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

Hannes Leeb and Benedikt M. Pötscher
Department of Statistics, University of Vienna

Preliminary version: May 25, 2012
This version: September 20, 2012

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 and show that some of the proposals in the econometrics literature lead to tests that do not have the claimed size properties.

1 Introduction

Suppose we have a sequence of statistical experiments $\mathfrak{E}_n$ 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 Euclidean space. Suppose the researcher wants to base a test for the null-hypothesis $H_0 : \alpha = \alpha_0$ on 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 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., Bickel and Doksum (1977), p.170) to deal with this problem is to choose as critical value

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

[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.

where $0 < \delta < 1$ and where $c_{n,\beta}(\delta)$ satisfies $P_{n,\alpha_0,\beta}\left(T_n(\alpha_0) > c_{n,\beta}(\delta)\right) = \delta$, 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.] 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 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 (and an asymptotic variant thereof) is precisely what Andrews and Guggenberger (2009) call a "size-corrected fixed critical value" test.

An obvious alternative idea, 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. 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 (1985). A precise result in this direction, which is a variation of Theorem 2.1 in Loh (1985) 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*

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

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 also holds if the random critical values* $\hat{c}_n$ *also depend 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 not less than

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

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 $P_{n,\alpha_0,\beta_n^{\max}}(c_{n,\hat{\beta}_n}(\delta) < c_{n,\sup}(\delta))$ is positive and therefore we can expect condition (4) to hold, 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 $\beta_n^{\max}$ is not unique but $P_{n,\alpha_0,\beta_n^{\max}}(\hat{\beta}_n \notin \arg\max c_{n,\beta}(\delta)) > 0$ holds.

In the same vein, it is also useful to note that Condition (4) can equivalently be stated as follows: The conditional distribution $P_{n,\alpha_0,\beta_n^{\max}}\left(T_n(\alpha_0) \leq \cdot \mid \hat{\beta}_n\right)$ 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 $\hat{\beta}_n$ 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 distribution $P_{n,\alpha_0,\beta_n^{\max}}\left(T_n(\alpha_0) \leq \cdot \mid \hat{\beta}_n\right)$ 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 4 in the Appendix for a formal statement.

The observation, that the test (3) based on the critical value $c_{n,\hat{\beta}_n}(\delta)$ typically will not be a level δ test, has led Loh (1985) and subsequently Berger and Boos (1994) and Silvapulle (1996) to consider the following procedure (or variants thereof) which leads to a level δ test, but is somewhat less "conservative" than the test given by (2):

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 $0 \leq \eta_n < \delta$. I.e., 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

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

Then we have

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

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

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

Hence, the critical value $c_{n,\eta_n,Loh}(\delta)$ 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

$$\begin{aligned}
&\sup_{\beta \in B} P_{n,\alpha_0,\beta}\left(T_n(\alpha_0) > c_{n,\eta_n,Loh}(\delta)\right) \\
\geq\ &\sup_{\beta \in B} P_{n,\alpha_0,\beta}\left(T_n(\alpha_0) > \sup_{\beta \in B} c_{n,\beta}(\delta - \eta_n)\right) \\
=\ &\sup_{\beta \in B} P_{n,\alpha_0,\beta}\left(T_n(\alpha_0) > c_{n,\beta_n^{\max}(\delta-\eta_n)}(\delta - \eta_n)\right) \\
\geq\ &P_{n,\alpha_0,\beta_n^{\max}(\delta-\eta_n)}\left(T_n(\alpha_0) > c_{n,\beta_n^{\max}(\delta-\eta_n)}(\delta - \eta_n)\right) \\
=\ &\delta - \eta_n. \qquad (7)
\end{aligned}$$

Apparently unaware of Loh (1985), Berger and Boos (1994), and Silvapulle (1996), the just given construction of a critical value has also been suggested by DiTraglia (2011) and McCloseky (2011).

The test based on the random critical value $c_{n,\eta_n,Loh}(\delta)$ 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 $c_{n,\eta_n,Loh}(\delta)$ 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.

In an attempt to get the power advantages of both tests, McCloseky (2011) suggested (in the context of testing post model selection) to use random critical values of the form

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

In fact, his proposal, $\hat{c}_{n,McC}(\delta)$ say, is even smaller, and obtained by taking the minimum of critical values of the form (8) when η_n runs through a finite set of values. However, 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 not to deliver a test that has level δ, but has a size that exceeds δ. [In fact, McCloseky's (2011) suggestion $\hat{c}_{n,McC}(\delta)$ being even less than or equal to $\hat{c}_{n,\eta_n,\min}$ exacerbates the problem.] So while McCloseky's proposal will 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 $c_{n,\eta_n,Loh}(\delta)$ is used (since $\hat{c}_{n,\eta_n,\min}(\delta) \leq c_{n,\eta_n,Loh}(\delta)$ holds).

While the above proposition tells us that the test based on the random critical values figuring in the proposition, 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 and hence the test would be at least asymptotically of level δ. In sufficiently "regular" testing problems this will indeed be the case. However, we next provide an 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

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

under the "textbook" assumptions that the errors ϵ_t are i.i.d. $N(0, \sigma^2)$, $\sigma^2 > 0$, 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

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

The elements of the limit of this matrix will be denoted by $\sigma^2_{\alpha,\infty}$, 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 Leeb and Pötscher (2005) 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 the test-statistic $|\sqrt{n}\hat{\beta}(U_n)/\sigma_{\beta,n}| > c$ 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

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

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 as a test-statistic

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

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 is clearly 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., Kabaila and Leeb (2006) and 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 in 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 Leeb and Pötscher (2005) show that the finite-sample distribution of $T_n(\alpha_0)$ under H_0 has a density that is given by

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

where $\Delta(a, b) = \Phi(a + b) - \Phi(a - b)$. Let $H_{n,\beta}$ denote the corresponding cumulative distribution function.

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

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

denote the conservative critical value (the supremum is actually a maximum in the interesting case $\delta \leq 1/2$ in view of Lemmata 5 and 6 in the Appendix). Let $c_{n,\hat{\beta}_n(U)}(\delta)$ be the parametric bootstrap based random critical value. With η satisfying $0 \leq \eta < \delta$, we also consider the random critical value

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

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 McCloseky (2011) and DiTraglia (2011) and comment on sample size dependent η below. Furthermore define

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

In the context of testing post-model-selection (an asymptotic version of) the critical value given in (10) has been considered in Andrews and Guggenberger (2009) and the corresponding test is called a "size-corrected fixed-critical-value test". In the same context, the critical value (11) is considered in DiTraglia (2011) and McCloseky (2011), and the critical value (12) is proposed in McCloseky (2011) (more precisely, the critical value $\hat{c}_{n,McC}(\delta)$ defined in McCloseky (2011) is less than or equal to (12)). In the closely related context of testing post-model-averaging, Liu (2011) considered the parametric bootstrap-based critical value, i.e., the analogue of $c_{n,\hat{\beta}_n(U)}(\delta)$.

While the critical values (10) and (11) lead to tests that are valid level δ tests (and have been proposed in the statistics literature much earlier as discussed in the preceding section), we next show that – as suggested by the discussion in the preceding section – the proposals by McCloseky (2011) and Liu (2011) do not lead to tests that have level δ; furthermore, we not only show that the overshoot of the size of these tests over δ is strictly positive, we also show that the overshoot does *not* converge to zero as sample size goes to infinity. [In this preliminary version we show this only for some choices of η for McCloseky' procedure, but a more general result can be established.] This casts severe doubt on the results in Liu (2011) and McCloseky (2011). For simplicity the next theorem considers 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*

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

Furthermore, there exist $0 < \eta < \delta$ such that we have

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

and consequently also

$$\inf_{n>1} \sup_{\beta\in\mathbb{R}} P_{n,\alpha_0,\beta}\left(T_n(\alpha_0) > \hat{c}_{n,McC}(\delta)\right) > \delta. \tag{15}$$

In fact, the suprema in the above displays actually do not *depend on* n.

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 $0 \le \eta < \delta$, observe that $c_{n,\eta,Loh}(\delta) = \sup_{\beta\in I_n} c_{n,\beta}(\delta-\eta)$ can be rewritten as

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

Now define $\gamma^{\max} = \gamma^{\max}(\delta)$ as a value of γ such that $\bar{c}_{\gamma^{\max}}(\delta) = \bar{c}_{\sup}(\delta) := \sup_{\gamma\in\mathbb{R}} \bar{c}_\gamma(\delta)$. That such a maximizer exists follows form Lemmata 5 and 6 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 $\bar{c}_{\sup}(\delta) = \bar{c}_{\gamma^{\max}}(\delta)$ is strictly larger than

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

in view of Lemmata 5 and 6 in the Appendix, we can find a (sufficiently small) η, $0 < \eta < \delta$, such that

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

Let now $\varepsilon > 0$ satisfy $\varepsilon < \bar{c}_{\sup}(\delta) - \Phi^{-1}(1-(\delta-\eta))$. By continuity of $\bar{c}_\gamma(\delta-\eta)$ w.r.t. γ we see that there exists $M = M(\varepsilon) > 0$ such that for $|\gamma| > M$ we have $\bar{c}_\gamma(\delta-\eta) < \bar{c}_{\sup}(\delta) - \varepsilon$. 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 $\hat{c}_{n,\eta,\min}(\delta) \le \bar{c}_{\sup}(\delta) - \varepsilon$. Furthermore, noting that $P_{n,\alpha_0,\beta_n^{\max}}\left(T_n(\alpha_0) > c_{n,\sup}(\delta)\right) = P_{n,\alpha_0,\beta_n^{\max}}\left(T_n(\alpha_0) > \bar{c}_{\sup}(\delta)\right) = \delta$, we have

$$\begin{aligned}
& \sup_{\beta\in\mathbb{R}} P_{n,\alpha_0,\beta}\left(T_n(\alpha_0) > \hat{c}_{n,\eta,\min}(\delta)\right) \ge P_{n,\alpha_0,\beta_n^{\max}}\left(T_n(\alpha_0) > \hat{c}_{n,\eta,\min}(\delta)\right) = \\
& P_{n,\alpha_0,\beta_n^{\max}}\left(T_n(\alpha_0) > \bar{c}_{\sup}(\delta)\right) + P_{n,\alpha_0,\beta_n^{\max}}\left(\hat{c}_{n,\eta,\min}(\delta) < T_n(\alpha_0) \le \bar{c}_{\sup}(\delta)\right) \\
\ge\ & \delta + P_{n,\alpha_0,\beta_n^{\max}}\left(\hat{c}_{n,\eta,\min}(\delta) < T_n(\alpha_0) \le \bar{c}_{\sup}(\delta), \hat{\gamma}(U) \in A\right) \\
\ge\ & \delta + P_{n,\alpha_0,\beta_n^{\max}}\left(\bar{c}_{\sup}(\delta) - \varepsilon < T_n(\alpha_0) \le \bar{c}_{\sup}(\delta), \hat{\gamma}(U) \in A\right).
\end{aligned}$$

We are hence done if we can show that the probability in the last line is positive, observing that this probability clearly is independent of n since under $P_{n,\alpha_0,\beta_n^{\max}}$

the statistic $T_n(\alpha_0)$ has cdf $\bar{H}_{\gamma^{\max}}$. Now a simple (perhaps loose) lower bound is as follows

$$\begin{aligned}
& P_{n,\alpha_0,\beta_n^{\max}}\left(\bar{c}_{\sup}(\delta)-\varepsilon < T_n(\alpha_0) \leq \bar{c}_{\sup}(\delta), \hat{\gamma}(U) \in A\right) \\
= \; & P_{n,\alpha_0,\beta_n^{\max}}\left(\bar{c}_{\sup}(\delta)-\varepsilon < T_n(\alpha_0) \leq \bar{c}_{\sup}(\delta), \hat{\gamma}(U) \in A, |\hat{\gamma}(U)| \leq c\right) \\
& + P_{n,\alpha_0,\beta_n^{\max}}\left(\bar{c}_{\sup}(\delta)-\varepsilon < T_n(\alpha_0) \leq \bar{c}_{\sup}(\delta), \hat{\gamma}(U) \in A, |\hat{\gamma}(U)| > c\right) \\
= \; & P_{n,\alpha_0,\beta_n^{\max}}\Big(\bar{c}_{\sup}(\delta) \geq n^{1/2}\left(\hat{\alpha}(R)-\alpha_0\right) / \left(\sigma_{\alpha,n}\left(1-\rho^2\right)^{1/2}\right) > \\
& \qquad \bar{c}_{\sup}(\delta)-\varepsilon, \hat{\gamma}(U) \in A, |\hat{\gamma}(U)| \leq c\Big) \\
& + P_{n,\alpha_0,\beta_n^{\max}}\Big(\bar{c}_{\sup}(\delta) \geq n^{1/2}\left(\hat{\alpha}(U)-\alpha_0\right)/\sigma_{\alpha,n} > \\
& \qquad \bar{c}_{\sup}(\delta)-\varepsilon, \hat{\gamma}(U) \in A, |\hat{\gamma}(U)| > c\Big) \\
= \; & \left[\Phi(\bar{c}_{\sup}(\delta)) - \Phi(\bar{c}_{\sup}(\delta)-\varepsilon)\right] \Pr\left(Z_2 \in A, |Z_2| \leq c\right) \\
& + \Pr\left(\bar{c}_{\sup}(\delta) \geq Z_1 > \bar{c}_{\sup}(\delta)-\varepsilon, Z_2 \in A, |Z_2| > c\right)
\end{aligned}$$

where we have made use of independence of $\hat{\alpha}(R)$ and $\hat{\gamma}(U)$, cf. Lemma A.1 in Leeb and Pötscher (2003), and where

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

is a non-singular normal distribution since $|\rho| < 1$. Now it is obvious that the probability in the last line of the above display is strictly positive and is independent of n. This proves (14). Since $\hat{c}_{n,McC}(\delta) \leq \hat{c}_{n,\eta,\min}(\delta)$ the result (15) follows immediately.

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 5 and 6 in the Appendix. Choose $\varepsilon > 0$ to satisfy $\varepsilon < \bar{c}_{\sup}(\delta) - \Phi^{-1}(1-\delta)$. By continuity of $\bar{c}_{\gamma}(\delta)$ w.r.t. γ 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

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

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

Remark 3 If we allow η to depend on n, we may choose $\eta = \eta_n \to 1$ as $n \to \infty$. Then the test based on $\hat{c}_{n,\eta,\min}(\delta)$ still has a positive overshoot for every sample size, but the overshoot will go to zero as $n \to \infty$. But the test then "approaches" the conservative test that uses $c_{\sup}(\delta)$, and does not respect the level for any finite sample size. Contrast this with the test based on $c_{n,\eta,Loh}(\delta)$ which holds the level at each sample size, and also "approaches" the conservative test if $\eta_n \to 1$. Hence, there seems to be little reason for preferring $\hat{c}_{n,\eta,\min}(\delta)$ to $c_{n,\eta,Loh}(\delta)$ in this scenario.

3 References

Andrews, D. W. K. & P. Guggenberger (2009): Hybrid and Size-Corrected Subsampling Methods. *Econometrica* 77, 721-762.

Bickel, P. J. & K. A. Doksum (1977): *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden-Day, Oakland.

Berger, R. L. & D. D. Boos (1994): P Values Maximized Over a Confidence Set for the Nuisance Parameter, *Journal of the American Statistical Association* 89, 1012-1016.

DiTraglia, F. J. (2011): Using Invalid Instruments on Purpose: Focused Moment Selection and Averaging GMM. Working Paper, Version November 9, 2011.

Kabaila, P. & H. Leeb (2006): On the Large-Sample Minimal Coverage Probability of Confidence Intervals after Model Selection. *Journal of the American Statistical Association* 101, 619-629.

Leeb, H. & B. M. Pötscher (2003): The Finite-Sample Distribution of Post-Model-Selection Estimators and Uniform Versus Non-Uniform Approximations, *Econometric Theory* 19, 100-142.

Leeb, H. & B. M. Pötscher (2005): Model Selection and Inference: Facts and Fiction, *Econometric Theory* 21, 29-59.

Loh, W.-Y. (1985): A New Method for Testing Separate Families of Hypotheses, *Journal of the American Statistical Association* 80, 362-368.

Liu, C.-A. (2011): A Plug-In Averaging Estimator for Regressions with Heteroskedastic Errors, Working Paper, Version October 29, 2011.

McCloseky, A. (2011): Powerful Procedures with Correct Size for Test Statistics with Limit Distributions that are Discontinuous in Some Parameters. Working Paper, Version October 2011.

Silvapulle, M. J. (1996): A Test in the Presence of Nuisance Parameters, *Journal of the American Statistical Association* 91, 1690-1693. (Correction, ibidem 92 (1997), 801.)

A Appendix

Lemma 4 *Suppose a random variable* $\hat{c}$ *satisfies* $\Pr(\hat{c}_n \leq c^*) = 1$ *for some real number* c^* *as well as* $\Pr(\hat{c}_n < c^*) > 0$. *Let* S *be real-valued random variable. If for every non-empty interval* J *in the real line*

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

holds almost surely, then

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

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

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

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) and that the event $\{\hat{c}_n < c^*\}$ has positive probability by assumption. ∎

Lemma 5 *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 6 *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 then $\Phi^{-1}(1-v)$.

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

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., Leeb and Pötscher (2003), 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. Now we can write

$$\begin{aligned}
\Pr(T' \leq q) &= \Pr\left(\sqrt{1-\rho^2}W + \rho Z \leq q\right) - \Pr\left(|Z+\gamma| \leq c, W \leq \frac{q-\rho Z}{\sqrt{1-\rho^2}}\right) \\
&\quad + \Pr\left(|Z+\gamma| \leq c, W \leq q + \frac{\rho\gamma}{\sqrt{1-\rho^2}}\right) \\
&= \Phi(q) - \Pr(A) + \Pr(B).
\end{aligned}$$

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 only consider the case where $\rho > 0$ in the following, because the case where $\rho < 0$ is treated similarly, mutatis mutandis. We 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

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

and

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

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, it follows that $\Pr(B \backslash A) - \Pr(A \backslash B) < 0$, as required.

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

Remark 7 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 8 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.