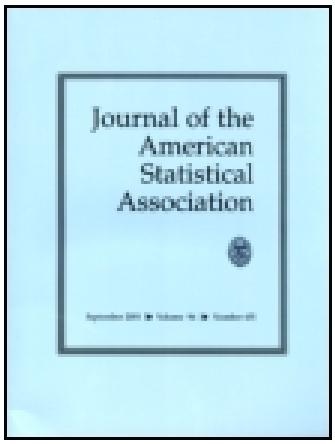

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Selection Adjusted Confidence Intervals With More Power to Determine the Sign

Asaf Weinstein [a], William Fithian [b] & Yoav Benjamini [a]
[a] Department of Statistics, University of Pennsylvania, Philadelphia, PA, 19104
[b] Department of Statistics, Stanford University, Stanford, CA, 4305
[c] Department of Statistics and Operations Research, The Sackler Faculty of Exact Sciences and The Sagol School of Neurosciences, Tel Aviv University, Tel Aviv, Israel
Accepted author version posted online: 17 Oct 2012.Published online: 15 Mar 2013.

To cite this article: Asaf Weinstein, William Fithian & Yoav Benjamini (2013) Selection Adjusted Confidence Intervals With More Power to Determine the Sign, Journal of the American Statistical Association, 108:501, 165-176, DOI: 10.1080/01621459.2012.737740

To link to this article: http://dx.doi.org/10.1080/01621459.2012.737740

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA

Selection Adjusted Confidence Intervals With More Power to Determine the Sign

Asaf WEINSTEIN, William FITHIAN, and Yoav BENJAMINI

In many current large-scale problems, confidence intervals (CIs) are constructed only for the parameters that are large, as indicated by their estimators, ignoring the smaller parameters. Such selective inference poses a problem to the usual marginal CIs that no longer offer the right level of coverage, not even on the average over the selected parameters. We address this problem by developing three methods to construct short and valid CIs for the location parameter of a symmetric unimodal distribution, while conditioning on its estimator being larger than some constant threshold. In two of these methods, the CI is further required to offer early sign determination, that is, to avoid including parameters of both signs for relatively small values of the estimator. One of the two, the Conditional Quasi-Conventional CI, offers a good balance between length and sign determination while protecting from the effect of selection. The CI is not symmetric, extending more toward 0 than away from it, nor is it of constant shape. However, when the estimator is far away from the threshold, the proposed CI tends to the usual marginal one. In spite of its complexity, it is specified by closed form expressions, up to a small set of constants that are each the solution of a single variable equation.

When multiple testing procedures are used to control the false discovery rate or other error rates, the resulting threshold for selecting may be data dependent. We show that conditioning the above CIs on the data-dependent threshold still offers false coverage-statement rate (FCR) for many widely used testing procedures. For these reasons, the conditional CIs for the parameters selected this way are an attractive alternative to the available general FCR adjusted intervals. We demonstrate the use of the method in the analysis of some 14,000 correlations between hormone change and brain activity change in response to the subjects being exposed to stressful movie clips. Supplementary materials for this article are available online.

KEY WORDS: Conditional confidence intervals; False coverage rate; Selective inference.

1. INTRODUCTION

Throughout this article, let $Y = \theta + Z$, where the density of the random variable Z is known, unimodal and symmetric about 0, and assume that we are interested in the value of the parameter only if $|Y|$ is big enough, say bigger than c. Alternatively, we only observe $X \overset{d}{=} (Y||Y| \geq c)$. This conditional distribution depends on $\theta = E(Y)$, and a confidence interval (CI) should be constructed for θ upon observing X. We will assume that $Z \sim N(0, 1)$, but unless specifically noted otherwise, the results are directly applicable to any unimodal and symmetric distribution.

Denote, as usual, the density of Z by $\phi(z)$ and its cumulative distribution function by $\Phi(z)$, and let $z_\alpha = \Phi^{-1}(1-\alpha)$ be the $(1-\alpha)$ quantile of Z. Furthermore, let $Q_c(\theta) = \mathrm{P}_\theta(|Y| \geq c) = 2 - \Phi(c-\theta) - \Phi(c+\theta)$. Then the probability density of X is

$$f_\theta(x) = \begin{cases} \phi(x-\theta)/Q_c(\theta), & c \leq |x| \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

In other words, the density of the observed random variable X is zero on $(-c, c)$, and is proportional to that of Y elsewhere. Note that the conditioning alters the role of the parameter θ: while it is a mere location parameter for Y (i.e., $Y - \theta$ has a distribution invariant of θ), it reflects both the location and the shape of X.

In this article, we suggest three different procedures for constructing two-sided confidence intervals for $\theta = E(Y)$. All three use the general duality between a family of α level tests and a $(1-\alpha)$ level confidence procedure (Lehmann 1986): If for each value $\theta \in \Theta$ of the parameter, $A(\theta)$ is a $1-\alpha$ level acceptance region, that is, $\mathrm{P}_\theta(X \in A(\theta)) \geq 1-\alpha$, then $S(X) = \{\theta : X \in A(\theta)\}$ is a $1-\alpha$ confidence set for θ.

In the first procedure, the acceptance region corresponding to θ has the shortest possible total length, just as the conventional acceptance region in the unconditional case. The other two follow the principle setup by Benjamini, Hochberg, and Stark (1998, hereafter BH&S) whereby a CI can serve the dual goal of (1) bounding the parameter within a short interval while (2) avoiding parameters of opposite signs, and a useful CI should balance the two. The tension between the two goals is evident when comparing two-sided intervals with one-sided intervals, which give up on finite length to be confined as much as possible to one side of the parameter space.

Property (2) is referred to as offering weak sign determination. Both opponents and even some proponents of hypothesis testing agree that a null hypothesis such as $\theta = 0$ is an ideal never to be found in practice (Pratt 1961; Tukey 1991; Benjamini, Hochberg, and Stark 1998). Therefore, a CI that includes no negative values has weakly determined the sign to be positive,

Asaf Weinstein is Professor, Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104 (E-mail: *asafw@wharton.upenn.edu*). William Fithian is Professor, Department of Statistics, Stanford University, Stanford, CA 94305 (E-mail: *wfithian@stanford.edu*). Yoav Benjamini is Professor, Department of Statistics and Operations Research, The Sackler Faculty of Exact Sciences and The Sagol School of Neurosciences, Tel Aviv University, Tel Aviv, Israel (E-mail: *ybenja@tau.ac.il*). We are thankful to Professor Larry Brown who, following the presentation of the FCR controlling intervals at the 2005 IMS/Bernoulli meeting in Barcelona, suggested constructing conditional confidence intervals as an alternative, and to Dr. Dani Yekutieli for the enlightening discussions on selective inference. We are thankful to Professor Talma Hendler from Tel Aviv University who lead the imaging research and with whom the first author is cooperating, and to Sharon Vaisvaser and to Yonatan Weinetraub who participated in that research. Y.B. thanks Stanford University for a Sabbatical stay during which the cooperation with W.F. started. The research of Y.B. was partially supported by grants from the Israeli Science Foundation and the European Research Council Advanced Researcher grant (PSARPS). W.F. was supported by the VIGRE DMS-0502385 grant from the National Science Foundation. The data analysis of the example was supported by a grant from the U.S. Department of Defense.

DOI: 10.1080/01621459.2012.737740

and it is an error only if $\theta < 0$. We too adopt this point of view.

The proposed CIs answer another rarely addressed yet important concern inherent in many current large-scale problems. In these problems, CIs are constructed for a limited set of parameters of interest, which are selected from the much larger pool of potential parameters based on the value of their estimators. Some examples are CIs constructed for associations of genetic markers with disease, only for the markers with p-values below a specified threshold; genes are chosen based on whether their expression levels differ between two groups by more than three-fold; risk factors for disease are chosen based on significance, and then their effect is estimated; regions in the brain that are highly correlated with a response are chosen and an estimator of the response is given only for the selected ones. In these situations, and in many others, a parameter is selected only if its associated estimator exceeds (in its absolute value) some constant threshold or, similarly, if the corresponding p-value is small enough.

The ongoing practice is to construct the usual (marginally correct) 95% CIs. When there is no selection involved, marginal CIs offer "coverage on the average" over all parameters estimated, and since there is no interest in simultaneous inference, their level of coverage suffices: we are accustomed to the 5% level of error, so here too we can bear the fact that 5% of the CIs will not cover their respective parameters (see, e.g., Gelman and Hill 2009).

Benjamini and Yekutieli (2005) addressed this issue in generality, considering a setting where a selection procedure takes place, and CIs are constructed only for the selected parameters. They demonstrated that if we restrict our attention to the subset of selected parameters, the marginal 95% CIs no longer provide their nominal coverage—not even on the average. The proportion of selected parameters not covered by their respective CIs can be much higher than 0.05 and even can approach 1. Hence, even if the strong protection of simultaneous coverage is not our aim, the prevailing practice does not in general offer "coverage on the average over the selected parameters." Formally, they introduce the false coverage-statement rate (FCR),

$$\mathrm{FCR} = E\left(\frac{\text{No. of noncovering intervals}}{\text{No. of intervals constructed}}\right),$$

where the ratio is 0 if no interval is constructed. They then addressed the problem of how to construct a set of intervals that offer FCR $\le q$, regardless of what selection procedure is used.

Using any of the three methods, we offer here to construct $1-q$ CIs for each selected parameter indeed assures FCR $\le q$. By constructing each interval based on the conditional distribution, the effect of selection is incorporated into the marginal coverage probability: conditional on any nonempty subset of selected parameters, S, we have

$$E\left(\frac{\text{No. of noncovering intervals of the } |S| \text{ constructed}}{|S|}\middle| S\right) \le \frac{1}{|S|}|S|q = q,$$

and since $\mathrm{FCR} \le E(\frac{\text{No. of noncovering intervals}}{|S|}||S| > 0)$, the FCR is controlled at q. Interestingly, the FCR CIs offered by Benjamini and Yekutieli (2005) were the initial motivation for the current work. For $Y_i \sim N(\theta_i, 1)$, $i = 1, \ldots, m$, consider the selection rule picking only those θ_i for which $|y_i| > c$, and denote by m' the size of this subset. In their procedure, a $(1 - \frac{m'\alpha}{m})$ interval is constructed for each of the selected parameters, regardless of the size of $|y_i|$.

In contrast, from a conditional point of view, when $|y_i| \gg c$ even the $y_i - z_{1-\alpha/2} < \theta_i < y_i + z_{1-\alpha/2}$ has approximately $1 - \alpha$ coverage, so an interval at a level close to $1 - \alpha$ should be sufficient. By the conditional constructions offered in this article we attempt to avoid the uniform inflation of their CIs. At the same time, we wanted the conditional intervals to (weakly) determine the sign for relatively small values of the estimator. Combining the approach of BH&S with the conditional approach led to the rest of the developments offered here.

To make the two approaches really comparable, we remove the restriction that the threshold should be predetermined. We show that conditioning on being larger than a *data-dependent* constant still offers FCR control for many selection rules, including those based on stepwise multiple testing procedures that control the familywise error rate or the false discovery rate.

The work by Finner (1994) addresses a similar concern to ours, that of CI following the rejection of a two-sided hypothesis, and the approach is similar in that acceptance regions are constructed and inverted. Alas Finner (1994) had constructed one-sided intervals making use of the unconditional distribution, giving up entirely on their length (being always infinite). Interestingly, for large values of the observable X, his CI reverts to the usual one-sided interval, just as the CIs constructed here revert in the same situation to the regular two-sided interval.

A construction of CIs of bounded length for $\theta = E(Y)$ based on a conditional distribution is given in Zhong and Prentice (2008). The methods proposed here are different in two essential ways: first, our methods yield exact CIs, while Zhong and Prentice provide "asymptotic" CIs, in that they assume asymptotic distributions for certain terms used to obtain their intervals. Second, while it is not clear what the properties of these asymptotic intervals are, the two main procedures we propose inherently possess favorable sign determination properties, that is, when using the CI to infer about the sign, they are more powerful.

2. SHORTEST ACCEPTANCE REGION

Throughout this section and the following ones, $|\Omega|$ will denote the Lebesgue measure of a set $\Omega \subset \mathbb{R}$, and (for convenience) will be referred to as the "length" of Ω (hence if Ω is an interval, $|\Omega|$ is just its length, and if Ω is a union of disjoint intervals, $|\Omega|$ is the sum of their lengths).

As a first attempt at constructing a family of acceptance regions, we associate each value of $\theta = E(X)$ with the shortest possible region of the observation space that captures a probability of $1 - \alpha$. In general, for each value assumed by θ, this is the set $A(\theta) = \{x : f_\theta(x) > \xi_\theta\}$, where ξ_θ is such that $\mathrm{P}(A(\theta)) = 1 - \alpha$.

In the usual normal case, where $X \stackrel{d}{=} Y$, these shortest regions are symmetric around θ, $A(\theta) = \{x : z_{1-\alpha/2} < x - \theta < z_{1-\alpha/2}\}$, and when inverted, yield the conventional, symmetric CI, $x \pm z_{1-\alpha/2}$. In our case, θ is no longer a simple location parameter, and the form of these retention regions is not as trivial. In particular, we lose the symmetry that characterized the former situation, where once we found the shortest region for a particular value of $\theta \neq 0$, we have essentially found all of them (because, except for $\theta = 0$, they differ by a translation

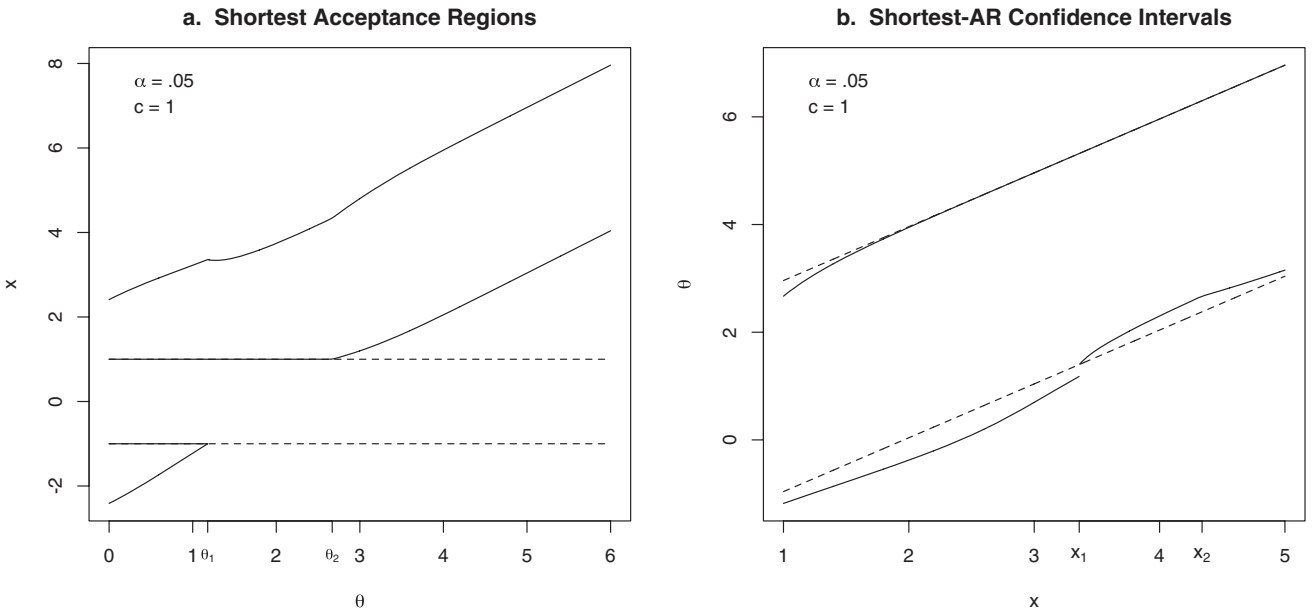

Figure 1. Acceptance regions and confidence intervals for the shortest acceptance region method, $c = 1, \alpha = 0.05$. (a) Acceptance regions. The region lies above c for $\theta \geq \theta_1 = 1.18$. (b) Confidence intervals. Weak (and strict) sign determination occurs for $x \geq x_1 = 3.36$. The boundaries of the standard (unadjusted) confidence interval are plotted with dashed lines.

and perhaps a reflection about the origin). Still, the fact that the original distribution (that of Y) is symmetric and unimodal makes it easier to obtain these regions even in the conditional case. Indeed, using the fact that on the support of f_θ, the density of X is proportional to that of Y (for fixed c and θ) is key to constructing $\{A(\theta)\}$.

As proved in the Appendix, for a given cutoff c the shortest acceptance region, $A_{\text{Srt}}(\theta)$, is given as follows:

For $0 \leq \theta < \theta_1$,

$$A_{\text{Srt}}(\theta) = \left\{\theta \pm \Phi^{-1}\left(1 - \frac{\alpha}{2} Q_c(\theta)\right)\right\} \setminus (-c, c);$$

For $\theta_1 \leq \theta < \theta_2$,

$$A_{\text{Srt}}(\theta) = \left(c, \theta + \Phi^{-1}\left[\,\Phi(c-\theta) + (1-\alpha)Q_c(\theta)\,\right]\right); \quad (2)$$

For $\theta_2 \leq \theta$,

$$A_{\text{Srt}}(\theta) = \left\{\theta \pm \Phi^{-1}\left[\frac{1}{2}\left(1 + (1-\alpha)Q_c(\theta)\right)\right]\right\},$$

with $A_{\text{Srt}}(\theta) = -A_{\text{Srt}}(-\theta)$ for $\theta < 0$. The parameters θ_1 and θ_2 are the solutions to

$$\Phi(c + \theta_1) - \Phi(c - \theta_1) = (1-\alpha)Q_c(\theta_1) \quad (3)$$

and

$$2\Phi(\theta_2 - c) - 1 = (1-\alpha)Q_c(\theta_2). \quad (4)$$

See Figure 1(a) for the shape of these acceptance regions as a function of θ.

Upon inverting these acceptance regions, the marginal confidence set may consist of disjoint intervals. Realizing that in most situations interest will be in CIs, we take the convex hull of the confidence region to get the interval $S_{\text{Srt}}(X) = (l(X), u(X))$. The resulting CI satisfies $S_{\text{Srt}}(-X) = -S_{\text{Srt}}(X)$ (where for a set $U \subset \mathbb{R}$, we denote $-U = \{-u : u \in U\}$), and for $c < X$ its bounds are as follows. The upper end $u(X)$ is the value of θ, which solves

$$2\Phi(X - \theta) = \alpha Q_c(\theta),$$

and the lower end, $l(X)$, is given as follows:

For $c \leq X < x_1$, $l(X)$ is the value of θ s.t.

$$2(1 - \Phi(X - \theta)) = \alpha Q_c(\theta);$$

For $x_1 \leq X < x_2$, $l(X)$ is the maximum value of θ s.t.

$$\Phi(X - \theta) - \Phi(c - \theta) = (1-\alpha)Q_c(\theta)\}; \quad (5)$$

For $x_2 \leq X$, $l(X)$ is the value of θ s.t.

$$2\Phi(X - \theta) = (1-\alpha)Q_c(\theta),$$

where

$$x_1 = \sup\{x : x \in A_{\text{Srt}}(\theta_1)\} = c + 2\theta_1 \quad (6)$$

and

$$x_2 = \sup\{x : x \in A_{\text{Srt}}(\theta_2)\} = 2\theta_2 - c. \quad (7)$$

See Figure 1(b) for the shape of these confidence intervals as a function of x.

3. A CONDITIONAL MODIFIED PRATT (CMP) PROCEDURE

With sign determination as a goal, inverting most powerful tests against the alternative being at zero should be optimal. For a unimodal, symmetric random variable, Pratt (1961) obtained the interval $(0, X + z_{1-\alpha})$ for $z_{1-\alpha/2} < x$ and $(X - z_{1-\alpha}, 0)$ for $x < -z_{1-\alpha/2}$, by inverting such tests. Unfortunately, the interval is of unbounded length when $|x|$ grows. A modification of Pratt's procedure is suggested in BH&S to bound the length. Specifically, they derive the above most powerful test under the further restriction that the length of the acceptance region never exceed r times the length of the symmetric acceptance region,

$2z_{1-\alpha/2}$. The corresponding CIs also satisfy $|S(x)| \le r2z_{1-\alpha/2}$ for every x.

A similar modification of Pratt's CI for the conditional problem raises a difficulty. While in the unimodal, symmetric case, the shortest $(1-\alpha)$ region for every θ is of constant length $2z_{1-\alpha/2}$, in the conditional case the length of the shortest region for θ, $A_{\mathrm{Srt}}(\theta)$, is not constant and depends on θ. We therefore allow each $A(\theta)$ to extend as much as $r|A_{\mathrm{Srt}}(\theta)|$ for that specific θ, and among these choose the most powerful test for $EX = \theta$ against the alternative $EX = 0$. Denote these conditional modified Pratt (CMP) tests by $A_{\mathrm{CMP}}(\theta)$.

Since f_θ is proportional to $\phi(x-\theta)$ on $(-c,c)^c$ (for any fixed θ), the ordering of x values in $[-c,c]^c$ by the likelihood ratio (LR) under the conditional densities $f_\theta(x)$ is the same as the ordering by the LR of the unconditional densities. Making use of the fact that the Normal density is also log-concave, with no length constraints, for any $\theta > 0$ the desired region would be $\{x \in (-c,c)^c : t_\theta < x\}$, where t_θ is determined so that the size of the test is α. When the length is constrained, $A_{\mathrm{CMP}}(\theta)$ is of the form $\{x \in [-c,c]^c : \bar{t}_{\theta,r} < x < \tilde{t}_{\theta,r}\}$, and in accordance with the unconditional argument choose the largest $\bar{t}_{\theta,r}$ so that $\mathrm{P}(A_{\mathrm{CMP}}(\theta)) = 1-\alpha$ and $|A_{\mathrm{CMP}}(\theta)| = r|A_{\mathrm{S}}(\theta)|$. By symmetry, $A_{\mathrm{CMP}}(\theta) = -A_{\mathrm{CMP}}(-\theta)$, and $A_{\mathrm{CMP}}(0)$ is chosen to be symmetric around zero. Let $\tilde{\theta}_1$ be the value of $\theta \in (0,\theta_1)$, which solves

$$\Phi(c + r\,|A_{\mathrm{Srt}}(\theta)| - \theta) - \Phi(c-\theta) = (1-\alpha)Q_c(\theta), \tag{8}$$

and for $0 < \theta < \tilde{\theta}_1$ denote by $\tilde{a}_1(\theta)$ the value of $x \in (\inf A_{\mathrm{Srt}}(\theta), -c) = (\theta - \Phi^{-1}(1-\frac{\alpha}{2}Q_c(\theta)), -c)$ for which

$$\begin{aligned} &1 - \Phi(\theta - x) + 1 - \Phi(c + r\,|A_{\mathrm{Srt}}(\theta)| - (-c-x) - \theta) \\ &\quad = \alpha Q_c(\theta), \end{aligned} \tag{9}$$

and by $\tilde{a}_2(\theta)$ the biggest value of $x \in (c,\infty)$ for which

$$\Phi\left(x + r\,|A_{\mathrm{Srt}}(\theta)| - \theta\right) - \Phi(x-\theta) = (1-\alpha)Q_c(\theta). \tag{10}$$

As shown in the Appendix, the resulting acceptance region is given as follows:

For $\theta = 0$,

$$A_{\mathrm{CMP}}(\theta) = \left[-\Phi^{-1}\left(1-\frac{\alpha}{2}Q_c(\theta)\right), \Phi^{-1}\left(1-\frac{\alpha}{2}Q_c(\theta)\right)\right] \setminus (-c,c);$$

For $0 < \theta < \tilde{\theta}_1$,

$$\begin{aligned} A_{\mathrm{CMP}}(\theta) = (&\tilde{a}_1(\theta), c + r\,|A_{\mathrm{Srt}}(\theta)| - (-c - \tilde{a}_1(\theta))) \\ &\setminus (-c,c); \end{aligned} \tag{11}$$

For $\theta > \tilde{\theta}_1$,

$$A_{\mathrm{CMP}}(\theta) = \left(\tilde{a}_2(\theta),\ \tilde{a}_2(\theta) + r\,|A_{\mathrm{Srt}}(\theta)|\right),$$

with $A(\theta) = -A(-\theta)$ for $\theta < 0$. See Figure 2(a) for the shape of these acceptance regions as a function of θ.

The CI obtained by inverting this family of tests and taking its convex hull is

$$S_{\mathrm{CMP}}(X) = \begin{cases} (\tilde{l}_1(X), \tilde{u}(X)), & c \le X < \tilde{x}_1 \\ [0, \tilde{u}(X)), & \tilde{x}_1 \le X < z_{1-\alpha/2} \\ (0, \tilde{u}(X)), & z_{1-\alpha/2} \le X < \tilde{x}_2 \\ (\tilde{l}_2(X), \tilde{u}(X)), & \tilde{x}_2 \le X < \tilde{x}_3 \\ (\tilde{l}_3(X), \tilde{u}(X)), & \tilde{x}_3 < X \end{cases}, \tag{12}$$

with $S_{\mathrm{CMP}}(-X) = -S_{\mathrm{CMP}}(X)$. In the above equation, $\tilde{u}(x)$ is the value of θ such that $\tilde{a}_2(\theta) = x$; $\tilde{l}_1(x)$ is the value of $\theta \in (-\tilde{\theta}_1, 0)$ for which $\tilde{a}_1(\theta) = -x$; $\tilde{l}_2(x)$ is the value of $\theta \in (0, \tilde{\theta}_1)$ such that $\tilde{a}_1(\theta) = -c - [r|A_{\mathrm{Srt}}(\theta)| - (x-c)]$; $\tilde{l}_3(x)$ is the value of θ such that $\tilde{a}_2(\theta) = x - r|A_{\mathrm{Srt}}(\theta)|$; and $\tilde{x}_1 = -\tilde{a}_1(0)$; $\tilde{x}_2 = c + [r|A_{\mathrm{Srt}}(0)| - (-c - \tilde{a}_1(0))]$; $\tilde{x}_3 = c + r|A_{\mathrm{Srt}}(\tilde{\theta}_1)|$; $z_{1-\alpha/2} = \Phi^{-1}(1 - \frac{\alpha}{2}(1 - \Phi(c) + 1 - \Phi(c)))$. See Figure 2(b) for the shape of these confidence intervals as a function of x.

4. CONDITIONAL QUASI-CONVENTIONAL (CQC) CONFIDENCE INTERVALS

A different approach to compromise between CI length and more power to determine the sign of the parameter, following BH&S, is to minimize a weighted sum of the length of the acceptance region $|A(\theta)|$ and the extent to which the acceptance region crosses the origin, subject to a size constraint. As in the unconditional case, this approach yields conditional CIs which revert to the conventional symmetric interval when $|x|$ is large.

4.1 Deriving the Family of Acceptance Regions

Formally, we let $A(0)$ be symmetric around zero, and with any $\theta \neq 0$ we associate a region

$$A_{\mathrm{CQC}}(\theta) = \underset{A}{\mathrm{argmin}} \left\{ \lambda\,|A(\theta)| + \sup_{x \in A(\theta): \mathrm{sgn}(x) \neq \mathrm{sgn}(\theta)} |x| \right\}, \tag{13}$$

where A is any region, which satisfies $\mathrm{P}_\theta(X \in A) \ge 1-\alpha$.

In the usual unconditional case, the optimal set of acceptance regions corresponding to a CI with length that never exceeds r times the length of the conventional, symmetric one, has a relatively simple structure. In fact, for a given α, it is completely characterized by the quantity $\bar{c} = \inf_{\theta>0} \inf\{y : y \in A(\theta)\}$, which is determined by r (or, equivalently, by λ). We use the exact same optimization setup (13) to obtain the family of acceptance regions in our conditional setting. However, now there is considerable complexity in determining the value of λ corresponding to a resulting CI with a certain maximum length. Therefore, we first show how to obtain the acceptance regions for a prescribed λ, and later discuss the relation between λ and the corresponding maximum CI length.

Theorem 1. For any $\lambda > 0$, the solution to the optimization problem (13) is given by

$$A_{\mathrm{CQC}}(\theta) = \begin{cases} (-c - d^\star(\theta), c + a_1'(\theta)) \setminus (-c,c), & 0 < \theta < \theta_1' \\ (c, c + a_2'(\theta)), & \theta_1' \le \theta < \theta_1 \\ A_{\mathrm{Srt}}(\theta), & \theta_1 \le \theta, \end{cases} \tag{14}$$

with $A_{\mathrm{CQC}}(\theta) = -A_{\mathrm{CQC}}(-\theta)$ for $\theta < 0$, and where

1. θ_1' is the value of θ satisfying
$$1 + \lambda\left(1 - \frac{\phi(c+\theta)}{\phi(\Phi^{-1}(2 - \Phi(c+\theta) - \alpha Q_c(\theta)))}\right) = 0, \tag{15}$$
and θ_1 is given in Section 2.
2. $d^\star(\theta)$ is the value of d, which solves
$$1 + \lambda\left(1 - \frac{\phi(c+d+\theta)}{\phi(\Phi^{-1}(2 - \Phi(c+d+\theta) - \alpha Q_c(\theta)))}\right) = 0. \tag{16}$$

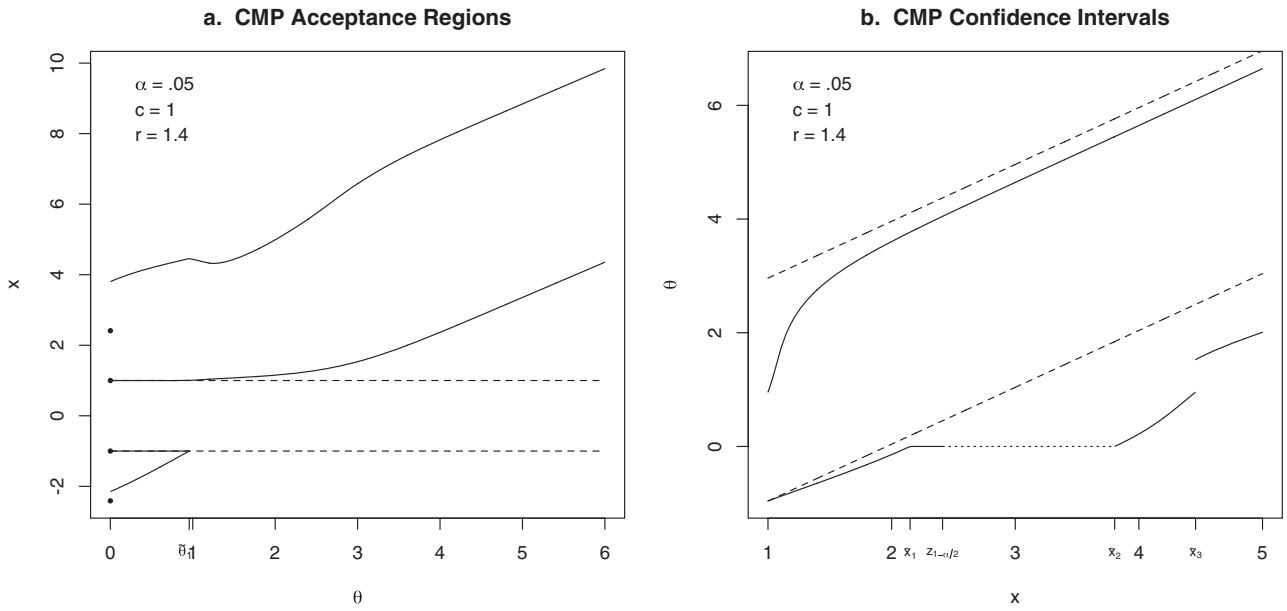

Figure 2. Acceptance regions and confidence intervals for the conditional modified Pratt (CMP) method, $c = 1, \alpha = 0.05, r = 1.4$. (a) Acceptance regions. The region lies above c for $\theta \geq \tilde{\theta}_1 = 0.96$. (b) Confidence intervals. A weak sign determination occurs for x as small as $\tilde{x}_1 = 2.15$ (compare with x_1 of the shortest acceptance region method), and, in turn, the interval separates from 0 only at $x \geq \tilde{x}_2 = 3.8$. Dotted line indicates that 0 is not included in the interval (but only in its closure). The boundaries of the standard (unadjusted) confidence interval are plotted with dashed lines. CMP interval length approaches $2rz_{1-\alpha/2} \approx 5.49$ as $x \to \infty$.

3. $a_1'(\theta)$ is the value of h, which solves

$$1 - \Phi(c + h - \theta) + 1 - \Phi(d^\star(\theta) + c + \theta) = \alpha Q_c(\theta). \tag{17}$$

4. $a_2'(\theta)$ is the value of h, which solves

$$\Phi(c + h - \theta) - \Phi(c - \theta) = (1 - \alpha)Q_c(\theta). \tag{18}$$

Proof. For $\theta \geq \theta_1$, it is obvious (from the discussion in Section 2) that the shortest region possible, $A_{\text{Srt}}(\theta)$, is optimal. □

For $0 < \theta < \theta_1$, denote by $d = d(\theta)$ the amount of extension of any candidate $A(\theta)$ to the left of $-c$, and by $l = l(\theta)$ the length, $|A(\theta)|$.

Lemma 1. Let $\psi(d, \theta, \lambda) = c + d + \lambda(d - c + \theta + \Phi^{-1}(2 - \Phi(d + c + \theta) - \alpha Q_c(\theta)))$. Furthermore, let $\underline{d}(\theta) = \max(-c - \theta + \Phi^{-1}\{1 - \alpha Q_c(\theta)\}, 0)$, $\overline{d}(\theta) = -c - \inf_x A_{\text{Srt}}(\theta)$, and $\theta^* = \inf\{\theta > 0 : \underline{d}(\theta) = 0\}$. Then:

(a) For $0 < \theta < \theta_1$, $A_{\text{CQC}}(\theta) = (-c - d^*, c + l^* - d^*) \setminus (-c, c)$, where

$$d^* = \underset{\underline{d}(\theta) \leq d \leq \overline{d}(\theta)}{\operatorname{argmin}} \psi(d, \theta, \lambda) \tag{19}$$

and where l^* is determined through

$$1 - \Phi(d^* + c + \theta) + 1 - \Phi(l^* - d^* + c - \theta) = \alpha Q_c(\theta),$$

is a solution to the original restricted optimization problem (13).

(b) On $G = \{(d, \theta) : \underline{d}(\theta) < d < \overline{d}(\theta), \quad \theta > 0\} \cup \{(0, \theta) : \theta^* < \theta < \theta_1\}$ and for any fixed $\lambda > 0$, the derivative of ψ with respect to d exists, is continuous in d and θ, and is a strictly increasing function in d for any fixed θ and in θ for any fixed d.

(c) For any fixed $\lambda > 0$, there exists $\theta \in (\theta^*, \theta_1)$ for which $\frac{\partial}{\partial d}\psi(d, \theta, \lambda)_{|_{d=0}} = 0$.

(d) For fixed $0 < \theta < \theta^*$ and for fixed $\lambda > 0$, there exists $d \in (\underline{d}(\theta), \overline{d}(\theta))$ such that $\frac{\partial}{\partial d}\psi(d, \theta, \lambda) = 0$.

Proof.

(a) Note that no region with a given extension d to the right of $-c$ is better, in terms of (13), than $(x_1, x_2) \setminus [-c, c]$, where $-c - x_1 = d$ and x_2 is then set to make $\mathrm{P}_\theta(A(\theta)) = 1 - \alpha$. That the part left of $-c$ should be an interval with a right end at $-c$ is obvious. As for the part, which lies to the right of c, first note that we may consider only $d < -c - \inf_S x A(\theta)$, since larger d would increase both terms in (13). Now, from Section 2, as long as $0 < \theta < \theta_1$, $A_{\text{Srt}}(\theta) = \{x : f_\theta(x) > \mathcal{K}_\theta\}$, where $\mathcal{K}_\theta < f_\theta(-c)$. Since for any $d < -c - \inf_x A_{\text{Srt}}(\theta)$,

$$(-c - d, -c) \cup \{x : f_\theta(x) > f_\theta(c)\} \subsetneq A_{\text{Srt}}(\theta),$$

we conclude that given any $d < -c - \inf_x A_{\text{Srt}}(\theta)$, the shortest region to the right of c we can possibly choose to hold the rest of the probability is an interval with a left end at $-c$.

Let there now be $0 < \theta < \theta_1$. By the discussion above, we may consider only candidates $A(\theta)$ of the form $(x_1, x_2) \setminus [-c, c]$, each of which is characterized by the extension d to the right of $-c$. The total length $l = l(\theta)$ of $A(\theta)$ is determined by d through

$$1 - \Phi(d + c + \theta) + 1 - \Phi(l - d + c - \theta) = \alpha Q_c(\theta). \tag{20}$$

Solving for l and plugging the expression in for $|A(\theta)|$, (13) can be rewritten in terms of d as

$$d^*(\theta) = \underset{d}{\operatorname{argmin}} \{c + d + \lambda [d - c + \theta + \Phi^{-1}(2 - \Phi(c + \theta + d) - \alpha p(c, \theta))]\}, \tag{21}$$

$$d \in \left[\max\left(-c-\theta+\Phi^{-1}\{1-\alpha Q_c(\theta)\}, 0\right), -c-\inf_x A_{\mathrm{Srt}}(\theta)\right].$$

where

$$\max(-c-\theta+\Phi^{-1}\{1-\alpha Q_c(\theta)\}, 0) \le d \le -c-\inf_x A_{\mathrm{Srt}}(\theta). \tag{22}$$

Here, the lower bound is the minimal value d has to take in order for $A(\theta)$ to satisfy the coverage requirement.

(b) Let $G_1 = \{(d,\theta) : \underline{d}(\theta) < d < \overline{d}(\theta),\ \theta > 0\}$ and $G_2 = \{(0,\theta) : \theta^* < \theta < \theta_1\}$. Then ψ is obviously defined on G_1, and since $\Phi(c+\theta)+\alpha Q_c(\theta) > 1$ for $\theta > \theta^*$, it is also defined on G_2. Hence ψ is defined on $G = G_1 \cup G_2$. Now, for any $(d,\theta) \in G$, we have

$$\frac{\partial}{\partial d}\psi(d,\theta,\lambda) = 1+\lambda\left(1-\frac{\phi(c+d+\theta)}{\phi(\Phi^{-1}(2-\Phi(c+d+\theta)-\alpha Q_c(\theta)))}\right), \tag{23}$$

which is a continuous function of d and θ. Moreover, one can easily verify that on G, the numerator of the quotient in (23) is strictly decreasing, while the denominator is strictly increasing, in d for any fixed θ and in θ for any fixed d.

(c) $\{\theta : (0,\theta) \in G\} = (\theta^*, \theta_1)$, and on that set

$$\frac{\partial}{\partial d}\psi(0,\theta,\lambda) = 1+\lambda\left(1-\frac{\phi(c+\theta)}{\phi(\Phi^{-1}(2-\Phi(c+\theta)-\alpha Q_c(\theta)))}\right). \tag{24}$$

Using the fact that θ^* and θ_1 are the values of θ, which solve, respectively,

$$1-\Phi(c+\theta) = \alpha Q_c(\theta) \quad \text{and}$$
$$2(1-\Phi(c+\theta)) = \alpha Q_c(\theta),$$

it is easy to verify that for any $\lambda > 0$, the expression on the right-hand side of (24) tends to $-\infty$ as $\theta \longrightarrow \theta^*$ from the right, and tends to 1 as $\theta \longrightarrow \theta_1$ from the left. By continuity, it must vanish for some intermediate value.

(d) Using the fact that $\underline{d}(\theta)$ is the value of d such that

$$1-\Phi(c+d+\theta) = \alpha Q_c(\theta)$$

and $\overline{d}(\theta)$ is the value of d such that

$$2(1-\Phi(d+c+\theta)) = \alpha Q_c(\theta),$$

one can easily verify that $\frac{\partial}{\partial d}\psi(d) \longrightarrow -\infty$ as $d \longrightarrow \underline{d}(\theta)$ from the right, and $\frac{\partial}{\partial d}\psi(d) \longrightarrow 1$ as $d \longrightarrow \overline{d}(\theta)$ from the left. Therefore, there must be some $d \in (\underline{d}(\theta), \overline{d}(\theta))$ such that $\frac{\partial}{\partial d}\psi(d) = 0$. □

Returning to the proof of Theorem 1, for any $0 < \theta < \theta_1$ we first need, by (a), to obtain the minimizer in (13). From (c) and (b), there exists a unique value of θ, denoted by θ_1', which satisfies (15). It also follows from (c) and (b) that for $\theta_1' < \theta < \theta_1$, $\frac{\partial}{\partial d}\psi(d,\theta,\lambda) > 0$ for any $d > 0$. Therefore, in that case the minimizer is necessarily $d = 0$. As for $0 < \theta < \theta_1'$, we distinguish between two cases. If $0 < \theta < \theta^*$, by (b) and (d), the value of d for which $\frac{\partial}{\partial d}\psi(d,\theta,\lambda) = 0$ is the minimum. If $\theta^* < \theta < \theta_1'$, then by (b) we have that $\frac{\partial}{\partial d}\psi(0,\theta,\lambda) < \frac{\partial}{\partial d}\psi(0,\theta_1',\lambda) = 0$ and $\frac{\partial}{\partial d}\psi(d,\theta,\lambda) \longrightarrow 1$ as $d \longrightarrow \overline{d}(\theta)$ from the left (as shown in the proof of (d)), thus, by (b), the value of d for which the derivative vanishes is again the unique minimum. The expressions in (14) for $0 < \theta < \theta_1$ and those in (17) and (18) are an immediate consequence of the description in (a) of an optimal acceptance region once the minimizer d^* is known. See Figure 3(a) for the shape of these acceptance regions as a function of θ.

Having obtained the above expressions for $A_{\mathrm{CQC}}(\theta)$, we now derive the corresponding CIs by inversion.

4.2 Inverting the Acceptance Regions

The convex hull of the set $\{\theta : X \in A_{\mathrm{CQC}}(\theta)\}$, where $A_{\mathrm{CQC}}(\theta)$ is given in (14), is

$$S_{\mathrm{CQC}}(X) = \begin{cases} (-\theta_1', \theta_1'), & X = c \\ (-l_1'(X), u'(X)), & 0 < X < x_1'(0) \\ [0, u'(X), & x_1'(0) \le X < z_{1-\alpha/2} \\ (0, u'(X)), & z_{1-\alpha/2} \le X < x_2' \\ (l_2'(X), u'(X)), & x_2' \le X < x_3' \\ S_{\mathrm{Srt}}(X), & x_3' \le X, \end{cases} \tag{25}$$

with $S_{\mathrm{CQC}}(X) = -S_{\mathrm{CQC}}(-X)$ for $X < 0$ and where

1. $x_1'(0) = \sup_{\theta<0} \sup A(\theta)$ and is the value of x for which
$$1+\lambda\left(1-\frac{\phi(c+d)}{\phi(\Phi^{-1}(2-\Phi(c+d)-\alpha Q_c(0)))}\right)\Big|_{d=x-c} = 0.$$
2. $z_{1-\alpha/2} = \sup_x A(0)$.
3. $x_2' = \inf_{\theta>0} \sup_x A(\theta) = \Phi^{-1}\{2-\Phi(x_1'(0))-\alpha Q_c(0)\}$.
4. $x_3' = \sup_x A(\theta_1') = \theta_1' + \Phi^{-1}\{\Phi(c-\theta_1') + (1-\alpha)Q_c(\theta_1')\}$.
5. $l_1'(x)$ is the value of θ such that
$$1+\lambda\left(1-\frac{\phi(c+d+\theta)}{\phi(\Phi^{-1}(2-\Phi(c+d+\theta)-\alpha Q_c(\theta)))}\right)\Big|_{d=x-c} = 0.$$
6. $l_2'(x)$ is the value of θ for which
$$1+\lambda\left(1-\frac{\phi(c+d+\theta)}{\phi(\Phi^{-1}(2-\Phi(c+d+\theta)-\alpha Q_c(\theta)))}\right)\Big|_{d=d(\theta,x)} = 0,$$
where $d(\theta,x) = -c-\theta+\Phi^{-1}\{2-\Phi(x-\theta)-\alpha Q_c(\theta)\}$.
7. $u'(x)$ is the value of θ for which $2(1-\Phi(\theta-x)) = \alpha Q_c(\theta)$. See Figure 3(b) for the shape of the confidence intervals as a function of x.

4.3 Relationship Between λ and the Maximum Length of the Confidence Interval

We specified the penalty λ to derive a family of acceptance regions, which was then inverted to CIs. Ideally, though, we would like to constraint the maximal length of the interval, which is conceptually easier to quantify, and let it dictate a corresponding value for λ.

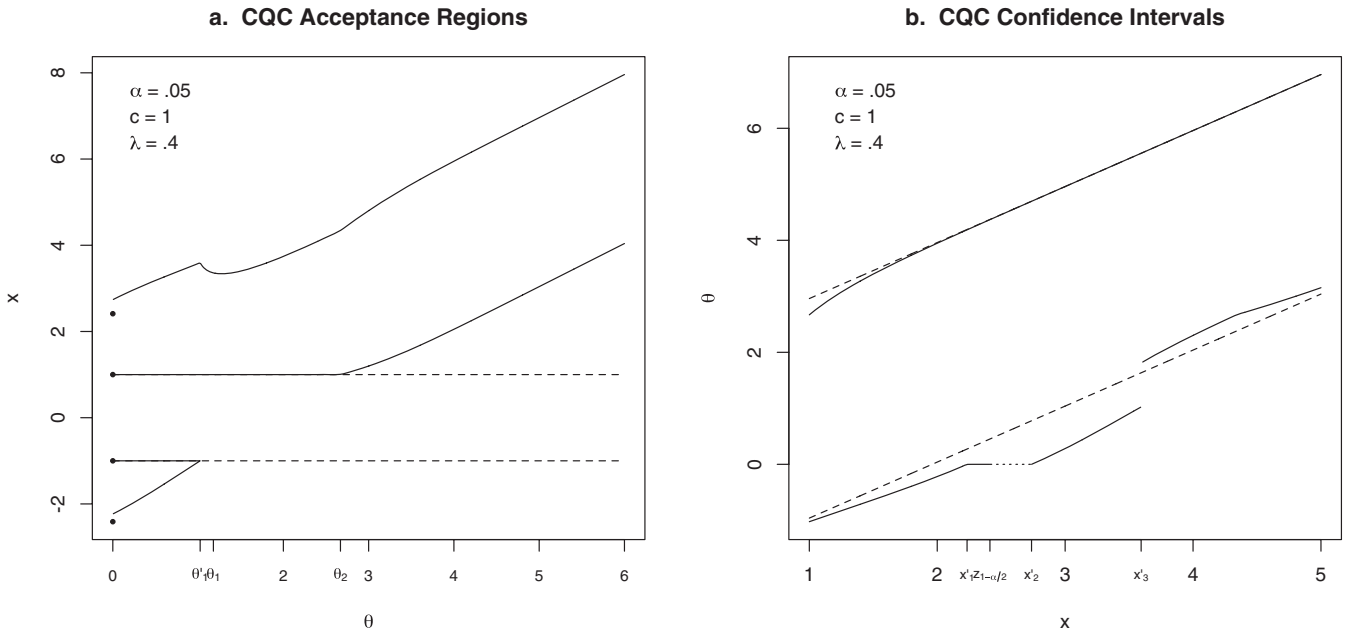

Figure 3. Acceptance regions and confidence intervals for the Conditional Quasi-Conventional (CQC) method, $c = 1, \alpha = 0.05, \lambda = 0.4$. (a) Acceptance regions. The region lies above c for θ as small as $\theta_1' = 1.02$. (b) Confidence intervals. A weak sign determination occurs for x as small as $x_1' = 2.23$, which is smaller than x_1 of the shortest acceptance region method; the interval separates from 0 at $x_3' = 3.59$, which is bigger than $z_{1-\alpha/2}$ but smaller than the point where the CMP with $r = 1.4$ interval separates from 0. While the length of the CMP interval is approximately $2rz_{1-\alpha/2}$ for big x, the CQC interval approaches the length of the standard, unadjusted interval, $2z_{1-\alpha/2} = 3.92$ as $x \to \infty$. Dotted line indicates that 0 is not included in the interval (but only in its closure). The boundaries of the standard (unadjusted) confidence interval are plotted with dashed lines.

In the unconditional case, because of the relatively simple structure of the optimal acceptance regions, it was possible to directly derive the family of these regions, which results in a maximum inflation factor of r for the CI, so it was unnecessary to make use of the corresponding value of λ. In our conditional case, θ is no longer a mere location parameter, and the family of optimal acceptance regions no longer admits the simple structure as before. This makes the relationship between the maximal length of the CQC interval and λ much more complicated and difficult to analyze analytically. Nevertheless, since it is reasonable to expect the CQC interval to be at least as long as the shortest acceptance region interval "for most x," it is interesting to numerically investigate how much worse it really does. Recall that our reference shortest acceptance region interval length varies with x, hence, we choose as a measure of comparison the quantity

$$w_c(\lambda) = \sup_x\{|S_{\mathrm{CQC}}(x)|/|S_{\mathrm{Srt}}(x)|\} \tag{26}$$

and numerically evaluate it over a range of λ values (Figure 4).

Note that for small λ, small changes have large effect on $w_c(\lambda)$, while for large values of λ, $w_c(\lambda)$ is much less sensitive to λ. For λ bigger than 0.4, the maximal increase in length is less than 25% for c = 2, and less than 20% for c = 1.

Remark. For each of the three methods discussed above, if we denote by $L(x)$ the lower end and by $U(x)$ the upper end of the CI, then these functions are monotone nondecreasing for $x > c$ (and $x < -c$), as can be observed in the above figures. Apart from the fact that the convex hull is taken upon inversion of the acceptance regions, this property is ensured by the facts that (1) $A(\theta) \cap \{x > c\}$ is always an interval and (2) $l(\theta) = \inf\{x > c : x \in A(\theta)\}$ is nondecreasing with $\lim_{\theta\to-\infty} l(\theta) = c$. That $U(x)$ is nondecreasing is trivial from (2); this is also the case for $L(x)$, since, letting $\bar{u}(\theta) = \sup_{\theta'\le\theta} \sup_x A(\theta')$ (the upper envelope of $u(\theta)$), we have from (1) and (2) that $L(x) = \inf\{\theta : x \in A(\theta)\} = \inf\{\theta : \bar{u}(\theta) > x\}$, which is a nondecreasing function of x since $\bar{u}$ is nondecreasing.

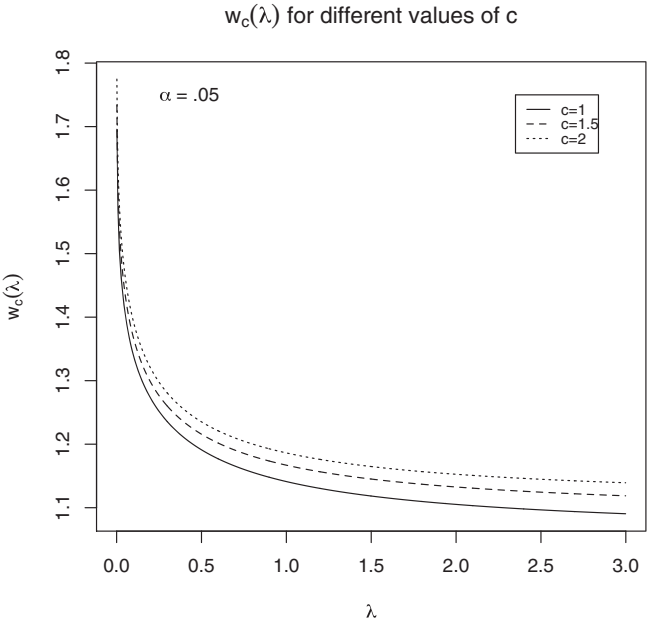

Figure 4. $w_c(\lambda)$ as a function of λ for $c = 1$, 1.5, and 2. The quantity $|S_{\mathrm{CQC}}(x^*)|/|S_{\mathrm{Srt}}(x^*)|$ was evaluated numerically.

5. EXAMPLE

In an ongoing experiment on the response to stress as reflected in brain activity and connectivity as measured by functional magnetic resonance imaging, subjects were exposed to stressful movie segments. Both the activity at voxels in the brain and the level of cortisol in their blood were recorded before and while being exposed. The cortisol levels, which are known to reflect the level of stress exposure, were log-transformed before taking the difference per subject. The resulting distribution of the estimates is quite Gaussian. The difference in activity level was estimated from a generalized linear model. The results per voxel, as inspected for a sample of voxels, are again quite symmetric and close to Gaussian. One of the questions that interests us is the correlation between the difference in activity and the difference in cortisol levels in the promising voxels. The results for 16 subjects are available at this stage, as the study is still ongoing (more subjects' data will become available). For the same reason, the study has not been reported yet, so we shall avoid giving further details about the experiment and the analysis leading to the correlations. Thus, we proceed from the 14,756 correlations we have—one for each voxel.

Interest lies only with voxels for which the absolute correlation is high, in this case larger than 0.6. There were 15 positive and 21 negative such correlations. The correlations were Fisher transformed into Gaussian variates, then CQC intervals were calculated using the values $\sigma = 1/\sqrt{16-3}$ and $\lambda = 0.4$. The standard unadjusted 95% marginal CIs were also calculated on this scale. Both sets of intervals were back transformed into the correlation scale and are presented in Figure 5. The CQC intervals are given by the lines and the standard intervals are marked by the tick marks. All 36 selected standard intervals do not cover 0 and seemingly offer evidence for both high positive correlations and high negative ones. In contrast, only the two largest correlations and the two smallest correlations have conditional intervals that exclude correlations of opposite signs, where the two largest ones include 0 and the two smallest ones exclude 0. The other 32 conditional intervals extend to beyond 0, indicating there is no evidence to exclude correlations of the other sign, based on the results at hand. Considering 36 intervals at 0.95 level, we expect noncoverage by 1.8 intervals, so we do not have yet an evidence of either positive or negative correlation from the conditional CIs. This point of view is the more realistic one in our case.

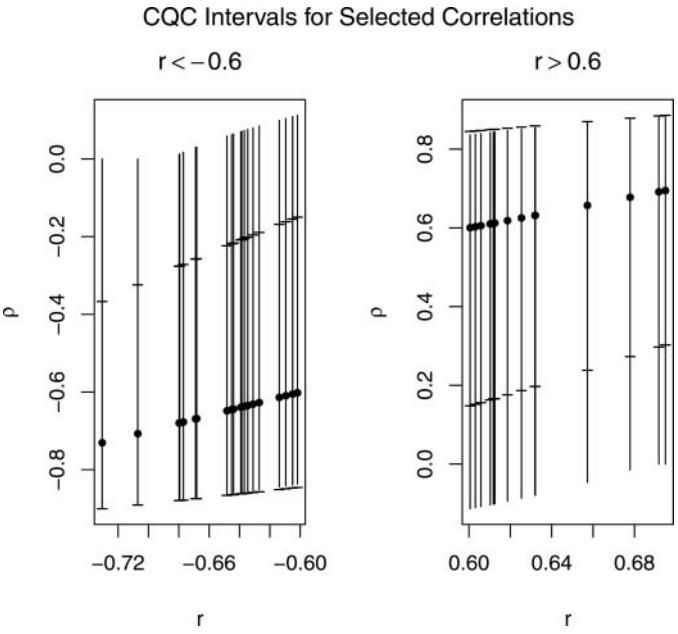

Figure 5. CQC and standard (unadjusted) intervals for correlations between cortisol level changes in the blood and activity changes in the brain at various locations, when viewing stressful movie segments. Only correlations that are greater in absolute value than 0.6 are displayed. CQC intervals are given by the vertical lines and the endpoints of the unadjusted standard are indicated by tick marks. While all of the standard intervals are separated from zero, only the two largest correlations and the two smallest ones have conditional intervals that exclude values of opposite signs.

Note also that the outer sides of the CQC CIs are almost the same as those of the standard ones, and even a bit shorter very close to the conditioned upon value.

6. COMPARISON WITH OTHER METHODS

In their work concerning selection bias of estimators in genome-wide association studies, Zhong and Prentice (2008) suggested methods to obtain CIs for $\theta = E(Y)$, where $Y \sim N(\theta, \sigma^2)$, upon observing $X \stackrel{d}{=} Y|(|Y| \geq c\sigma)$. Because σ is assumed to be known, there is no loss of generality in setting $\sigma = 1$.

First, since, asymptotically, $2\ln\{f_{\hat{\theta}_{\mathrm{MLE}}}(X)/f_\theta(X)\} \sim \chi^2_1$, a CI can be obtained as

$$S_{\mathrm{LR}}(X) = \left\{\theta : \ln f_{\hat{\theta}_{\mathrm{MLE}}}(X)) - \chi^2_{1;1-\alpha}/2 \leq \ln f_\theta(X)\right\}.$$

This approximation is obviously not supposed to hold for a small sample sizes. Even when many observations are combined into a single estimator $\hat{\theta}$, but then θ is estimated conditionally on $|\hat{\theta}| \geq \sigma c$, we are practically at a situation where we attempt to construct the confidence set from just a single observation.

A second, quantile-based (QB), CI is proposed by having each acceptance region leave an $\alpha/2$ probability on each tail of the conditional distribution.

Examining the acceptance regions of each of the above two methods, we can notice that both yield CIs that are roughly symmetric around x for large values of x, as displayed by Figure 6. Notice also that the QB acceptance regions are the same as those of the shortest acceptance region ones for small θ ($0 \leq \theta \leq \theta_1$, θ_1 defined in (3)), and in particular both methods make (weak) sign determination starting at the same value of the observed x. We can further see in Figure 6 that, similar to the CMP CI, the QB interval has a sharp drop toward the origin, which is due to the separation of the acceptance regions from c at $\theta = \theta_1$. In Figure 7, we can see the trade-off between early sign determination and late separation from zero: the CQC interval, as expected, weakly determines the sign earlier than does the QB interval, but does it at the cost of a later separation from zero.

Table 1 presents estimated values for the length, coverage probability, and probability of making a weak sign determination per each of the methods discussed in the article. We used $\theta = -10^{-6}$ with the notation θ^- instead of $\theta = 0$ to distinguish between a correct sign determination and a wrong one. In terms of coverage, all methods except LR always offer the right coverage, even if sometimes conservatively so. The LR

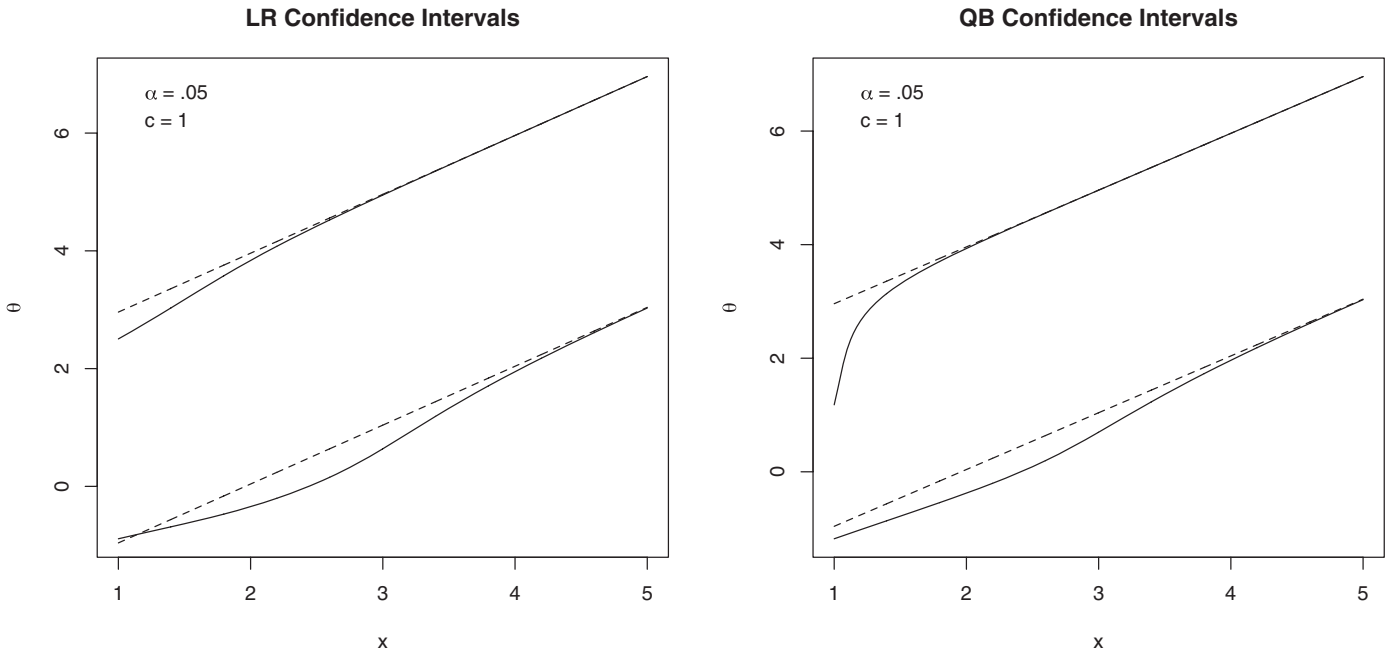

Figure 6. Confidence intervals for the likelihood ratio and quantile-based methods.

is unstable in terms of coverage, sometimes lower than needed ($c = 1,\ \theta = 1$), sometimes higher ($c = 3,\ \theta = 1$). As can be seen, the CMP procedure has shorter average length when θ is very small, but has significantly inflated average length when θ is big compared with the other methods (up to 30% longer). It is interesting to note that the CQC interval has a smaller expected length for big θ ($\theta = 5$) than do the LR and QB methods, while for small θ its expected length is very close to that of these two methods.

The three conditional intervals offered here always enjoy, as expected, higher power to determine the sign than do the LR and the QB methods; with the choices we made here for the parameters r and λ, the CMP has higher power than the other two procedures. However, the power of the CQC approaches that of the CMP as θ increases, and never falls by much.

In summary, it seems that the CQC method, while enjoying exact coverage properties, also reaches a good overall compromise between sign determination and expected length, across values that the conditioning constant c and the penalty term λ may take, and across θ where it matters.

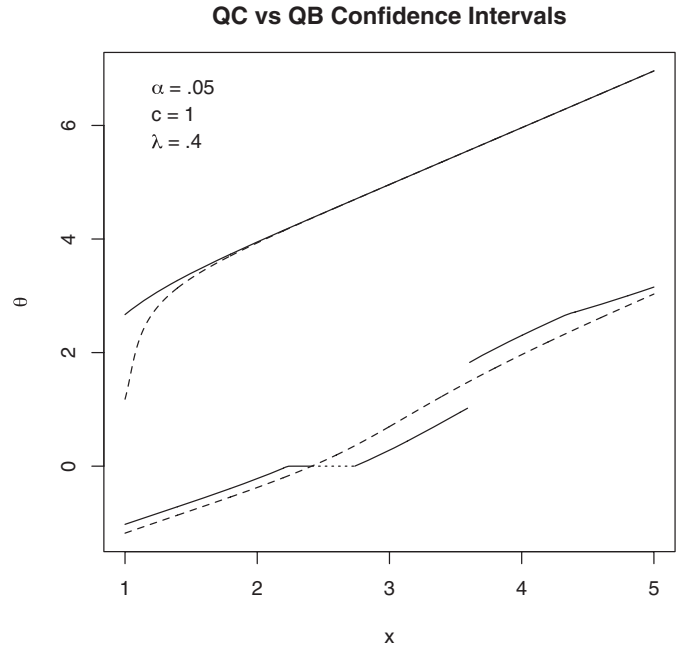

Figure 7. Quantile-based (dashed line) versus Conditional Quasi-Conventional (CQC) (solid line) confidence intervals. Parameter value used for the CQC interval is $\lambda = 0.4$. The CQC interval makes an earlier (weak) sign determination at the expense of a later separation from zero, and is approximately the standard, symmetric interval for large x.

7. FALSE COVERAGE-STATEMENT RATE WITH CONDITIONAL CONFIDENCE INTERVALS

Our main motivation for conditional CIs was their use in selective inference, where we select parameters with large estimators, and then construct CIs only for the selected ones. We showed in Section 1 that when selection is via comparison to a fixed threshold (e.g., fixed-level testing), the FCR is controlled, but we have not yet addressed selection via a data-dependent threshold (e.g., the BH procedure, Benjamini and Hochberg 1995).

Suppose we have m independent test statistics $\mathbf{Y} = (Y_1, \ldots, Y_m)$, where $Y_i \sim F_i(y - \theta_i)$ and each F_i satisfies the conditions as before. A selection procedure is called simple if for each component i, conditioning on the values of the other components $\mathbf{Y}^{(i)}$, the total number selected is constant whenever component i is among the selected. That is, conditional on $\mathbf{Y}^{(i)}$ and on the event $\{i \in S(\mathbf{Y}^{(i)}, Y_i)\}$, $|S(\mathbf{Y}^{(i)}, Y_i)| = R_i(\mathbf{Y}^{(i)})$ is constant with respect to Y_i.

This notion was introduced by Benjamini and Yekutieli (2005) who showed that step-up and step-down multiple testing procedures using the p-values (in particular the BH procedure)

Table 1. Simulation estimates of the expected length (Lng), probability for weak sign determination (PSD), and coverage probability (Cvr) for the shortest acceptance region (Srt), CMP, CQC, likelihood-ratio (LR), and quantile-based (QB) confidence intervals, and for different values of θ. The conditioning constants are $c = 1$ and $c = 3$. The parameters of CMP and CQC are $r = 1.5$ and $\lambda = 0.4$, respectively. The number of simulation repetitions is 2000, the standard error of Cvr is bounded by 0.005, that of PSD and of Lng by 0.01

	$\theta = 0^-$			$\theta = 1$			$\theta = 3$			$\theta = 5$		
	Lng	PSD	Cvr	Lng	PSD	Cvr	Lng	PSD	Cvr	Lng	PSD	Cvr
						$c = 1$						
Srt	4.127	0.025	0.95	4.181	0.146	0.955	4.07	0.743	0.954	3.826	0.996	0.948
CMP	3.259	0.051	0.95	3.53	0.236	0.951	4.615	0.813	0.948	5.196	0.999	0.951
CQC	3.998	0.04	0.952	4.098	0.203	0.95	4.229	0.791	0.954	3.872	0.998	0.948
LR	3.784	0.021	0.955	3.922	0.141	0.937	4.143	0.736	0.945	3.969	0.995	0.948
QB	3.833	0.025	0.95	3.997	0.146	0.955	4.158	0.743	0.951	3.963	0.996	0.948
						$c = 3$						
Srt	5.448	0.026	0.949	5.45	0.11	0.954	5.242	0.414	0.94	4.216	0.902	0.954
CMP	4.186	0.056	0.95	4.467	0.18	0.961	5.046	0.511	0.95	5.182	0.934	0.955
CQC	5.394	0.042	0.951	5.422	0.154	0.975	5.318	0.473	0.94	4.301	0.924	0.954
LR	4.703	0.018	0.962	4.837	0.095	0.976	5.063	0.384	0.971	4.574	0.891	0.943
QB	4.779	0.026	0.949	4.966	0.11	0.947	5.146	0.414	0.945	4.529	0.902	0.956

are simple selection procedures. Many of these testing procedures have the further property that a parameter i is selected (rejected) if its estimator passes a data-dependent threshold: $|Y_i| > C_i(\mathbf{Y}^{(i)})$. We shall call these *simple thresholding selection procedures*. Choosing the k largest or k most significant estimators also falls into this category.

Theorem 2. Let $S(\mathbf{Y})$ be a simple thresholding selection procedure, with $C_i(\mathbf{Y}^{(i)})$ as thresholds, and let $\mathrm{CI}_i(\alpha)$ for each $i \in S(\mathbf{Y})$ be a level $1-\alpha$ conditional CI for θ_i given $|Y_i| > C_i(\mathbf{Y}^{(i)})$. Then this set of CIs controls the FCR at level α.

Proof. Define

$$A_r^{(i)} = \{\mathbf{Y}^{(i)} : \text{ if } i \in S(\mathbf{Y}^{(i)}, Y_i) \text{ then } |S(\mathbf{Y}^{(i)}, Y_i)| = r\}$$

Since S is simple, for each i the $A_r^{(i)}$ partition all possible values of $\mathbf{Y}^{(i)}$. Using an identity from Benjamini and Yekutieli (2005), we have

$$\mathrm{FCR} = \sum_{r=1}^{m} \frac{1}{r} \sum_{i=1}^{m} \mathrm{P}\left(\theta_i \notin CI_i(\alpha), i \in S(\mathbf{Y}), |S(\mathbf{Y})| = r\right) \quad (27)$$

$$= \sum_{r=1}^{m} \frac{1}{r} \sum_{i=1}^{m} \mathrm{P}(\theta_i \notin CI_i(\alpha), |Y_i| > C_i(\mathbf{Y}^{(i)}), \mathbf{Y}^{(i)} \in A_r^{(i)}). \quad (28)$$

Because Y_i is independent of $\mathbf{Y}^{(i)}$, the $\mathrm{CI}_i(\alpha)$ enjoy conditional coverage just as if the threshold $C_i(\mathbf{Y}^{(i)})$ were fixed. That is,

$$\mathrm{P}(\theta_i \notin \mathrm{CI}_i(\alpha) \mid |Y_i| > C_i(\mathbf{Y}^{(i)}), \mathbf{Y}^{(i)} \in A_r^{(i)}) \le \alpha. \quad (29)$$

Using (29), along with

$$\sum_i \mathrm{P}(i \in S(\mathbf{Y}), |S(\mathbf{Y})| = r) = r\mathrm{P}(|S(\mathbf{Y})| = r),$$

$$\mathrm{FCR} \le \sum_{r=1}^{m} \frac{1}{r} \sum_{i=1}^{m} \alpha \mathrm{P}(i \in S(\mathbf{Y}), |S(\mathbf{Y})| = r) \quad (30)$$

$$= \alpha \sum_{r=1}^{m} \frac{1}{r} r\mathrm{P}(|S(\mathbf{Y})| = r) \quad (31)$$

$$= \alpha \mathrm{P}(|S(\mathbf{Y})| \ge 1) \le \alpha. \quad (32)$$

The above equation further implies that when we use any simple thresholding selection procedure, the FCR is also controlled at level α conditional on constructing at least one interval. □

8. GENERALIZATIONS

It is worth emphasizing that the only properties of the Gaussian distribution, which were used in the derivation of the CIs are symmetry and unimodality, and for the CMP log-concavity of the density. Hence, CIs can be constructed with the relevant constants calculated for any other distribution enjoying these properties.

The CMP and the CQC intervals can be constructed for Gaussian distributions with known standard error other than 1 by dividing the estimator and the conditioned upon constant c by σ, calculating the CIs and reinflating them by σ. Yet care has to be taken as to the interpretation of the constant c. If c serves as a condition on the value of the original (unnormalized) estimator, the above is fine. If c expresses significance, say being $z_{1-\alpha/2}$, it should not be divided by σ (or equivalently should be first multiplied by σ before being divided).

In spite of these two observations, the Gaussian case of unknown σ with small sample size being estimated from the data using $\widehat{\sigma}$ presents a challenge. Finner (1994) addressed similar issues at length regarding his construction of his one-sided CIs. He makes the important distinction between the above two cases. When interest is with significance the problem involves using properties of the noncentral t-distribution; when interest is with a large estimated value, his solution is more complicated. In both cases the problem we face is more difficult, because after the conditioning we can no longer rely on the symmetry of the unconditional distribution. We therefore leave this problem for future research.

Nevertheless, when the standard errors are estimated but can be used under the asymptotic Gaussian regime, the proposed CI can be used. This would include one-sample mean, two-sample means difference, and regression coefficients, with large enough sample sizes; correlation coefficients after using Fisher's

transformation; the coefficients in a logistic regression; coefficients in survival analysis, and such. Finally, we have discussed the control of FCR for simple thresholding selection procedures and independent estimators only. Generalization to thresholding procedures that are not simple are being developed. It seems that further generalization to positive regression dependent estimators, as done in Benjamini and Yekutieli (2005), is also feasible. Both will be explored in a later study.

9. DISCUSSION

When facing a family of m parameters, it is often the case that the parameters of interest are only the large ones, or those significantly different from 0 at some given level. All three types of conditional intervals can be used then to set CIs for these parameters only, and the set of intervals still assures control over the FCR.

The selection rules can be as simple as selecting the largest k-estimators. It can follow individual testing, multiple testing based rule such as the fixed-level Bonferroni as well step-up and step-down procedures that control the FDR and some of its variants.

There is some difficulty that arises from the dual interpretation of selection via testing, which is already evident in the single parameter case. When selecting a parameter by (unconditional) testing whether it is significantly different from 0, the conditional CI at the same level falls below zero when $|x|$ is close to the threshold, meaning it includes values on both sides of 0. Benjamini and Yekutieli (2005) discussed and demonstrated the inevitability of this phenomenon, and further studied its limiting implications when addressing multiple parameters. This is the motivation behind their introduction of the FCR as a goal in selective inference, rather than the conditional coverage. Indeed, with this goal in mind, they show that constructing $1 - \frac{Rq}{m}$ level CIs for the R parameters selected through the procedure of Benjamini and Hochberg for multiple testing at FDR level q controls the FCR at q while ensuring that none of the constructed intervals includes zero.

In view of that, the challenge that motivated this work is still intriguing if but from a new perspective. If the estimators are independent, then following testing via the method of Benjamini and Hochberg, both the method using Benjamini and Yekutieli's FCR-adjustment of the marginal CIs and the method using the conditional CIs developed here, offers FCR-controlling CIs for the selected parameters. The first produces intervals that avoid parameter values of opposite signs, but they are inflated, even for large values of the estimator; the latter provides intervals that are short when the estimator is large, and keep the FCR closer to the desired level, but include values of opposite signs when the estimator is close enough to the testing threshold. It is therefore interesting to investigate whether a method exists, which admits both desired properties.

APPENDIX

A.1 Obtaining the Shortest Acceptance Regions

The goal is to obtain, for each value of θ, a formal expression for the set $A(\theta) = \{x : f_\theta(x) > \xi_\theta\}$, where ξ_θ is such that $\mathrm{P}_\theta(A(\theta)) = 1 - \alpha$. Let $g_\theta(y) = \phi(y - \theta)$ be the (unconditional) distribution of Y. Then, since the densities of X and Y are proportional on the support of X, $A(\theta) = \{x : g_\theta(x) > \xi'_\theta\}$, where ξ'_θ is such that $\mathrm{P}_\theta(A(\theta)) = (1 - \alpha)Q_c(\theta)$. Moreover, note that

1. if $g_\theta(c) \le \xi'_\theta$, $A(\theta)$ lies to the right of c and is of the form $(\theta - t, \theta + t)$,
2. if $g_\theta(-c) \le \xi'_\theta < g_\theta(c)$, $A(\theta)$ lies to the right of c and is of the form (c, t),
3. if $0 < \xi'_\theta \le g_\theta(-c)$, $A(\theta)$ intersects both the region to the left of $-c$ and the region to the right of c, and is of the form $(\theta - t, \theta + t)$,

where in each of the cases above the value t depends on θ. Observing that, necessarily, $A(-\theta) = A(\theta)$, it is left to determine which of (1)–(3) is the case for each value of $\theta \ge 0$. Let

$$S_1(\theta) = \{x : g_\theta(x) > g_\theta(-c)\}, \qquad S_2(\theta) = \{x : g_\theta(x) > g_\theta(c)\}.$$

Then

$$\mathrm{P}((c, c + 2\theta)) = \frac{\Phi(c + \theta) - \Phi(c - \theta)}{Q_c(\theta)}, \quad \mathrm{P}(S_2(\theta)) = \frac{2\Phi(\theta - c) - 1}{Q_c(\theta)},$$

and

(a) $\mathrm{P}_\theta(S_1(\theta)) \to 1$, $\mathrm{P}_\theta(S_2(\theta)) \to 1$ as $\theta \to \infty$, $\mathrm{P}_{\theta=0}(S_1(0)) = \mathrm{P}_{\theta=0}(S_2(0)) = 0$, and both are strictly increasing functions on $\{\theta \ge 0\}$, except maybe on the regions where they vanish, which can be verified by taking derivatives.

(b) $\mathrm{P}_\theta(S_1(\theta)) < \mathrm{P}_\theta(S_2(\theta))$ for all $\theta > 0$ since for any such θ, $S_1(\theta) \subsetneq S_2(\theta)$.

By (a) and (b), there exist unique values θ_1 and θ_2, $\theta_1 < \theta_2$, such that $\xi'_{\theta_1} = g_{\theta_1}(-c)$ and $\xi'_{\theta_2} = g_{\theta_1}(c)$, and, further, the case in (1) holds when $\theta_2 \ge \theta$; the case in (2) holds when $\theta_1 \le \theta < \theta_2$; and the case in (3) holds when $0 \le \theta < \theta_1$. Also, from (1) and (2) we conclude that θ_1 is the solution to

$$\Phi(c + \theta) - \Phi(c - \theta) = (1 - \alpha)Q_c(\theta), \tag{A.1}$$

and θ_2 is the solution to

$$2\Phi(\theta - c) - 1 = (1 - \alpha)Q_c(\theta). \tag{A.2}$$

Now that the form among those described in (1)–(3) is known for each θ, the value of t is determined through the requirement that $\mathrm{P}_\theta(A(\theta)) = 1 - \alpha$, which yields the expression in (2) of Section 2.

A.2 Obtaining the CMP Acceptance Regions

The discussion in Section 3, which precedes the formal statement of the CMP acceptance regions, gives a qualitative description of $A(\theta)$. To obtain the exact expression in (11), we need to determine for which θ values $A(\theta)$ intersects $(-\infty, -c)$ and when it is entirely contained in (c, ∞), and then, distinguishing between these two cases, obtain the desired boundaries of $A(\theta)$ as the proper roots of the respective equations. For this purpose, observe that

(a) For $\theta_1 < \theta$, $A(\theta)$ is contained in (c, ∞).

(b) $\mathrm{P}_\theta(c, c + r|A_{\mathrm{Srt}}(\theta)|)$ is strictly increasing in θ on $0 < \theta < \theta_1$.

(c) For $0 < \theta < \theta_1$, $A(\theta) \cap (-\infty, -c) \ne \emptyset \iff \mathrm{P}_\theta\ (c, c + r|A_{\mathrm{Srt}}(\theta)|) < 1 - \alpha$.

(d) $\mathrm{P}_\theta\{(x, c + r|A_{\mathrm{Srt}}(\theta)| - (-c - x)) \setminus [-c, c]\}$ is strictly increasing in x on $x \in (\inf_x A_{\mathrm{Srt}}(\theta), -c)$.

In the above, (a) is because even the shortest acceptance region is contained in (c, ∞); (c) is because for $0 < \theta < \theta_1$, the probability of

the interval in the right-hand side is the biggest among all intervals of length $r|A_{\mathrm{Srt}}(\theta)|$, which intersect (c, ∞); and (b) and (d) can be verified by writing out the explicit expression for the probability, and taking the derivative.

For $\alpha < 0.5$, if we denote $h(\theta) = \mathrm{P}_\theta(c, c + r|A_{\mathrm{Srt}}(\theta)|)$, then by continuity of h on $0 \le \theta \le \theta_1$ and the facts that $h(0) < \mathrm{P}_{\theta=0}(c, \infty) = 0.5 < 1 - \alpha$ and $h(\theta_1) > \mathrm{P}_{\theta_1}(c, c + |A_{\mathrm{Srt}}(\theta_1)|) = 1 - \alpha$, it follows that there exists $0 < \tilde{\theta}_1 < \theta_1$ such that $\mathrm{P}_{\tilde{\theta}_1}(c, c + r|A_{\mathrm{Srt}}(\tilde{\theta}_1)|) = 1 - \alpha$. By (b), we conclude that

$$\begin{cases} \mathrm{P}_\theta\left(c, c + r\,|A_{\mathrm{Srt}}(\theta)|\right) < 1 - \alpha, & 0 < \theta < \tilde{\theta}_1 \\ \mathrm{P}_\theta\left(c, c + r\,|A_{\mathrm{Srt}}(\theta)|\right) > 1 - \alpha, & \tilde{\theta}_1 < \theta < \theta_1. \end{cases}$$

It now follows from (a) and (c) that $A(\theta)$ intersects $(-\infty, -c)$ when $0 < \theta < \tilde{\theta}_1$ and is entirely contained in (c, ∞) for $\theta > \tilde{\theta}_1$. Using the discussion in Section 3, (11) is true for $\tilde{a}_1(\theta)$ and $\tilde{a}_2(\theta)$, which satisfy (9) and (10), respectively. Finally, (d) implies that there is indeed a unique solution to (9) in $(\inf A_{\mathrm{Srt}}(\theta), -c)$.

SUPPLEMENTARY MATERIALS

R code containing functions for computing the acceptance regions and confidence intervals for the suggested methods.

[Received August 2010. Revised September 2012.]

REFERENCES

Benjamini, Y., and Hochberg, Y. (1995), "Controlling the False Discovery Rate in Multiple Comparisons: A Practical and Powerful Approach," *Journal of the Royal Statistical Society,* Series B, 57, 289–300. [173]

Benjamini, Y., Hochberg, Y., and Stark, P. B. (1998), "Confidence Intervals With More Power to Determine the Sign: Two Ends Constrain the Means," *Journal of the American Statistical Association*, 93, 309–317. [165]

Benjamini, Y., and Yekutieli, D. (2005), "False Discovery Rate–Adjusted Multiple Confidence Intervals for Selected Parameters," *Journal of the American Statistical Association*, 100, 71–81. [166,173,174,175]

Finner, H. (1994), "Two-Sided Tests and One Sided Confidence Bounds," *The Annals of Statistics*, 22, 1502–1516. [166,174]

Gelman, A., and Hill, J. (2009), *Data Analysis Using Regression and Multilevel/Hierarchical Models* (2nd ed.), New York: Cambridge University Press. [166]

Lehmann, E. (1986), *Testing Statistical Hypotheses*, New York: Wiley. [165]

Pratt, J. W. (1961), "Length of Confidence Intervals," *Journal of the American Statistical Association*, 56, 549–567. [165,167]

Tukey, J. (1991), "The Philosophy of Multiple Comparisons," *Statistical Science*, 6, 100–116. [165]

Zhong, H., and Prentice, R. L. (2008), "Bias-Reduced Estimators and Confidence Intervals for Odds Ratios in Genome-Wide Association Studies," *Biostatistics*, 9, 621–634. [166,172]