Testing in the Presence of Nuisance Parameters: Some Comments on Tests Post-Model-Selection and Random Critical Values

Hannes Leeb and Benedikt M. Pötscher

Abstract We point out that the ideas underlying some test procedures recently proposed for testing post-model-selection (and for some other test problems) in the econometrics literature have been around for quite some time in the statistics literature. We also sharpen some of these results in the statistics literature. Furthermore, we show that some intuitively appealing testing procedures, that have found their way into the econometrics literature, lead to tests that do not have desirable size properties, not even asymptotically.

1 Introduction

Suppose we have a sequence of statistical experiments given by a family of probability measures $\{P_{n,\alpha,\beta} : \alpha \in A, \beta \in B\}$ where α is a "parameter of interest", and β is a "nuisance-parameter". Often, but not always, A and B will be subsets of the Euclidean space. Suppose the researcher wants to test the null hypothesis $H_0 : \alpha = \alpha_0$ using the real-valued test-statistic $T_n(\alpha_0)$, with large values of $T_n(\alpha_0)$ being taken as indicative for violation of H_0.[1] Suppose further that the distribution of $T_n(\alpha_0)$ under H_0 depends on the nuisance parameter β. This leads to the key question: How should the critical value then be chosen? [Of course, if another, pivotal, test-statistic is available, this one could be used. However, we consider here the case where a (non-trivial) pivotal test-statistic either does not exist or where the researcher—for better or worse—insists on using $T_n(\alpha_0)$.] In this situation a standard way (see, e.g., [3, p.170]) to deal with this problem is to choose as critical

---

[1]This framework obviously allows for "one-sided" as well as for "two-sided" alternatives (when these concepts make sense) by a proper definition of the test-statistic.

H. Leeb (✉) • B.M. Pötscher
Department of Statistics, University of Vienna, Vienna, Austria
e-mail: hannes.leeb@univie.ac.at

value

$$c_{n,\sup}(\delta) = \sup_{\beta \in B} c_{n,\beta}(\delta), \tag{1}$$

where $0 < \delta < 1$ and where $c_{n,\beta}(\delta)$ satisfies $P_{n,\alpha_0,\beta}\left(T_n(\alpha_0) > c_{n,\beta}(\delta)\right) = \delta$ for each $\beta \in B$, i.e., $c_{n,\beta}(\delta)$ is a $(1-\delta)$-quantile of the distribution of $T_n(\alpha_0)$ under $P_{n,\alpha_0,\beta}$. [We assume here the existence of such a $c_{n,\beta}(\delta)$, but we do not insist that it is chosen as the smallest possible number satisfying the above condition, although this will usually be the case.] In other words, $c_{n,\sup}(\delta)$ is the "worst-case" critical value. While the resulting test, which rejects H_0 for

$$T_n(\alpha_0) > c_{n,\sup}(\delta), \tag{2}$$

certainly is a level δ test (i.e., has size $\leq \delta$), the conservatism caused by taking the supremum in (1) will often result in poor power properties, especially for values of β for which $c_{n,\beta}(\delta)$ is much smaller than $c_{n,\sup}(\delta)$. The test obtained from (1) and (2) above (more precisely, an asymptotic variant thereof) is what Andrews and Guggenberger [1] call a "size-corrected fixed critical value" test.[2]

An alternative idea, which has some intuitive appeal and which is much less conservative, is to use $c_{n,\hat{\beta}_n}(\delta)$ as a random critical value, where $\hat{\beta}_n$ is an estimator for β (taking its values in B), and to reject H_0 if

$$T_n(\alpha_0) > c_{n,\hat{\beta}_n}(\delta) \tag{3}$$

obtains (measurability of $c_{n,\hat{\beta}_n}(\delta)$ being assumed). This choice of critical value can be viewed as a parametric bootstrap procedure. Versions of $c_{n,\hat{\beta}_n}(\delta)$ have been considered by Williams [14] or, more recently, by Liu [9]. However,

$$P_{n,\alpha_0,\beta}\left(T_n(\alpha_0) > c_{n,\hat{\beta}_n}(\delta)\right) \geq P_{n,\alpha_0,\beta}\left(T_n(\alpha_0) > c_{n,\sup}(\delta)\right)$$

clearly holds for every β, indicating that the test using the random critical value $c_{n,\hat{\beta}_n}(\delta)$ may *not* be a level δ test, but may have size larger than δ. This was already noted by Loh [8]. A precise result in this direction, which is a variation of Theorem 2.1 in [8], is as follows:

Proposition 1 *Suppose that there exists a* $\beta_n^{\max} = \beta_n^{\max}(\delta)$ *such that* $c_{n,\beta_n^{\max}}(\delta) = c_{n,\sup}(\delta)$. *Then*

$$P_{n,\alpha_0,\beta_n^{\max}}\left(c_{n,\hat{\beta}_n}(\delta) < T_n(\alpha_0) \leq c_{n,\sup}(\delta)\right) > 0 \tag{4}$$

---

[2]While Andrews and Guggenberger [1] do not consider a finite-sample framework but rather a "moving-parameter" asymptotic framework, the underlying idea is nevertheless exactly the same.

implies

$$\sup_{\beta \in B} P_{n,\alpha_0,\beta} \left(T_n(\alpha_0) > c_{n,\hat{\beta}_n}(\delta) \right) > \delta, \tag{5}$$

i.e., the test using the random critical value $c_{n,\hat{\beta}_n}(\delta)$ does not have level δ. More generally, if $\hat{c}_n$ is any random critical value satisfying $\hat{c}_n \leq c_{n,\beta_n^{\max}}(\delta)(= c_{n,\sup}(\delta))$ with $P_{n,\alpha_0,\beta_n^{\max}}$-probability 1, then (4) still implies (5) if in both expressions $c_{n,\hat{\beta}_n}(\delta)$ is replaced by $\hat{c}_n$. [The result continues to hold if the random critical value $\hat{c}_n$ also depends on some additional randomization mechanism.]

Proof Observe that $c_{n,\hat{\beta}_n}(\delta) \leq c_{n,\sup}(\delta)$ always holds. But then the l.h.s. of (5) is bounded from below by

$$P_{n,\alpha_0,\beta_n^{\max}} \left(T_n(\alpha_0) > c_{n,\hat{\beta}_n}(\delta) \right)$$

$$= P_{n,\alpha_0,\beta_n^{\max}} \left(T_n(\alpha_0) > c_{n,\sup}(\delta) \right) + P_{n,\alpha_0,\beta_n^{\max}} \left(c_{n,\hat{\beta}_n}(\delta) < T_n(\alpha_0) \leq c_{n,\sup}(\delta) \right)$$

$$= P_{n,\alpha_0,\beta_n^{\max}} \left(T_n(\alpha_0) > c_{n,\beta_n^{\max}}(\delta) \right) + P_{n,\alpha_0,\beta_n^{\max}} \left(c_{n,\hat{\beta}_n}(\delta) < T_n(\alpha_0) \leq c_{n,\sup}(\delta) \right)$$

$$= \delta + P_{n,\alpha_0,\beta_n^{\max}} \left(c_{n,\hat{\beta}_n}(\delta) < T_n(\alpha_0) \leq c_{n,\sup}(\delta) \right) > \delta,$$

the last inequality holding in view of (4). The proof for the second claim is completely analogous. ∎

To better appreciate condition (4) consider the case where $c_{n,\beta}(\delta)$ is uniquely maximized at $\beta_n^{\max}$ and $P_{n,\alpha_0,\beta_n^{\max}}(\hat{\beta}_n \neq \beta_n^{\max})$ is positive. Then

$$P_{n,\alpha_0,\beta_n^{\max}}(c_{n,\hat{\beta}_n}(\delta) < c_{n,\sup}(\delta)) > 0$$

holds and therefore we can expect condition (4) to be satisfied, unless there exists a quite strange dependence structure between $\hat{\beta}_n$ and $T_n(\alpha_0)$. The same argument applies in the more general situation where there are multiple maximizers $\beta_n^{\max}$ of $c_{n,\beta}(\delta)$ as soon as $P_{n,\alpha_0,\beta_n^{\max}}(\hat{\beta}_n \notin \arg\max c_{n,\beta}(\delta)) > 0$ holds for one of the maximizers $\beta_n^{\max}$.

In the same vein, it is also useful to note that Condition (4) can equivalently be stated as follows: The conditional cumulative distribution function $P_{n,\alpha_0,\beta_n^{\max}}(T_n(\alpha_0) \leq \cdot \mid \hat{\beta}_n)$ of $T_n(\alpha_0)$ given $\hat{\beta}_n$ puts positive mass on the interval $(c_{n,\hat{\beta}_n}(\delta), c_{n,\sup}(\delta)]$ for a set of $\hat{\beta}_n$'s that has positive probability under $P_{n,\alpha_0,\beta_n^{\max}}$. [Also note that Condition (4) implies that $c_{n,\hat{\beta}_n}(\delta) < c_{n,\sup}(\delta)$ must hold with positive $P_{n,\alpha_0,\beta_n^{\max}}$-probability.] A sufficient condition for this then clearly is that for a set of $\hat{\beta}_n$'s of positive $P_{n,\alpha_0,\beta_n^{\max}}$-probability we have that (i) $c_{n,\hat{\beta}_n}(\delta) < c_{n,\sup}(\delta)$, and (ii) the conditional cumulative distribution function $P_{n,\alpha_0,\beta_n^{\max}}(T_n(\alpha_0) \leq \cdot \mid \hat{\beta}_n)$ puts positive mass on *every* non-empty interval. The analogous result holds for the

case where $\hat{c}_n$ replaces $c_{n,\hat{\beta}_n}(\delta)$ (and conditioning is w.r.t. $\hat{c}_n$), see Lemma 5 in the Appendix for a formal statement.

The observation that the test (3) based on the random critical value $c_{n,\hat{\beta}_n}(\delta)$ typically will not be a level δ test has led Loh [8] and subsequently Berger and Boos [2] and Silvapulle [13] to consider the following procedure (or variants thereof) which leads to a level δ test that is somewhat less "conservative" than the test given by (2)[3]: Let I_n be a random set in B satisfying

$$\inf_{\beta \in B} P_{n,\alpha_0,\beta} \left(\beta \in I_n \right) \geq 1 - \eta_n,$$

where $0 \leq \eta_n < \delta$. That is, I_n is a confidence set for the nuisance parameter β with infimal coverage probability not less than $1 - \eta_n$ (provided $\alpha = \alpha_0$). Define a random critical value via

$$c_{n,\eta_n,\text{Loh}}(\delta) = \sup_{\beta \in I_n} c_{n,\beta}(\delta - \eta_n). \tag{6}$$

Then we have

$$\sup_{\beta \in B} P_{n,\alpha_0,\beta} \left(T_n(\alpha_0) > c_{n,\eta_n,\text{Loh}}(\delta) \right) \leq \delta.$$

This can be seen as follows: For every $\beta \in B$

$$\begin{aligned} P_{n,\alpha_0,\beta} \left(T_n(\alpha_0) > c_{n,\eta_n,\text{Loh}}(\delta) \right) &= P_{n,\alpha_0,\beta} \left(T_n(\alpha_0) > c_{n,\eta_n,\text{Loh}}(\delta), \beta \in I_n \right) \\ &\quad + P_{n,\alpha_0,\beta} \left(T_n(\alpha_0) > c_{n,\eta_n,\text{Loh}}(\delta), \beta \notin I_n \right) \\ &\leq P_{n,\alpha_0,\beta} \left(T_n(\alpha_0) > c_{n,\beta}(\delta - \eta_n), \beta \in I_n \right) + \eta_n \\ &\leq P_{n,\alpha_0,\beta} \left(T_n(\alpha_0) > c_{n,\beta}(\delta - \eta_n) \right) + \eta_n \\ &= \delta - \eta_n + \eta_n = \delta. \end{aligned}$$

Hence, the random critical value $c_{n,\eta_n,\text{Loh}}(\delta)$ results in a test that is guaranteed to be level δ. In fact, its size can also be lower bounded by $\delta - \eta_n$ provided there exists a $\beta_n^{\max}(\delta - \eta_n)$ satisfying $c_{n,\beta_n^{\max}(\delta-\eta_n)}(\delta - \eta_n) = \sup_{\beta \in B} c_{n,\beta}(\delta - \eta_n)$: This follows since

$$\sup_{\beta \in B} P_{n,\alpha_0,\beta} \left(T_n(\alpha_0) > c_{n,\eta_n,\text{Loh}}(\delta) \right)$$

$$\geq \sup_{\beta \in B} P_{n,\alpha_0,\beta} \left(T_n(\alpha_0) > \sup_{\beta \in B} c_{n,\beta}(\delta - \eta_n) \right)$$

---

[3] Loh [8] actually considers the random critical value $c_{n,\eta_n,\text{Loh}^*}(\delta)$ given by $\sup_{\beta \in I_n} c_{n,\beta}(\delta)$, which typically does not lead to a level δ test in finite samples in view of Proposition 1 (since $c_{n,\eta_n,\text{Loh}^*}(\delta) \leq c_{n,\sup}(\delta)$). However, Loh [8] focuses on the case where $\eta_n \to 0$ and shows that then the size of the test converges to δ; that is, the test is asymptotically level δ if $\eta_n \to 0$. See also Remark 4.

$$= \sup_{\beta \in B} P_{n,\alpha_0,\beta} \left(T_n(\alpha_0) > c_{n,\beta_n^{\max}(\delta - \eta_n)}(\delta - \eta_n) \right)$$

$$\geq P_{n,\alpha_0,\beta_n^{\max}(\delta - \eta_n)} \left(T_n(\alpha_0) > c_{n,\beta_n^{\max}(\delta - \eta_n)}(\delta - \eta_n) \right)$$

$$= \delta - \eta_n. \tag{7}$$

The critical value (6) (or asymptotic variants thereof) has also been used in econometrics, e.g., by DiTraglia [4], McCloskey [10, 11], and Romano et al. [12].

The test based on the random critical value $c_{n,\eta_n,\mathrm{Loh}}(\delta)$ may have size strictly smaller than δ. This suggests that this test will not improve over the conservative test based on $c_{n,\sup}(\delta)$ for *all* values of β: We can expect that the test based on (6) will sacrifice some power when compared with the conservative test (2) when the true β is close to $\beta_n^{\max}(\delta)$ or $\beta_n^{\max}(\delta - \eta_n)$; however, we can often expect a power gain for values of β that are "far away" from $\beta_n^{\max}(\delta)$ and $\beta_n^{\max}(\delta - \eta_n)$, as we then typically will have that $c_{n,\eta_n,\mathrm{Loh}}(\delta)$ is smaller than $c_{n,\sup}(\delta)$. Hence, each of the two tests will typically have a power advantage over the other in certain parts of the parameter space B.

It is thus tempting to try to construct a test that has the power advantages of both these tests by choosing as a critical value the smaller one of the two critical values, i.e., by choosing

$$\hat{c}_{n,\eta_n,\min}(\delta) = \min \left(c_{n,\sup}(\delta), c_{n,\eta_n,\mathrm{Loh}}(\delta) \right) \tag{8}$$

as the critical value. While both critical values $c_{n,\sup}(\delta)$ and $c_{n,\eta_n,\mathrm{Loh}}(\delta)$ lead to level δ tests, this is, however, unfortunately not the case in general for the test based on the random critical value (8). To see why, note that by construction the critical value (8) satisfies

$$\hat{c}_{n,\eta_n,\min}(\delta) \leq c_{n,\sup}(\delta),$$

and hence can be expected to fall under the wrath of Proposition 1 given above. Thus it can be expected to not deliver a test that has level δ, but has a size that exceeds δ. So while the test based on the random critical value proposed in (8) will typically reject more often than the tests based on (2) or on (6), it does so by violating the size constraint. Hence it suffers from the same problems as the parametric bootstrap test (3). [We make the trivial observation that the lower bound (7) also holds if $\hat{c}_{n,\eta_n,\min}(\delta)$ instead of $c_{n,\eta_n,\mathrm{Loh}}(\delta)$ is used, since $\hat{c}_{n,\eta_n,\min}(\delta) \leq c_{n,\eta_n,\mathrm{Loh}}(\delta)$ holds.] As a point of interest we note that the construction (8) has actually been suggested in the literature, see McCloskey [10].[4] In fact, McCloskey [10] suggested a random critical value $\hat{c}_{n,\mathrm{McC}}(\delta)$ which is the minimum of critical values of the form (8) with η_n running through a finite set of values; it is thus less than or equal to the individual $\hat{c}_{n,\eta_n,\min}$'s, which exacerbates the size distortion problem even further.

---

[4]This construction is no longer suggested in [11].

While Proposition 1 shows that tests based on random critical values like $c_{n,\hat{\beta}_n}(\delta)$ or $\hat{c}_{n,\eta_n,\min}(\delta)$ will typically not have level δ, it leaves open the possibility that the overshoot of the size over δ may converge to zero as sample size goes to infinity, implying that the test would then be at least *asymptotically* of level δ. In sufficiently "regular" testing problems this will indeed be the case. However, for many testing problems where nuisance parameters are present such as testing post-model- selection, it turns out that this is typically *not* the case: In the next section we illustrate this by providing a prototypical example where the overshoot does *not* converge to zero for the tests based on $c_{n,\hat{\beta}_n}(\delta)$ or $\hat{c}_{n,\eta_n,\min}(\delta)$, and hence these tests are not level δ *even asymptotically*.

2 An Illustrative Example

In the following we shall—for the sake of exposition—use a very simple example to illustrate the issues involved. Consider the linear regression model

$$y_t = \alpha x_{t1} + \beta x_{t2} + \epsilon_t \qquad (1 \le t \le n) \tag{9}$$

under the "textbook" assumptions that the errors ϵ_t are i.i.d. $N(0, \sigma^2)$, $\sigma^2 > 0$, and the nonstochastic $n \times 2$ regressor matrix X has full rank (implying $n > 1$) and satisfies $X'X/n \to Q > 0$ as $n \to \infty$. The variables y_t, x_{ti}, as well as the errors ϵ_t can be allowed to depend on sample size n (in fact may be defined on a sample space that itself depends on n), but we do not show this in the notation. For simplicity, we shall also assume that the error variance σ^2 is known and equals 1. It will be convenient to write the matrix $(X'X/n)^{-1}$ as

$$(X'X/n)^{-1} = \begin{pmatrix} \sigma_{\alpha,n}^2 & \sigma_{\alpha,\beta,n} \\ \sigma_{\alpha,\beta,n} & \sigma_{\beta,n}^2 \end{pmatrix}.$$

The elements of the limit of this matrix will be denoted by $\sigma_{\alpha,\infty}^2$, etc. It will prove useful to define $\rho_n = \sigma_{\alpha,\beta,n}/(\sigma_{\alpha,n}\sigma_{\beta,n})$, i.e., ρ_n is the correlation coefficient between the least-squares estimators for α and β in model (9). Its limit will be denoted by ρ_∞. Note that $|\rho_\infty| < 1$ holds, since $Q > 0$ has been assumed.

As in [7] we shall consider two candidate models from which we select on the basis of the data: The unrestricted model denoted by U which uses both regressors x_{t1} and x_{t2}, and the restricted model denoted by R which uses only the regressor x_{t1} (and thus corresponds to imposing the restriction $\beta = 0$). The least-squares estimators for α and β in the unrestricted model will be denoted by $\hat{\alpha}_n(U)$ and $\hat{\beta}_n(U)$, respectively. The least-squares estimator for α in the restricted model will be denoted by $\hat{\alpha}_n(R)$, and we shall set $\hat{\beta}_n(R) = 0$. We shall decide between the competing models U and R depending on whether $|\sqrt{n}\hat{\beta}(U_n)/\sigma_{\beta,n}| > c$ or not, where $c > 0$ is a user-specified cut-off point independent of sample size (in line

with the fact that we consider conservative model selection). That is, we select the model $\hat{M}_n$ according to

$$\hat{M}_n = \begin{cases} U \text{ if } |\sqrt{n}\hat{\beta}_n(U)/\sigma_{\beta,n}| > c, \\ R \qquad \text{otherwise.} \end{cases}$$

We now want to test the hypothesis $H_0 : \alpha = \alpha_0$ versus $H_1 : \alpha > \alpha_0$ and we insist, for better or worse, on using the test-statistic

$$T_n(\alpha_0) = \left[n^{1/2} \left(\hat{\alpha}(R) - \alpha_0\right) / \left(\sigma_{\alpha,n} \left(1 - \rho_n^2\right)^{1/2}\right)\right] \mathbf{1}(\hat{M}_n = R)$$
$$+ \left[n^{1/2} \left(\hat{\alpha}(U) - \alpha_0\right) / \sigma_{\alpha,n}\right] \mathbf{1}(\hat{M}_n = U).$$

That is, depending on which of the two models has been selected, we insist on using the corresponding textbook test-statistic (for the known-variance case). While this could perhaps be criticized as somewhat simple-minded, it describes how such a test may be conducted in practice when model selection precedes the inference step. It is well known that if one uses this test-statistic and naively compares it to the usual normal-based quantiles acting as if the selected model were given a priori, this results in a test with severe size distortions, see, e.g., [5] and the references therein. Hence, while sticking with $T_n(\alpha_0)$ as the test-statistic, we now look for appropriate critical values in the spirit of the preceding section and discuss some of the proposals from the literature. Note that the situation just described fits into the framework of the preceding section with β as the nuisance parameter and $B = \mathbb{R}$.

Calculations similar to the ones in [7] show that the finite-sample distribution of $T_n(\alpha_0)$ under H_0 has a density that is given by

$$h_{n,\beta}(u) = \Delta \left(n^{1/2}\beta/\sigma_{\beta,n}, c\right) \phi \left(u + \rho_n \left(1 - \rho_n^2\right)^{-1/2} n^{1/2}\beta/\sigma_{\beta,n}\right)$$
$$+ \left(1 - \Delta \left(\left(1 - \rho_n^2\right)^{-1/2} \left(n^{1/2}\beta/\sigma_{\beta,n} + \rho_n u\right), \left(1 - \rho_n^2\right)^{-1/2} c\right)\right) \phi(u),$$

where $\Delta(a, b) = \Phi(a + b) - \Phi(a - b)$ and where ϕ and Φ denote the density and cdf, respectively, of a standard normal variate. Let $H_{n,\beta}$ denote the cumulative distribution function (cdf) corresponding to $h_{n,\beta}$.

Now, for given significance level δ, $0 < \delta < 1$, let $c_{n,\beta}(\delta) = H_{n,\beta}^{-1}(1 - \delta)$ as in the preceding section. Note that the inverse function exists, since $H_{n,\beta}$ is continuous and is strictly increasing as its density $h_{n,\beta}$ is positive everywhere. As in the preceding section let

$$c_{n,\text{sup}}(\delta) = \sup_{\beta \in \mathbb{R}} c_{n,\beta}(\delta) \qquad (10)$$

denote the conservative critical value (the supremum is actually a maximum in the interesting case $\delta \leq 1/2$ in view of Lemmata 6 and 7 in the Appendix).

Let $c_{n,\hat{\beta}_n(U)}(\delta)$ be the parametric bootstrap-based random critical value. With η satisfying $0 < \eta < \delta$, we also consider the random critical value

$$c_{n,\eta,\text{Loh}}(\delta) = \sup_{\beta \in I_n} c_{n,\beta}(\delta - \eta) \tag{11}$$

where

$$I_n = \left[\hat{\beta}_n(U) \pm n^{-1/2}\sigma_{\beta,n}\Phi^{-1}(1 - (\eta/2)) \right]$$

is an $1 - \eta$ confidence interval for β. [Again the supremum is actually a maximum.] We choose here η *independent* of n as in [4, 10, 11] and comment on sample size dependent η below. Furthermore define

$$\hat{c}_{n,\eta,\min}(\delta) = \min\left(c_{n,\text{sup}}(\delta), c_{n,\eta,\text{Loh}}(\delta)\right). \tag{12}$$

Recall from the discussion in Sect. 1 that these critical values have been used in the literature in the contexts of testing post-model-selection, post-moment-selection, or post-model-averaging. Among the critical values $c_{n,\text{sup}}(\delta)$, $c_{n,\hat{\beta}_n(U)}(\delta)$, $c_{n,\eta,\text{Loh}}(\delta)$, and $\hat{c}_{n,\eta,\min}(\delta)$, we already know that $c_{n,\text{sup}}(\delta)$ and $c_{n,\eta,\text{Loh}}(\delta)$ lead to tests that are valid level δ tests. We next confirm—as suggested by the discussion in the preceding section—that the random critical values $c_{n,\hat{\beta}_n(U)}(\delta)$ and $\hat{c}_{n,\eta,\min}(\delta)$ (at least for some choices of η) do *not* lead to tests that have level δ (i.e., their size is strictly larger than δ). Moreover, we also show that the sizes of the tests based on $c_{n,\hat{\beta}_n(U)}(\delta)$ or $\hat{c}_{n,\eta,\min}(\delta)$ do *not* converge to δ as $n \to \infty$, implying that the asymptotic sizes of these tests exceed δ. These results a fortiori also apply to any random critical value that does not exceed $c_{n,\hat{\beta}_n(U)}(\delta)$ or $\hat{c}_{n,\eta,\min}(\delta)$ (such as $\hat{c}_{n,\text{McC}}(\delta)$ in [10] or $c_{n,\eta,\text{Loh}^*}(\delta)$). In the subsequent theorem we consider for simplicity only the case $\rho_n \equiv \rho$, but the result extends to the more general case where ρ_n may depend on n.

Theorem 2 *Suppose $\rho_n \equiv \rho \neq 0$ and let $0 < \delta \leq 1/2$ be arbitrary. Then*

$$\inf_{n>1} \sup_{\beta \in \mathbb{R}} P_{n,\alpha_0,\beta}\left(T_n(\alpha_0) > c_{n,\hat{\beta}_n(U)}(\delta)\right) > \delta. \tag{13}$$

Furthermore, for each fixed η, $0 < \eta < \delta$, that is sufficiently small we have

$$\inf_{n>1} \sup_{\beta \in \mathbb{R}} P_{n,\alpha_0,\beta}\left(T_n(\alpha_0) > \hat{c}_{n,\eta,\min}(\delta)\right) > \delta. \tag{14}$$

Proof We first prove (14). Introduce the abbreviation $\gamma = n^{1/2}\beta/\sigma_{\beta,n}$ and define $\hat{\gamma}(U) = n^{1/2}\hat{\beta}(U)/\sigma_{\beta,n}$. Observe that the density $h_{n,\beta}$ (and hence the cdf $H_{n,\beta}$) depends on the nuisance parameter β only via γ, and otherwise is independent of sample size n (since $\rho_n = \rho$ is assumed). Let $\bar{h}_\gamma$ be the density of $T_n(\alpha_0)$ when expressed in the reparameterization γ. As a consequence, the quantiles satisfy

$c_{n,\beta}(v) = \bar{c}_\gamma(v)$ for every $0 < v < 1$, where $\bar{c}_\gamma(v) = \bar{H}_\gamma^{-1}(1-v)$ and $\bar{H}_\gamma$ denotes the cdf corresponding to $\bar{h}_\gamma$. Furthermore, for $0 < \eta < \delta$, observe that $c_{n,\eta,\text{Loh}}(\delta) = \sup_{\beta \in I_n} c_{n,\beta}(\delta - \eta)$ can be rewritten as

$$c_{n,\eta,\text{Loh}}(\delta) = \sup_{\gamma \in \left[\hat{\gamma}(U) \pm \Phi^{-1}(1-(\eta/2))\right]} \bar{c}_\gamma(\delta - \eta).$$

Now define $\gamma^{\max} = \gamma^{\max}(\delta)$ as a value of γ such that $\bar{c}_{\gamma^{\max}}(\delta) = \bar{c}_{\text{sup}}(\delta) := \sup_{\gamma \in \mathbb{R}} \bar{c}_\gamma(\delta)$. That such a maximizer exists follows from Lemmata 6 and 7 in the Appendix. Note that $\gamma^{\max}$ does not depend on n. Of course, $\gamma^{\max}$ is related to $\beta_n^{\max} = \beta_n^{\max}(\delta)$ via $\gamma^{\max} = n^{1/2}\beta_n^{\max}/\sigma_{\beta,n}$. Since $\bar{c}_{\text{sup}}(\delta) = \bar{c}_{\gamma^{\max}}(\delta)$ is strictly larger than

$$\lim_{|\gamma| \to \infty} \bar{c}_\gamma(\delta) = \Phi^{-1}(1-\delta)$$

in view of Lemmata 6 and 7 in the Appendix, we have for all sufficiently small η, $0 < \eta < \delta$, that

$$\lim_{|\gamma| \to \infty} \bar{c}_\gamma(\delta - \eta) = \Phi^{-1}(1-(\delta-\eta)) < \bar{c}_{\text{sup}}(\delta) = \bar{c}_{\gamma^{\max}}(\delta). \tag{15}$$

Fix such an η. Let now $\varepsilon > 0$ satisfy $\varepsilon < \bar{c}_{\text{sup}}(\delta) - \Phi^{-1}(1-(\delta-\eta))$. Because of the limit relation in the preceding display, we see that there exists $M = M(\varepsilon) > 0$ such that for $|\gamma| > M$ we have $\bar{c}_\gamma(\delta - \eta) < \bar{c}_{\text{sup}}(\delta) - \varepsilon$. Define the set

$$A = \left\{ x \in \mathbb{R} : |x| > \Phi^{-1}(1-(\eta/2)) + M \right\}.$$

Then on the event $\{\hat{\gamma}(U) \in A\}$ we have that $\hat{c}_{n,\eta,\min}(\delta) \leq \bar{c}_{\text{sup}}(\delta) - \varepsilon$. Furthermore, noting that $P_{n,\alpha_0,\beta_n^{\max}}\left(T_n(\alpha_0) > c_{n,\text{sup}}(\delta)\right) = P_{n,\alpha_0,\beta_n^{\max}}\left(T_n(\alpha_0) > \bar{c}_{\text{sup}}(\delta)\right) = \delta$, we have

$$\sup_{\beta \in \mathbb{R}} P_{n,\alpha_0,\beta}\left(T_n(\alpha_0) > \hat{c}_{n,\eta,\min}(\delta)\right) \geq P_{n,\alpha_0,\beta_n^{\max}}\left(T_n(\alpha_0) > \hat{c}_{n,\eta,\min}(\delta)\right)$$

$$= P_{n,\alpha_0,\beta_n^{\max}}\left(T_n(\alpha_0) > \bar{c}_{\text{sup}}(\delta)\right) + P_{n,\alpha_0,\beta_n^{\max}}\left(\hat{c}_{n,\eta,\min}(\delta) < T_n(\alpha_0) \leq \bar{c}_{\text{sup}}(\delta)\right)$$

$$\geq \delta + P_{n,\alpha_0,\beta_n^{\max}}\left(\hat{c}_{n,\eta,\min}(\delta) < T_n(\alpha_0) \leq \bar{c}_{\text{sup}}(\delta), \hat{\gamma}(U) \in A\right)$$

$$\geq \delta + P_{n,\alpha_0,\beta_n^{\max}}\left(\bar{c}_{\text{sup}}(\delta) - \varepsilon < T_n(\alpha_0) \leq \bar{c}_{\text{sup}}(\delta), \hat{\gamma}(U) \in A\right).$$

We are hence done if we can show that the probability in the last line is positive and independent of n. But this probability can be written as follows:[5]

$$P_{n,\alpha_0,\beta_n^{\max}} \left(\bar{c}_{\sup}(\delta) - \varepsilon < T_n(\alpha_0) \leq \bar{c}_{\sup}(\delta), \hat{\gamma}(U) \in A \right)$$

$$= P_{n,\alpha_0,\beta_n^{\max}} \left(\bar{c}_{\sup}(\delta) - \varepsilon < T_n(\alpha_0) \leq \bar{c}_{\sup}(\delta), \hat{\gamma}(U) \in A, |\hat{\gamma}(U)| \leq c \right)$$

$$+ P_{n,\alpha_0,\beta_n^{\max}} \left(\bar{c}_{\sup}(\delta) - \varepsilon < T_n(\alpha_0) \leq \bar{c}_{\sup}(\delta), \hat{\gamma}(U) \in A, |\hat{\gamma}(U)| > c \right)$$

$$= P_{n,\alpha_0,\beta_n^{\max}} \left(\bar{c}_{\sup}(\delta) \geq n^{1/2} \left(\hat{\alpha}(R) - \alpha_0 \right) / \left(\sigma_{\alpha,n} \left(1 - \rho^2 \right)^{1/2} \right) \right.$$

$$\left. > \bar{c}_{\sup}(\delta) - \varepsilon, \hat{\gamma}(U) \in A, |\hat{\gamma}(U)| \leq c \right)$$

$$+ P_{n,\alpha_0,\beta_n^{\max}} \left(\bar{c}_{\sup}(\delta) \geq n^{1/2} \left(\hat{\alpha}(U) - \alpha_0 \right) / \sigma_{\alpha,n} \right.$$

$$\left. > \bar{c}_{\sup}(\delta) - \varepsilon, \hat{\gamma}(U) \in A, |\hat{\gamma}(U)| > c \right)$$

$$= \left[\Phi(\bar{c}_{\sup}(\delta) + \rho \left(1 - \rho^2 \right)^{-1/2} \gamma^{\max}) - \Phi(\bar{c}_{\sup}(\delta) + \rho \left(1 - \rho^2 \right)^{-1/2} \gamma^{\max} - \varepsilon) \right]$$

$$\times \Pr \left(Z_2 \in A, |Z_2| \leq c \right) + \Pr \left(\bar{c}_{\sup}(\delta) \geq Z_1 > \bar{c}_{\sup}(\delta) - \varepsilon, Z_2 \in A, |Z_2| > c \right),$$

where we have made use of independence of $\hat{\alpha}(R)$ and $\hat{\gamma}(U)$, cf. Lemma A.1 in [6], and of the fact that $n^{1/2} \left(\hat{\alpha}(R) - \alpha_0 \right)$ is distributed as $N(-\sigma_{\alpha,n}\rho\gamma^{\max}, \sigma_{\alpha,n}^2 \left(1 - \rho^2 \right))$ under $P_{n,\alpha_0,\beta_n^{\max}}$. Furthermore, we have used the fact that $\left(n^{1/2} \left(\hat{\alpha}(U) - \alpha_0 \right) / \sigma_{\alpha,n}, \hat{\gamma}(U) \right)'$ is under $P_{n,\alpha_0,\beta_n^{\max}}$ distributed as $(Z_1, Z_2)'$ where

$$(Z_1, Z_2)' \sim N \left((0, \gamma^{\max})', \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

which is a non-singular normal distribution, since $|\rho| < 1$. It is now obvious from the final expression in the last but one display that the probability in question is strictly positive and is independent of n. This proves (14).

We turn to the proof of (13). Observe that $c_{n,\hat{\beta}_n(U)}(\delta) = \bar{c}_{\hat{\gamma}(U)}(\delta)$ and that

$$\bar{c}_{\sup}(\delta) = \bar{c}_{\gamma^{\max}}(\delta) > \lim_{|\gamma| \to \infty} \bar{c}_\gamma(\delta) = \Phi^{-1}(1 - \delta)$$

in view of Lemmata 6 and 7 in the Appendix. Choose $\varepsilon > 0$ to satisfy $\varepsilon < \bar{c}_{\sup}(\delta) - \Phi^{-1}(1 - \delta)$. Because of the limit relation in the preceding display, we see that there exists $M = M(\varepsilon) > 0$ such that for $|\gamma| > M$ we have $\bar{c}_\gamma(\delta) < \bar{c}_{\sup}(\delta) - \varepsilon$. Define the set

$$B = \{x \in \mathbb{R} : |x| > M\}.$$

---

[5]The corresponding calculation in previous versions of this paper had erroneously omitted the term $\rho \left(1 - \rho^2 \right)^{-1/2} \gamma^{\max}$ from the expression on the far right-hand side of the subsequent display. This is corrected here by accounting for this term. Alternatively, one could drop the probability involving $|\hat{\gamma}(U)| \leq c$ altogether from the proof and work with the resulting lower bound.

Then on the event $\{\hat{\gamma}(U) \in B\}$ we have that $c_{n,\hat{\beta}_n(U)}(\delta) = \bar{c}_{\hat{\gamma}(U)}(\delta) \le \bar{c}_{\mathrm{sup}}(\delta) - \varepsilon$. The rest of the proof is then completely analogous to the proof of (14) with the set A replaced by B.   ∎

Remark 3

(i) Inspection of the proof shows that (14) holds for every η, $0 < \eta < \delta$, that satisfies (15).

(ii) It is not difficult to show that the suprema in (13) and (14) actually do *not* depend on n.

Remark 4 If we allow η to depend on n, we may choose $\eta = \eta_n \to 0$ as $n \to \infty$. Then the test based on $\hat{c}_{n,\eta_n,\mathrm{min}}(\delta)$ still has a size that strictly overshoots δ for every n, but the overshoot will go to zero as $n \to \infty$. While this test then "approaches" the conservative test that uses $c_{n,\mathrm{sup}}(\delta)$, it does not respect the level for any finite-sample size. [The same can be said for Loh's [8] original proposal $c_{n,\eta_n,\mathrm{Loh}^*}(\delta)$, cf. footnote 3.] Contrast this with the test based on $c_{n,\eta_n,\mathrm{Loh}}(\delta)$ which holds the level for each n, and also "approaches" the conservative test if $\eta_n \to 0$. Hence, there seems to be little reason for preferring $\hat{c}_{n,\eta_n,\mathrm{min}}(\delta)$ (or $c_{n,\eta_n,\mathrm{Loh}^*}(\delta)$) to $c_{n,\eta_n,\mathrm{Loh}}(\delta)$ in this scenario where $\eta_n \to 0$.

Appendix

Lemma 5 *Suppose a random variable $\hat{c}_n$ satisfies $\mathrm{Pr}\,(\hat{c}_n \le c^*) = 1$ for some real number c^* as well as $\mathrm{Pr}\,(\hat{c}_n < c^*) > 0$. Let S be real-valued random variable. If for every non-empty interval J in the real line*

$$\mathrm{Pr}\,(S \in J \mid \hat{c}_n) > 0 \tag{16}$$

holds almost surely, then

$$\mathrm{Pr}\,(\hat{c}_n < S \le c^*) > 0.$$

The same conclusion holds if in (16) the conditioning variable $\hat{c}_n$ is replaced by some variable w_n, say, provided that $\hat{c}_n$ is a measurable function of w_n.

Proof Clearly

$$\mathrm{Pr}\,(\hat{c}_n < S \le c^*) = E\left[\mathrm{Pr}\,(S \in (\hat{c}_n, c^*] \mid \hat{c}_n)\right] = E\left[\mathrm{Pr}\,(S \in (\hat{c}_n, c^*] \mid \hat{c}_n)\,\mathbf{1}\,(\hat{c}_n < c^*)\right],$$

the last equality being true, since the first term in the product is zero on the event $\hat{c}_n = c^*$. Now note that the first factor in the expectation on the far right-hand side of the above equality is positive almost surely by (16) on the event $\{\hat{c}_n < c^*\}$, and that the event $\{\hat{c}_n < c^*\}$ has positive probability by assumption.   ∎

Recall that $\bar{c}_\gamma(v)$ has been defined in the proof of Theorem 2.

Lemma 6 *Assume $\rho_n \equiv \rho \neq 0$. Suppose $0 < v < 1$. Then the map $\gamma \to \bar{c}_\gamma(v)$ is continuous on $\mathbb{R}$. Furthermore, $\lim_{\gamma\to\infty} \bar{c}_\gamma(v) = \lim_{\gamma\to-\infty} \bar{c}_\gamma(v) = \Phi^{-1}(1-v)$.*

Proof If $\gamma_l \to \gamma$, then $\bar{h}_{\gamma_l}$ converges to $\bar{h}_\gamma$ pointwise on $\mathbb{R}$. By Scheffé's Lemma, $\bar{H}_{\gamma_l}$ then converges to $\bar{H}_\gamma$ in total variation distance. Since $\bar{H}_\gamma$ is strictly increasing on $\mathbb{R}$, convergence of the quantiles $\bar{c}_{\gamma_l}(v)$ to $\bar{c}_\gamma(v)$ follows. The second claim follows by the same argument observing that $\bar{h}_\gamma$ converges pointwise to a standard normal density for $\gamma \to \pm\infty$. ∎

Lemma 7 *Assume $\rho_n \equiv \rho \neq 0$.*

(i) *Suppose $0 < v \leq 1/2$. Then for some $\gamma \in \mathbb{R}$ we have that $\bar{c}_\gamma(v)$ is larger than $\Phi^{-1}(1-v)$.*

(ii) *Suppose $1/2 \leq v < 1$. Then for some $\gamma \in \mathbb{R}$ we have that $\bar{c}_\gamma(v)$ is smaller than $\Phi^{-1}(1-v)$.*

Proof Standard regression theory gives

$$\hat{\alpha}_n(U) = \hat{\alpha}_n(R) + \rho\sigma_{\alpha,n}\hat{\beta}_n(U)/\sigma_{\beta,n},$$

with $\hat{\alpha}_n(R)$ and $\hat{\beta}_n(U)$ being independent; for the latter cf., e.g., [6], Lemma A.1. Consequently, it is easy to see that the distribution of $T_n(\alpha_0)$ under $P_{n,\alpha_0,\beta}$ is the same as the distribution of

$$T' = T'(\rho,\gamma) = \left(\sqrt{1-\rho^2}W + \rho Z\right)\mathbf{1}\{|Z+\gamma| > c\}$$

$$+ \left(W - \rho\frac{\gamma}{\sqrt{1-\rho^2}}\right)\mathbf{1}\{|Z+\gamma| \leq c\},$$

where, as before, $\gamma = n^{1/2}\beta/\sigma_{\beta,n}$, and where W and Z are independent standard normal random variables.

We now prove (i): Let q be shorthand for $\Phi^{-1}(1-v)$ and note that $q \geq 0$ holds by the assumption on v. It suffices to show that $\Pr(T' \leq q) < \Phi(q)$ for some γ. We can now write

$$\Pr\left(T' \leq q\right) = \Pr\left(\sqrt{1-\rho^2}W + \rho Z \leq q\right) - \Pr\left(|Z+\gamma| \leq c, W \leq \frac{q-\rho Z}{\sqrt{1-\rho^2}}\right)$$

$$+ \Pr\left(|Z+\gamma| \leq c, W \leq q + \frac{\rho\gamma}{\sqrt{1-\rho^2}}\right)$$

$$= \Phi(q) - \Pr(A) + \Pr(B).$$

Here, A and B are the events given in terms of W and Z. Picturing these two events as subsets of the plane (with the horizontal axis corresponding to Z and the vertical axis corresponding to W), we see that A corresponds to the vertical band where

$|Z + \gamma| \leq c$, truncated above the line where $W = (q - \rho Z)/\sqrt{1 - \rho^2}$; similarly, B corresponds to the same vertical band $|Z + \gamma| \leq c$, truncated now above the horizontal line where $W = q + \rho\gamma/\sqrt{1 - \rho^2}$.

We first consider the case where $\rho > 0$ and distinguish two cases:

Case 1: $\rho c \leq \left(1 - \sqrt{1 - \rho^2}\right) q.$

In this case the set B is contained in A for every value of γ, with $A \backslash B$ being a set of positive Lebesgue measure. Consequently, $\Pr(A) > \Pr(B)$ holds for every γ, proving the claim.

Case 2: $\rho c > \left(1 - \sqrt{1 - \rho^2}\right) q.$

In this case choose γ so that $-\gamma - c \geq 0$, and, in addition, such that also $(q - \rho(-\gamma - c))/\sqrt{1 - \rho^2} < 0$, which is clearly possible. Recalling that $\rho > 0$, note that the point where the line $W = (q - \rho Z)/\sqrt{1 - \rho^2}$ intersects the horizontal line $W = q + \rho\gamma/\sqrt{1 - \rho^2}$ has as its first coordinate $Z = -\gamma + (q/\rho)(1 - \sqrt{1 - \rho^2})$, implying that the intersection occurs in the right half of the band where $|Z + \gamma| \leq c$. As a consequence, $\Pr(B) - \Pr(A)$ can be written as follows:

$$\Pr(B) - \Pr(A) = \Pr(B \backslash A) - \Pr(A \backslash B)$$

where

$$B \backslash A = \left\{ -\gamma + (q/\rho)(1 - \sqrt{1 - \rho^2}) \leq Z \leq -\gamma + c, \right.$$
$$\left. (q - \rho Z)/\sqrt{1 - \rho^2} < W \leq q + \rho\gamma/\sqrt{1 - \rho^2} \right\}$$

and

$$A \backslash B = \left\{ -\gamma - c \leq Z \leq -\gamma + (q/\rho)(1 - \sqrt{1 - \rho^2}), \right.$$
$$\left. q + \rho\gamma/\sqrt{1 - \rho^2} < W \leq (q - \rho Z)/\sqrt{1 - \rho^2} \right\}.$$

Picturing $A \backslash B$ and $B \backslash A$ as subsets of the plane as in the preceding paragraph, we see that these events correspond to two triangles, where the triangle corresponding to $A \backslash B$ is larger than or equal (in Lebesgue measure) to that corresponding to $B \backslash A$. Since γ was chosen to satisfy $-\gamma - c \geq 0$ and $(q - \rho(-\gamma - c))/\sqrt{1 - \rho^2} < 0$, we see that each point in the triangle corresponding to $A \backslash B$ is closer to the origin than any point in the triangle corresponding to $B \backslash A$. Because the joint Lebesgue density of (Z, W), i.e., the bivariate standard Gaussian density, is spherically symmetric and radially monotone, it follows that $\Pr(B \backslash A) - \Pr(A \backslash B) < 0$, as required.

The case $\rho < 0$ follows because $T'(\rho, \gamma)$ has the same distribution as $T'(-\rho, -\gamma)$.

Part (ii) follows, since $T'(\rho, \gamma)$ has the same distribution as $-T'(-\rho, \gamma)$.   ■

Remark 8 If $\rho_n \equiv \rho \neq 0$ and $v = 1/2$, then $\bar{c}_0(1/2) = \Phi^{-1}(1/2) = 0$, since $\bar{h}_0$ is symmetric about zero.

Remark 9 If $\rho_n \equiv \rho = 0$, then $T_n(\alpha_0)$ is standard normally distributed for every value of β, and hence $\bar{c}_\gamma(v) = \Phi^{-1}(1 - v)$ holds for every γ and v.

References

1. Andrews, D.W.K., Guggenberger, P.: Hybrid and size-corrected subsampling methods. Econometrica **77**, 721–762 (2009)
2. Berger, R.L., Boos, D.D.: P values maximized over a confidence set for the nuisance parameter. J. Am. Stat. Assoc. **89** 1012–1016 (1994)
3. Bickel, P.J., Doksum, K.A.: Mathematical Statistics: Basic Ideas and Selected Topics. Holden-Day, Oakland (1977)
4. DiTraglia, F.J.: Using invalid instruments on purpose: focused moment selection and averaging for GMM. Working Paper, Version November 9, 2011 (2011)
5. Kabaila, P., Leeb, H.: On the large-sample minimal coverage probability of confidence intervals after model selection. J. Am. Stat. Assoc. **101**, 619–629 (2006)
6. Leeb, H., Pötscher, B.M.: The finite-sample distribution of post-model-selection estimators and uniform versus non-uniform approximations. Economet. Theor. **19**, 100–142 (2003)
7. Leeb, H., Pötscher, B.M.: Model selection and inference: facts and fiction. Economet. Theor. **21**, 29–59 (2005)
8. Loh, W.-Y.: A new method for testing separate families of hypotheses. J. Am. Stat. Assoc. **80**, 362–368 (1985)
9. Liu, C.-A.: A plug-in averaging estimator for regressions with heteroskedastic errors. Working Paper, Version October 29, 2011 (2011)
10. McCloskey, A.: Powerful procedures with correct size for test statistics with limit distributions that are discontinuous in some parameters. Working Paper, Version October 2011 (2011)
11. McCloskey, A.: Bonferroni-based size correction for nonstandard testing problems. Working Paper, Brown University (2012)
12. Romano, J.P., Shaikh, A., Wolf, M.: A practical Two-step method for testing moment inequalities. Econometrica **82**, 1979–2002 (2014)
13. Silvapulle, M.J.: A test in the presence of nuisance parameters. J. Am. Stat. Assoc. **91**, 1690–1693 (1996) (Correction, ibidem 92 (1997) 801)
14. Williams, D.A.: Discrimination between regression models to determine the pattern of enzyme synthesis in synchronous cell cultures. Biometrics **26**, 23–32 (1970)