 shows. The second approximation alternative adds the $O(m^{-2})$ term to the approximation. The third alternative uses the critical value from the Student’s t -distribution with the degrees of freedom chosen to match the variance. To compare, for various m and α , I simulated the two-sided rejection probability for a given critical value given $T_{m,\infty}$ from (4) as the true distribution. One million simulation replications per m and α were run; to gauge simulation error, I also include in the table the critical values given in Goh (2004) (who ran 500,000 replications each). Additional values of m and α are available in the online appendix.

APPENDIX B. SIGN OF $u_{3,0}(z)$ FOR COMMON DISTRIBUTIONS

Here, I only show the work for the normal distribution; the method is the same for the other distributions. Work for t -, exponential, χ^2 , and Fréchet distributions is in the supplemental appendix. For the uniform distribution, $u_{3,0} = 0$ since $f' = 0$ and $f'' = 0$.

Starting with the pdf and derivatives of the normal distribution,

$$\begin{aligned} f_n(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}, & f'_n(x) &= -(x-\mu)/\sigma^2 \cdot f_n(x) \\ f''_n(x) &= -f_n(x)/\sigma^2 - f'_n(x)(x-\mu)/\sigma^2 = -f_n(x)/\sigma^2 - [-(x-\mu)/\sigma^2 \cdot f_n(x)](x-\mu)/\sigma^2 \\ &= -f_n(x)/\sigma^2 + (x-\mu)^2/\sigma^4 \cdot f_n(x) = f_n(x)[(x-\mu)^2/\sigma^4 - 1/\sigma^2], \\ 3f'_n(x)^2 - f_n(x)f''_n(x) &= 3[-(x-\mu)/\sigma^2 \cdot f_n(x)]^2 - f_n(x) \cdot f_n(x)[(x-\mu)^2/\sigma^4 - 1/\sigma^2] \\ &= [f_n(x)]^2 \{2(x-\mu)^2/\sigma^4 + 1/\sigma^2\} \geq 0, \end{aligned}$$

for any μ and σ^2 . Thus the sign of $u_{3,0}(z)$ is always the sign of z for the normal distribution at any quantile.

APPENDIX C. CALCULATION OF TYPE II ERROR (PROPOSITION 5)

The type II error is the probability of not rejecting when H_0 is false. For a two-sided symmetric test, this is

$$P(|T_{m,n}| < z) = P(T_{m,n} < z) - P(T_{m,n} < -z).$$

Letting the corrected critical value $z_{\alpha,m} = z_{1-\alpha/2} + z_{1-\alpha/2}^3/(4m)$ as in (7), and expanding via (11) and (10), for $C > 0$,

$$\begin{aligned} P(|T_{m,n}| < z) &= L^+ + H_1^+ - H_2^+ + O(n^{-1/2}) + o(m^{-1} + (m/n)^2) \quad \text{with} \\ L^+ &= \Phi(z_{\alpha,m} + C) - \Phi(-z_{\alpha,m} + C), \\ H_1^+ &= \phi(z_{\alpha,m} + C) \left[m^{-1} u_{2,\gamma}(z_{\alpha,m} + C) + (m/n)^2 u_{3,\gamma}(z_{\alpha,m} + C) \right], \\ H_2^+ &= \phi(-z_{\alpha,m} + C) \left[m^{-1} u_{2,\gamma}(C - z_{\alpha,m}) + (m/n)^2 u_{3,\gamma}(C - z_{\alpha,m}) \right]. \end{aligned}$$

I write $O(n^{-1/2})$ for the $n^{-1/2}$ terms since they do not depend on m and thus do not affect the optimization problem for selecting m . Define L^- , H_1^- , and H_2^- similarly but with $-C < 0$ instead of $C > 0$, and thus $-\gamma = -C\sqrt{p(1-p)}/f(\xi_p)$ instead of

γ :

$$\begin{aligned} L^- &= \Phi(z_{\alpha,m} - C) - \Phi(-z_{\alpha,m} - C), \\ H_1^- &= \phi(z_{\alpha,m} - C) \left[m^{-1} u_{2,-\gamma}(z_{\alpha,m} - C) + (m/n)^2 u_{3,-\gamma}(z_{\alpha,m} - C) \right], \\ H_2^- &= \phi(-z_{\alpha,m} - C) \left[m^{-1} u_{2,-\gamma}(-C - z_{\alpha,m}) + (m/n)^2 u_{3,-\gamma}(-C - z_{\alpha,m}) \right]. \end{aligned}$$

I calculate average power where the alternatives $+C$ and $-C$ each have 0.5 probability. Using $\phi(-x) = \phi(x)$,

$$(17) \quad \begin{aligned} P(|T_{m,n}| < z_{\alpha,m}) &= \frac{1}{2} \left\{ (L^+ + L^-) + (H_1^+ + H_1^-) - (H_2^+ + H_2^-) \right. \\ &\quad \left. + O(n^{-1/2}) + o(m^{-1} + (m/n)^2) \right\}. \end{aligned}$$

For the first-order term L^+ ,

$$(18) \quad \begin{aligned} &\Phi(z_{\alpha,m} + C) - \Phi(-z_{\alpha,m} + C) \\ &= \Phi(C + z_{1-\alpha/2} + z_{1-\alpha/2}^3/(4m)) - \Phi(C - z_{1-\alpha/2} - z_{1-\alpha/2}^3/(4m)) \\ &= 0.5 + m^{-1} \frac{1}{4} z_{1-\alpha/2}^3 [\phi(z_{1-\alpha/2} + C) + \phi(C - z_{1-\alpha/2})] \end{aligned}$$

since in Proposition 5, C solves $0.5 = \Phi(z_{1-\alpha/2} + C) - \Phi(-z_{1-\alpha/2} + C)$. Since the fixed- m critical value is larger than the standard normal critical value, this term contributes additional type II error in the m^{-1} term. Similarly,

$$\begin{aligned} L^- &= \Phi(z_{\alpha,m} - C) - \Phi(-z_{\alpha,m} - C) \\ &= 0.5 + m^{-1} \frac{1}{4} z_{1-\alpha/2}^3 [\phi(C - z_{1-\alpha/2}) + \phi(-z_{1-\alpha/2} - C)] = L^+. \end{aligned}$$

Within the m^{-1} and $(m/n)^2$ terms, anything $o(1)$ will end up in the $o(m^{-1} + (m/n)^2)$ remainder. Thus, for those terms,

$$z_{\alpha,m} = z_{1-\alpha/2} + O(m^{-1}), \quad \phi(z_{\alpha,m} + C) = \phi(z_{1-\alpha/2} + C) + O(m^{-1}).$$

For the m^{-1} terms, since $\phi(x) = \phi(-x)$ and $C \equiv \gamma f(\xi_p) / \sqrt{p(1-p)}$, and letting $d_1 \equiv z_{\alpha,m} + C$, $d_2 \equiv z_{\alpha,m} - C$,

$$\begin{aligned} &u_{2,\gamma}(d_1) - u_{2,-\gamma}(-d_1) \\ &= -\frac{1}{4} d_1^3 - \frac{1}{4} C^2 d_1 + \frac{1}{4} 2C d_1^2 - \left[-\frac{1}{4} (-d_1)^3 - \frac{1}{4} (-C)^2 (-d_1) + \frac{1}{4} 2(-C)(-d_1)^2 \right] \end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{2}(d_1^3 + C^2 d_1 - 2C d_1^2) = -\frac{1}{2}(z_{1-\alpha/2} + C)z_{1-\alpha/2}^2 + O(m^{-1}), \\
u_{2,-\gamma}(d_2) - u_{2,\gamma}(-d_2) &= -\frac{1}{2}(d_2^3 + C^2 d_2 + 2C d_2^2) = -\frac{1}{2}(z_{1-\alpha/2} - C)z_{1-\alpha/2}^2 + O(m^{-1}), \\
\phi(z_{\alpha,m} + C)m^{-1}(u_{2,\gamma}(z_{\alpha,m} + C) - u_{2,-\gamma}(-z_{\alpha,m} - C)) \\
&+ \phi(z_{\alpha,m} - C)m^{-1}(u_{2,-\gamma}(z_{\alpha,m} - C) - u_{2,\gamma}(-z_{\alpha,m} + C)) \\
&= -\frac{1}{2}m^{-1}[\phi(z_{1-\alpha/2} + C)(z_{1-\alpha/2} + C)z_{1-\alpha/2}^2 + \phi(z_{1-\alpha/2} - C)(z_{1-\alpha/2} - C)z_{1-\alpha/2}^2] \\
(19) \quad &+ O(m^{-2}).
\end{aligned}$$

For the $(m/n)^2$ terms, writing $f(\xi_p)$ and its derivatives as f , f' , f'' , and letting $M \equiv [3(f')^2 - ff'']/(6f^4)$,

$$\begin{aligned}
&u_{3,\gamma}(z_{\alpha,m} + C) - u_{3,-\gamma}(-z_{\alpha,m} - C) \\
&= M(z_{\alpha,m} + C - C) - M(-z_{\alpha,m} - C - (-C)) = 2Mz_{1-\alpha/2} + O(m^{-1}), \\
&u_{3,-\gamma}(z_{\alpha,m} - C) - u_{3,\gamma}(-z_{\alpha,m} + C) \\
&= M(z_{\alpha,m} - C - (-C)) - M(-z_{\alpha,m} + C - C) = 2Mz_{1-\alpha/2} + O(m^{-1}), \\
&\frac{m^2}{n^2} \left\{ \phi(z_{\alpha,m} + C)[u_{3,\gamma}(z_{\alpha,m} + C) - u_{3,-\gamma}(-z_{\alpha,m} - C)] \right. \\
&\quad \left. + \phi(z_{\alpha,m} - C)[u_{3,-\gamma}(z_{\alpha,m} - C) - u_{3,\gamma}(-z_{\alpha,m} + C)] \right\} \\
(20) \quad &= (m/n)^2 2 \frac{3(f')^2 - ff''}{6f^4} z_{1-\alpha/2} [\phi(z_{1-\alpha/2} + C) + \phi(z_{1-\alpha/2} - C)] + O(m/n^2).
\end{aligned}$$

Combining (18), (19), and (20), (17) becomes

$$\begin{aligned}
P(|T_{m,n}| < z) &= \frac{1}{2} \left\{ 2(0.5 + m^{-1}(1/4)z_{1-\alpha/2}^3 [\phi(C + z_{1-\alpha/2}) + \phi(C - z_{1-\alpha/2})]) \right. \\
&\quad - 2(1/4)m^{-1}[\phi(z_{1-\alpha/2} + C)(z_{1-\alpha/2} + C)z_{1-\alpha/2}^2 \\
&\quad \quad \quad + \phi(z_{1-\alpha/2} - C)(z_{1-\alpha/2} - C)z_{1-\alpha/2}^2] \\
&\quad \left. + 2(m/n)^2 \frac{3(f')^2 - ff''}{6f^4} z_{1-\alpha/2} [\phi(z_{1-\alpha/2} + C) + \phi(z_{1-\alpha/2} - C)] \right\} \\
&\quad + O(n^{-1/2}) + o(m^{-1} + (m/n)^2),
\end{aligned}$$

which simplifies to the forms given in Proposition 5.

APPENDIX D. PROOF SKETCH OF EDGEWORTH EXPANSION (THEOREM 1) OF
 SBG TEST STATISTIC UNDER LOCAL ALTERNATIVE HYPOTHESIS

This is a sketch comparable to Hall & Sheather (1987, , “HS87” hereafter); full proof is in the supplemental appendix. The structure here follows HS88, taking their results as given (proofs in my full version). Full references for citations found only in this proof are in HS88. This proof of Theorem 1 uses the same setup and definitions as in the main text.

As in Section 2, the null hypothesis is $H_0 : \xi_p = \beta$, and the true $\xi_p = \beta - \gamma/\sqrt{n}$. I continue from (11), which showed that

$$P(T_{m,n} < z) = P\left(\frac{\sqrt{n}(X_{n,r} - \xi_p) + \gamma(S_{m,n}/S_0 - 1)}{S_{m,n}\sqrt{p(1-p)}} < z + C\right),$$

where $C \equiv \gamma f(\xi_p)/\sqrt{p(1-p)}$, $S_0 \equiv 1/f(\xi_p)$. I want to derive a higher-order expansion around the (shifted) standard normal distribution.

As in HS88, I first deal with centering $X_{n,r} = X_{n,[np]+1}$ since as n increases, $[np]$ increases in discrete steps of one. So instead of $X_{n,r} - \xi_p$ there is $X_{n,r} - \eta_p$, where

$$(21) \quad \eta_p \equiv F^{-1}\left[\exp\left(-\sum_r^n j^{-1}\right)\right].$$

This centering effect is no different when $\gamma \neq 0$; HS88 (p. 389) conclude that the effect is, defining $w_n \equiv [\epsilon_n - 1 + \frac{1}{2}(1-p)]/[p(1-p)]^{1/2}$,

$$n^{1/2}(X_{n,r} - \xi_p)/\hat{\tau} = n^{1/2}(X_{n,r} - \eta_p)/\hat{\tau} + n^{-1/2}w_n + O_p[n^{-1/2}m^{-1/2} + (m/n)^2n^{-1/2}].$$

Then the remainder $n^{-1/2}m^{-1/2} + (m/n)^2n^{-1/2} = o[m^{-1} + (m/n)^2]$, and so

$$\begin{aligned} P[n^{1/2}(X_{n,r} - \xi_p)/\tau \leq z] &= P[n^{1/2}(X_{n,r} - \eta_p)/\tau \leq z - n^{-1/2}w_n] + O(n^{-3/2}) \\ &= P[n^{1/2}(X_{n,r} - \eta_p)/\tau \leq z] - n^{-1/2}w_n\phi(z) + O(n^{-1}), \end{aligned}$$

$$(22) \quad P\left[\frac{n^{1/2}(X_{n,r} - \xi_p) + \gamma(S_{m,n}/S_0 - 1)}{\hat{\tau}} \leq z\right] = P\left[\frac{n^{1/2}(X_{n,r} - \eta_p) + \gamma(S_{m,n}/S_0 - 1)}{\hat{\tau}} \leq z\right] - n^{-1/2}w_n\phi(z) + o[m^{-1} + (m/n)^2].$$

As HS88 note, this is the so-called ‘delta method’ for deducing Edgeworth expansions; see Bhattacharya and Ghosh (1978, p. 438). Thus, instead of proving (10), the

theorem follows by proving

$$(23) \quad \left| \sup_{-\infty < z < \infty} P \left[\frac{n^{1/2}(X_{n,r} - \eta_p) + \gamma(S_{m,n}/S_0 - 1)}{\hat{\tau}} \leq z \right] - [\Phi(z) + n^{-1/2}u_{1,\gamma}^\dagger(z)\phi(z) + m^{-1}u_{2,\gamma}(z)\phi(z) + (m/n)^2u_{3,\gamma}(z)\phi(z)] \right| = o[m^{-1} + (m/n)^2],$$

where

$$\begin{aligned} u_{1,\gamma}^\dagger(z) &\equiv \frac{1}{6} \left(\frac{p}{1-p} \right)^{1/2} \frac{1+p}{p} (z^2 - 1) - \frac{\gamma f(\xi_p)}{p} \left(1 - \frac{p f'(\xi_p)}{[f(\xi_p)]^2} - \frac{1}{2(1-p)} \right) z \\ &\quad - \frac{1}{2} \left(\frac{p}{1-p} \right)^{1/2} \left(1 + \frac{f'(\xi_p)}{[f(\xi_p)]^2} (1-p) \right) z^2, \\ u_{2,\gamma}(z) &\equiv \frac{1}{4} \left[\frac{2\gamma f(\xi_p)}{[p(1-p)]^{1/2}} z^2 - \frac{\gamma^2 [f(\xi_p)]^2}{p(1-p)} z - z^3 \right], \\ u_{3,\gamma}(z) &\equiv \frac{f''(\xi_p) f(\xi_p) + 3[f'(\xi_p)]^2}{6[f(\xi_p)]^4} \left(\frac{\gamma f(\xi_p)}{[p(1-p)]^{1/2}} - z \right). \end{aligned}$$

Now to prove the above. As defined in HS88, let

$$(24) \quad G \equiv F^{-1}, g \equiv G', \text{ and } H(x) \equiv F^{-1}(e^{-x}),$$

and let W_1, W_2, \dots , be independent exponential random variables with unit mean. The sequence $\{X_{n,s}\}_{s=1}^n$ has the same joint distribution as $\{H(\sum_{s \leq j \leq n} j^{-1} W_j)\}_{s=1}^n$ (e.g. David, 1981, p. 21), and in a slight abuse of notation suggested by HS88, write $X_{n,s} = H(\sum_{s \leq j \leq n} j^{-1} W_j)$. Let $V_j \equiv W_j - 1$,

$$(25) \quad \begin{aligned} \Delta_1 &\equiv \sum_{r-m}^{r-1} j^{-1} V_j, \quad \Delta_2 \equiv \sum_r^{r+m-1} j^{-1} V_j, \quad \Delta_3 \equiv \sum_{r+m}^n j^{-1} V_j, \quad a_k \equiv H^{(k)} \left(\sum_r^n j^{-1} \right), \\ Z &\equiv [p(1-p)]^{1/2} [n^{1/2}(X_{n,r} - \eta_p) + \gamma(S_{m,n}/S_0 - 1)] / \hat{\tau} \\ &= [n^{1/2}(X_{n,r} - \eta_p) + \gamma(S_{m,n}/S_0 - 1)] [(n/2m)(X_{n,r+m} - X_{n,r-m})]^{-1}. \end{aligned}$$

Note that Δ_1 and Δ_2 are $O_p(n^{-1}m^{1/2})$ and Δ_3 is $O_p(n^{-1/2})$. With the above and Taylor expansions, $Z = Y + R$ where

$$(26) \quad \begin{aligned} Y &\equiv -pn^{1/2}[(1+\delta)(\Delta_2 + \Delta_3) + B], \quad \delta \equiv -(m/n)^2 g''(p)[6g(p)]^{-1}, \\ B &\equiv \delta\Psi + (n/m)(b_1\Delta_1 + b_2\Delta_2)\Delta_2 + (n/m)b_3(\Delta_1 + \Delta_2)(\Delta_3 + \Psi) \end{aligned}$$

$$(27) \quad + (n/m)^2 b_4 (\Delta_1 + \Delta_2)^2 (\Delta_3 + \Psi) + b_5 \Delta_3 (\Delta_3 + 2\Psi),$$

$$b_1 \equiv -p/2, b_2 \equiv -p/2, b_3 \equiv -p/2 - (m/n)(a_2/2a_1), b_4 \equiv (p/2)^2, b_5 \equiv -a_2/2a_1,$$

$$\Psi \equiv \gamma/(pg(p)\sqrt{n}), \quad \text{and}$$

$$(28) \quad R = O_p[n^{-1/2}m^{-1/2} + n^{-3/2}m + m^{-3/2} + (m/n)^{2+\epsilon}].$$

Here, Y is of the same form as HS88, but with Ψ now additionally showing up in the higher-order B terms. In the definition of Z , γ only appears in the numerator, not in the denominator. From the stochastic expansion of $S_{m,n}$, which is already required for the denominator of Z , $S_{m,n}/S_0 = 1 + \nu$, where ν contains the higher-order terms that are dropped for the first-order asymptotic result,

$$(29) \quad \nu = \frac{n}{2m}p(\Delta_1 + \Delta_2) + \frac{a_2}{2a_1}(\Delta_1 + \Delta_2 + 2\Delta_3) + (m/n)^2 \frac{g''(p)}{6g(p)}$$

$$+ O_p((m/n)^{2+\epsilon} + n^{-1/2}m^{-1/2} + mn^{-3/2}).$$

Thus, γ enters only in higher-order terms, through the numerator of Z .

Note the Taylor expansion, with sums all over some common range,

$$(30) \quad H\left(\sum W_j/j\right) - H\left(\sum 1/j\right) = \left(\sum V_j/j\right)H'\left(\sum 1/j\right)$$

$$+ (1/2)\left(\sum V_j/j\right)^2 H''\left(\sum 1/j\right) + \dots$$

For the numerator of Z , applying the expansion above yields

$$(31) \quad X_{n,r} - \eta_p = \left(\sum_{j=r}^n V_j/j\right)H'\left(\sum_r^n 1/j\right) + \frac{1}{2}\left(\sum_{j=r}^n V_j/j\right)^2 H''\left(\sum_r^n 1/j\right) + O_p(n^{-3/2})a_3$$

$$= (\Delta_2 + \Delta_3)a_1 + \frac{1}{2}(\Delta_2 + \Delta_3)^2 a_2 + O_p(n^{-3/2}),$$

since $X_{n,r} = H(\sum_{j \geq r} W_j/j)$, $\eta_p \equiv H(\sum_r^n 1/j)$, and $a_3 = O(1)$. Including γ as described earlier,

$$(32) \quad n^{1/2}(X_{n,r} - \eta_p) + \gamma(S_{m,n}/S_0 - 1)$$

$$= \sqrt{n}[(\Delta_2 + \Delta_3)a_1 + (1/2)(\Delta_2 + \Delta_3)^2 a_2 + O_p(n^{-3/2}) + \gamma\nu].$$

The denominator of Z does not include γ , so the stochastic expansion is identical to the intermediate results from HS87,

$$(33) \quad X_{n,r+m} - H\left(\sum_{r+m}^n j^{-1}\right) - \left\{X_{n,r-m} - H\left(\sum_{r-m}^n j^{-1}\right)\right\} \\ = -a_1(\Delta_1 + \Delta_2) - (m/np)a_2(\Delta_1 + \Delta_2 + 2\Delta_3) + O_p(n^{-3/2}m^{1/2} + n^{-5/2}m^2),$$

$$(34) \quad H\left(\sum_{r+m}^n j^{-1}\right) - H\left(\sum_{r-m}^n j^{-1}\right) \\ = (2m/n)[g(p) + (1/6)(m/n)^2g''(p) + O\{(m/n)^{2+\epsilon} + n^{-1}\}].$$

When $\gamma = 0$, plugging in (31), (33), and (34) yields

$$(35) \quad Z = -pn^{1/2}((\Delta_2 + \Delta_3)(-a_1/pg(p)) - \frac{1}{2pg(p)}(\Delta_2 + \Delta_3)^2a_2 + O_p(n^{-3/2})) \\ \times \left[\frac{n}{2m}(- (a_1/g(p))(\Delta_1 + \Delta_2) - (m/n)(a_2/pg(p))(\Delta_1 + \Delta_2 + 2\Delta_3) \right. \\ \left. + O_p(n^{-3/2}m^{1/2} + n^{-5/2}m^2)) \right. \\ \left. + 1 + (m/n)^2 \frac{g''(p)}{6g(p)} + O\{(m/n)^{2+\epsilon} + n^{-1}\} \right]^{-1} \\ \equiv -pn^{1/2}\Theta(1 + \nu)^{-1} = -pn^{1/2}\Theta(1 - \nu + \nu^2 + O(\nu^3)),$$

defining ν (same as above) and Θ for ease of notation.

If (32) (with the new γ term) is used instead of (31), and noting that above I had divided through by $g(p)$ in addition to factoring $-pn^{1/2}$ out of the numerator,

$$(36) \quad Z = [-pn^{1/2}\Theta + \gamma\nu][1 + \nu]^{-1} = -pn^{1/2}(\Theta - \Psi\nu)(1 + \nu)^{-1} \\ = -pn^{1/2}[\Theta - (\Theta + \Psi)\nu + (\Theta + \Psi)\nu^2 + O(n^{-1/2}\nu^3)],$$

with $\Psi \equiv \gamma/(pg(p)\sqrt{n})$ as in (27). Θ and Ψ are both $O_p(n^{-1/2})$, and $\nu = O_p(m^{-3/2})$, which is in R , so the remainder works out.

As given in HS88, with restrictions $n^\eta \leq m \leq n^{1-\eta}$ and $0 < \epsilon \leq 1/6$, (28) implies

$$(37) \quad R = O_p\{[m^{-1} + (m/n)^2]n^{-\epsilon\eta}\},$$

which they state may be strengthened to

$$P\{|R| > [m^{-1} + (m/n)^2]n^{-\zeta}\} = O(n^{-\lambda}), \text{ all } 0 < \zeta < \epsilon\eta \text{ and all } \lambda > 0,$$

since by Markov's and Rosenthal's inequalities (Burkholder (1973), p. 40),

$$\sum_{i=1}^2 P(|nm^{-1/2}\Delta_i| > n^\epsilon) + P(|n^{1/2}\Delta_3| > n^\epsilon) = O(n^{-\lambda}), \text{ all } \epsilon > 0 \text{ and all } \lambda > 0.$$

Thus, the theorem follows from proving the result for Y rather than for Z . As HS88 put it, "Since Δ_1, Δ_2 and Δ_3 are independent and have smooth distributions, it is elementary (but tedious¹⁰) to estimate the difference between $\exp(-t^2/2)$ and the characteristic function of Y , and then derive an Edgeworth expansion by following classical arguments, as in Petrov (1975) or Bhattacharya and Ghosh (1978). To simply identify terms in the expansion here, only moments of Y are needed," the result of which is, as stated in (3.2) of HS87 (it turns out that the difference with $\gamma \neq 0$ is that B contains more terms, but B enters the same way in these equations),

$$(38) \quad E[(-p^{-1}Y)^\ell] = z_1(\ell) + z_2(\ell) + z_3(\ell) + O\{m^{-3/2} + m^{-1/2}(m/n)^2\},$$

$$z_1 \equiv n^{\ell/2}E\{(1 + \delta)(\Delta_2 + \Delta_3)\}^\ell, \quad z_2 \equiv \ell n^{\ell/2}E\{(\Delta_2 + \Delta_3)^{\ell-1}B\},$$

$$z_3 \equiv \frac{1}{2}\ell(\ell - 1)n^{\ell/2}E\{(\Delta_2 + \Delta_3)^{\ell-2}B^2\}.$$

With the new Ψ terms, letting $D_i \equiv n^{1/2}\Delta_i$ to follow the notation of HS88,

$$(39) \quad z_2 = \ell \left\{ n^{-1/2} [p^{-2}b_2E(D_3^{\ell-1}) + 2b_5E(D_3^\ell)\Psi\sqrt{n}] \right.$$

$$+ 2m^{-1}p^{-2}b_4[E(D_3^\ell) + E(D_3^{\ell-1})\Psi\sqrt{n}] + n^{-1/2}b_5E(D_3^{\ell+1}) + \delta E(D_3^{\ell-1})\Psi\sqrt{n} \left. \right\}$$

$$+ \ell(\ell - 1)n^{-1/2}p^{-2}b_3(E(D_3^{\ell-1}) + E(D_3^{\ell-2})\Psi\sqrt{n}) + O(mn^{-3/2} + m^{-1/2}n^{-1/2}),$$

$$z_3 = \ell(\ell - 1)m^{-1}b_3^2p^{-2}[E(D_3^\ell) + 2E(D_3^{\ell-1})\Psi\sqrt{n} + E(D_3^{\ell-2})\Psi^2n]$$

$$(40) \quad + O(m^{-1/2}n^{-1/2} + m^{-3/2} + (m/n)^4).$$

Noting that $b_3 = b_2 + O(m/n) = -(p/2) + O(m/n)$ and $b_3^2 = b_4 + O(m/n) = (p/2)^2 + O(m/n)$,

$$z_2 + z_3 = \ell \left\{ n^{-1/2} [p^{-2}b_2E(D_3^{\ell-1}) + 2b_5E(D_3^\ell)\Psi\sqrt{n}] \right.$$

$$+ 2m^{-1}p^{-2}b_4[E(D_3^\ell) + E(D_3^{\ell-1})\Psi\sqrt{n}]$$

$$\left. + n^{-1/2}b_5E(D_3^{\ell+1}) + \delta E(D_3^{\ell-1})\Psi\sqrt{n} \right\}$$

¹⁰Again, please see full proof on my website for every "tedious" step.

$$\begin{aligned}
& + \ell(\ell - 1)n^{-1/2}p^{-2}b_3(E(D_3^{\ell-1}) + E(D_3^{\ell-2})\Psi\sqrt{n}) \\
& + \ell(\ell - 1)m^{-1}b_3^2p^{-2}[E(D_3^\ell) + 2E(D_3^{\ell-1})\Psi\sqrt{n} + E(D_3^{\ell-2})\Psi^2n] \\
& + O(mn^{-3/2} + m^{-1/2}n^{-1/2} + m^{-3/2} + (m/n)^4) \\
= & n^{-1/2}\ell\left\{ -(\ell - 1)(2p)^{-1}\Psi\sqrt{n}E(D_3^{\ell-2}) - \ell(2p)^{-1}E(D_3^{\ell-1}) \right. \\
& \quad \left. - (a_2/a_1)E(D_3^\ell)\Psi\sqrt{n} - (a_2/2a_1)E(D_3^{\ell+1}) \right\} \\
& + m^{-1}\ell\left\{ ((\ell - 1)/4)E(D_3^{\ell-2})\Psi^2n + (\ell/2)\Psi\sqrt{n}E(D_3^{\ell-1}) + ((\ell + 1)/4)E(D_3^\ell) \right\} \\
(41) \quad & + \delta\ell E(D_3^{\ell-1})\Psi\sqrt{n} + O(mn^{-3/2} + m^{-1/2}n^{-1/2} + m^{-3/2} + (m/n)^4).
\end{aligned}$$

Plugging in $z_2 + z_3$, (38) becomes

$$\begin{aligned}
E[(-p^{-1}Y)^\ell] & = E[(1 + \delta)(D_2 + D_3)^\ell] \\
& + n^{-1/2}\ell\left\{ -(\ell - 1)(2p)^{-1}\Psi\sqrt{n}E(D_3^{\ell-2}) - \ell(2p)^{-1}E(D_3^{\ell-1}) \right. \\
& \quad \left. - (a_2/a_1)E(D_3^\ell)\Psi\sqrt{n} - (a_2/2a_1)E(D_3^{\ell+1}) \right\} \\
& + m^{-1}\ell\left\{ \frac{\ell - 1}{4}E(D_3^{\ell-2})\Psi^2n + \frac{\ell}{2}\Psi\sqrt{n}E(D_3^{\ell-1}) + \frac{\ell + 1}{4}E(D_3^\ell) \right\} \\
(42) \quad & + \delta\ell E(D_3^{\ell-1})\Psi\sqrt{n} + o(m^{-1} + (m/n)^2).
\end{aligned}$$

As in HS88, let

$$(43) \quad K \equiv [p(1 - p)]^{-1/2}Y, \quad L \equiv -[p/(1 - p)]^{1/2}(1 + \delta)(D_2 + D_3),$$

and as they state,

$$(44) \quad E(D_3^{2k}) = [(1 - p)/p]^k(2k)!(k!2^k)^{-1} + O(n^{-1}), \quad E(D_3^{2k+1}) = O(n^{-1/2}),$$

where D_3 is asymptotically $N(0, (1 - p)/p)$. Multiplying equation (42) by $\{-[p/(1 - p)]^{1/2}\}^\ell(it)^\ell/\ell!$ and adding over ℓ , the formal expansion is

$$(45) \quad E(e^{itK}) = E(e^{itL}) + n^{-1/2}\alpha_1(t) + m^{-1}\alpha_2(t) + \delta\alpha_3(t) + o[m^{-1} + (m/n)^2],$$

where

$$\begin{aligned}
\alpha_1(t) & \equiv e^{-t^2/2} \left[\left(\frac{1}{2} - b_5(1 - p) \right) (p(1 - p))^{-1/2} ((it)^3 + (it)) \right. \\
& \quad \left. + (it)^2\Psi\sqrt{n} \left(2b_5 - \frac{1}{2(1 - p)} \right) \right],
\end{aligned}$$

$$\alpha_2(t) \equiv e^{-t^2/2} \frac{1}{4} \left[(it)^4 + (it)^2 \left(3 + \Psi^2 n \frac{p}{1-p} \right) - 2((it)^3 + (it)) \Psi \sqrt{n} \left(\frac{p}{1-p} \right)^{1/2} \right],$$

$$\alpha_3(t) \equiv e^{-t^2/2} (it) \left(-\Psi \sqrt{n} \left(\frac{p}{1-p} \right)^{1/2} \right),$$

which are, respectively, Fourier-Stieltjes transforms of

$$(46) \quad a_1(z) \equiv - \left[\Psi \sqrt{n} \left(2b_5 - \frac{1}{2(p-1)} \right) z + [(1/2) - b_5(1-p)] [p(1-p)]^{-1/2} z^2 \right] \phi(z),$$

$$(47) \quad a_2(z) \equiv \frac{1}{4} \left[-\Psi^2 n \frac{p}{1-p} z + 2\Psi \sqrt{n} \left(\frac{p}{1-p} \right)^{1/2} z^2 - z^3 \right] \phi(z), \text{ and}$$

$$(48) \quad a_3(z) \equiv \left[\Psi \sqrt{n} \left(\frac{p}{1-p} \right)^{1/2} \right] \phi(z).$$

The characteristic function of L is

$$(49) \quad E(e^{itL}) = \{1 + \delta(it)^2 - n^{-1/2} \frac{1}{6} [p/(1-p)]^{1/2} [(1+p)/p] (it)^3\} e^{-t^2/2} + O(\delta^2 + n^{-1}),$$

where the right-hand side is the Fourier-Stieltjes transform of

$$(50) \quad \Phi(z) - \delta z \phi(z) + n^{-1/2} \frac{1}{6} [p/(1-p)]^{1/2} [(1+p)/p] (z^2 - 1) \phi(z) + O(\delta^2 + n^{-1}).$$

To explicitly show the characteristic function of L , I take an expansion of the cumulant generating function, and use the known mean, variance, and third moment to plug in, calculating the third cumulant from the third moment. The fourth cumulant is also needed to show the bound on the continuation of the series expansion. Skipping the steps of calculation (found in the full proof), $E(L) = 0$, $E(L^2) = 1 + 2\delta + O(\delta^2 + n^{-1})$, $E(L^3) = -n^{-1/2}(1+p)p^{-1/2}(1-p)^{-1/2} + O(\delta^2 + n^{-1})$. Note, then, that $O((\mu'_2 - 1)^2) = O(\delta^2)$. Expanding the log of the characteristic function of L , which is the cumulant generating function,

$$\begin{aligned} \ln E(e^{itL}) &= \sum_{j=1}^{\infty} \frac{(it)^j}{j!} \kappa_j \quad (\kappa_j \equiv j\text{th cumulant}) \\ &= (it)\kappa_1 - (t^2/2)\kappa_2 + \frac{(it)^3}{6}\kappa_3 + \frac{(it)^4}{4!}\kappa_4 + \dots \end{aligned}$$

$$\begin{aligned}
&= (it)\mu'_1 - (t^2/2)(\mu'_2 - (\mu'_1)^2) + \frac{(it)^3}{6}(\mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3) \\
&\quad + \frac{(it)^4}{4!}(\mu'_4 - 4\mu'_3\mu'_1 - 3(\mu'_2)^2 + 12\mu'_2(\mu'_1)^2 - 6(\mu'_1)^4) + \dots \\
&= 0 - (t^2/2)\mu'_2 + \frac{(it)^3}{6}(\mu'_3) + \frac{(it)^4}{4!}(\mu'_4 - 3(\mu'_2)^2) + \dots, \text{ so} \\
E(e^{itL}) &= \exp\{- (t^2/2)(1 + (\mu'_2 - 1)) + \frac{(it)^3}{6}(\mu'_3) + \frac{(it)^4}{4!}(\mu'_4 - 3(\mu'_2)^2) + \dots\},
\end{aligned}$$

and plugging in leads to (49).

When simplifying terms, it can easily be verified that the Fourier-Stieltjes transform of the expression in (50),

$$\begin{aligned}
&\Phi(z) - \delta z\phi(z) + n^{-1/2}(1/6)(p/(1-p))^{1/2}((1+p)/p)(z^2 - 1)\phi(z) \\
&= \Phi(z) - \delta z\phi(z) + C(z^2 - 1)\phi(z) \\
&= \Phi(z) + \delta\phi'(z) + C\phi''(z),
\end{aligned}$$

is $e^{-t^2/2} + \delta(-1)^2(it)^2e^{-t^2/2} + C(-1)^3(it)^3e^{-t^2/2} = e^{-t^2/2}(1 + \delta(it)^2 - C(it)^3)$, matching (49). The proof in HS88 implies that the conditions are such that the remainder will be the same in both equations.

Combining results going back to (45) and noting that $b_5 = -(a_2/2a_1) = \frac{1}{2} + pg'(p)[2g(p)]^{-1} + O(n^{-1})$, the distribution function of K admits the Edgeworth expansion on the RHS of (23),

$$\begin{aligned}
(51) \quad P(K \leq z) &= \Phi(z) + \{n^{-1/2}u_{1,\gamma}^\dagger(z) + m^{-1}u_{2,\gamma}(z) + (m/n)^2u_{3,\gamma}(z)\}\phi(z) \\
&\quad + o\{m^{-1} + (m/n)^2\},
\end{aligned}$$

where

$$\begin{aligned}
u_{1,\gamma}^\dagger(z) &\equiv \frac{1}{6} \left(\frac{p}{1-p} \right)^{1/2} \frac{1+p}{p} (z^2 - 1) - \frac{\gamma f(\xi_p)}{p} \left(1 - \frac{pf'(\xi_p)}{[f(\xi_p)]^2} - \frac{1}{2(1-p)} \right) z \\
&\quad - \frac{1}{2} \left(\frac{p}{1-p} \right)^{1/2} \left(1 + \frac{f'(\xi_p)}{[f(\xi_p)]^2} (1-p) \right) z^2, \\
u_{2,\gamma}(z) &\equiv \frac{1}{4} \left[\frac{2\gamma f(\xi_p)}{[p(1-p)]^{1/2}} z^2 - \frac{\gamma^2 [f(\xi_p)]^2}{p(1-p)} z - z^3 \right], \text{ and} \\
u_{3,\gamma}(z) &\equiv \frac{3[f'(\xi_p)]^2 - f(\xi_p)f''(\xi_p)}{6[f(\xi_p)]^4} \left(z - \frac{\gamma f(\xi_p)}{[p(1-p)]^{1/2}} \right),
\end{aligned}$$

where the RHS of (51) is identical to the expansion indicated by (23).

Note that this is for η_p , not ξ_p (hence $u_{1,\gamma}^\dagger$ instead of $u_{1,\gamma}$), but implies the final result in Theorem 1 when combined with equation (22).

To show the final result, the inverse Fourier-Stieltjes transformed functions can be added to get the final answers since it is a linear transform (and consequently linear inverse).

Starting at (45), adding the inverse Fourier-Stieltjes transformed functions of the RHS leads to the distribution (cdf) of K :

$$\begin{aligned}
 P(K \leq z) &= \Phi(z) - \delta z \phi(z) + n^{-1/2} \frac{1}{6} \{p/(1-p)\}^{1/2} \{(1+p)/p\} (z^2 - 1) \phi(z) \\
 &\quad + O(\delta^2 + n^{-1}) + n^{-1/2} a_1(z) + m^{-1} a_2(z) + \delta a_3(z) + o\{m^{-1} + (m/n)^2\} \\
 &= \Phi(z) + \delta \phi(z) \left(\Psi \sqrt{n} \left(\frac{p}{1-p} \right)^{1/2} - z \right) \\
 &\quad + n^{-1/2} \phi(z) \left[\frac{1}{6} \left(\frac{p}{1-p} \right)^{1/2} \frac{1+p}{p} (z^2 - 1) - \Psi \sqrt{n} \left(2b_5 - \frac{1}{2(1-p)} \right) z \right. \\
 &\quad \quad \left. - ((1/2) - b_5(1-p))(p(1-p))^{-1/2} z^2 \right] \\
 &\quad + m^{-1} \phi(z) \frac{1}{4} \left[-\Psi^2 n \frac{p}{1-p} z + 2\Psi \sqrt{n} \left(\frac{p}{1-p} \right)^{1/2} z^2 - z^3 \right] \\
 &\quad + o\{m^{-1} + (m/n)^2\},
 \end{aligned}$$

and using $\delta \equiv -(m/n)^2 \frac{g''(p)}{6g(p)}$ and $\Psi = \gamma/(pg(p)\sqrt{n})$,

$$\begin{aligned}
 &= \Phi(z) + (m/n)^2 \frac{g''(p)}{6g(p)} \phi(z) \left(z - \frac{\gamma}{g(p)[p(1-p)]^{1/2}} \right) \\
 &\quad + n^{-1/2} \phi(z) \left[\frac{1}{6} \left(\frac{p}{1-p} \right)^{1/2} \frac{1+p}{p} (z^2 - 1) \right. \\
 &\quad \quad - \frac{\gamma}{pg(p)} \left(1 + \frac{pg'(p)}{g(p)} - \frac{1}{2(1-p)} \right) z \\
 &\quad \quad \left. - \frac{1}{2} \left(\frac{p}{1-p} \right)^{1/2} \left(1 - \frac{g'(p)}{g(p)} (1-p) \right) z^2 \right] \\
 &\quad + m^{-1} \phi(z) \frac{1}{4} \left[-\frac{\gamma^2}{[g(p)]^2 p(1-p)} z + \frac{2\gamma}{g(p)[p(1-p)]^{1/2}} z^2 - z^3 \right]
 \end{aligned}$$

$$+ o\{m^{-1} + (m/n)^2\},$$

and using $g(p) = 1/f(F^{-1}(p))$, $g'(p) = -f'(F^{-1}(p))/[f(F^{-1}(p))]^3$, and $g''(p) = [-f''(\xi_p)f(\xi_p) + 3[f'(\xi_p)]^2]/[f(\xi_p)]^5$, and thus $g'(p)/g(p) = -f'(\xi_p)/[f(\xi_p)]^2$,

$$\begin{aligned} &= \Phi(z) + (m/n)^2\phi(z)\frac{3[f'(\xi_p)]^2 - f(\xi_p)f''(\xi_p)}{6[f(\xi_p)]^4}\left(z - \frac{\gamma f(\xi_p)}{[p(1-p)]^{1/2}}\right) \\ &\quad + n^{-1/2}\phi(z)\left[\frac{1}{6}\left(\frac{p}{1-p}\right)^{1/2}\frac{1+p}{p}(z^2-1) \right. \\ &\quad\quad - \frac{\gamma f(\xi_p)}{p}\left(1 - \frac{pf'(\xi_p)}{[f(\xi_p)]^2} - \frac{1}{2(1-p)}\right)z \\ &\quad\quad \left. - \frac{1}{2}\left(\frac{p}{1-p}\right)^{1/2}\left(1 + \frac{f'(\xi_p)}{[f(\xi_p)]^2}(1-p)\right)z^2\right] \\ &\quad + m^{-1}\phi(z)\frac{1}{4}\left[\frac{2\gamma f(\xi_p)}{[p(1-p)]^{1/2}}z^2 - \frac{\gamma^2[f(\xi_p)]^2}{p(1-p)}z - z^3\right] + o\{m^{-1} + (m/n)^2\} \\ &= \Phi(z) + n^{-1/2}u_{1,\gamma}^\dagger(z)\phi(z) + (m/n)^2u_{3,\gamma}(z)\phi(z) + m^{-1}u_{2,\gamma}(z)\phi(z) \\ &\quad + o\{m^{-1} + (m/n)^2\} \end{aligned}$$

as in (51).

DEPARTMENT OF ECONOMICS, UNIVERSITY OF MISSOURI-COLUMBIA