SIZE CORRECTED POWER FOR BOOTSTRAP TESTS

Manuel A. Domínguez and Ignacio N. Lobato[1]

Instituto Tecnológico Autónomo de México

December, 2000.

Abstract

This note provides an alternative perspective for size-corrected power for a test. The advantage of this approach is that it allows the calculation of size-corrected power for bootstrap tests.

Keywords and Phrases: Size-adjusted power, Monte Carlo.

JEL classification number: C12.

[1] Corresponding author:

Ignacio N. Lobato

Address: Instituto Tecnológico Autónomo de México
Av. Camino a Santa Teresa #930
Col. Héroes de Padierna
10700 México, D.F., MEXICO
E-mail: ilobato@itam.mx
Phone:52-56284191
Fax:52-56284058

1 Introduction

This note is motivated by the following two observations:

a) When a researcher has several testing procedures available for a given problem, a common practice is to choose the one with the highest empirical power. However, this usage is difficultied because the empirical rejection probabilities (RP's) of the tests under the null hypothesis are different to the nominal level and among themselves. Hence, a widely employed practice in Econometrics is to report what is called *size-corrected power*, that is computing the empirical power with simulated critical values. See, for instance, Stock (1994), Hong and White (1995) or Dufour and Kiviet (1998).

b) Recently, there has been a big increase in the use of bootstrap tests in Econometrics. These tests employ simulated critical values (of the bootstrap estimate of the true distribution) instead of asymptotic critical values. See, for instance, Hall and Horowitz (1996), Hansen (1996) or Andrews (1997).

Given that the bootstrap test already computes simulated critical values, it does not appear to be obvious how to perform *size corrected power* for bootstrap tests. In the next Section we describe size corrected power for a typical asymptotic test and reinterpret it in an alternative form. In Section 3 we show how the alternative way can be applied to bootstrap tests. A word about notation. For any cumulative distribution function F and any $\alpha \in (0,1)$, $F^{-1}(\alpha)$ denotes the α-th quantile of F.

2 *Size corrected power* for an asymptotic test

Let $1-\alpha$ denote the nominal level (typically, $\alpha = 0.95$) at which the one-sided test is carried out. Let T_n be the relevant test statistic calculated with a sample of size n, and denote by F_{T_n} and F_{T_∞} the exact and asymptotic distribution functions of the statistic T_n under the null hypothesis, respectively. The standard asymptotic test consists in comparing the test statistic T_n against the appropriate asymptotic critical value, that is,

$$\text{reject the null hypothesis if } T_n > F_{T_\infty}^{-1}(\alpha) = q_1. \tag{1}$$

Since $F_{T_n} \neq F_{T_\infty}$ the RP of the test (1) is not exactly $1-\alpha$ under the null hypothesis but $1-\alpha^1$ where $\alpha^1 = F_{T_n}(F_{T_\infty}^{-1}(\alpha))$, (see Figure 1). Since F_{T_n} is unknown, in Monte Carlo experiments it has often been used the test

$$\text{reject the null hypothesis if } T_n > \widehat{F}_{T_n}^{-1}(\alpha) = q_2, \tag{2}$$

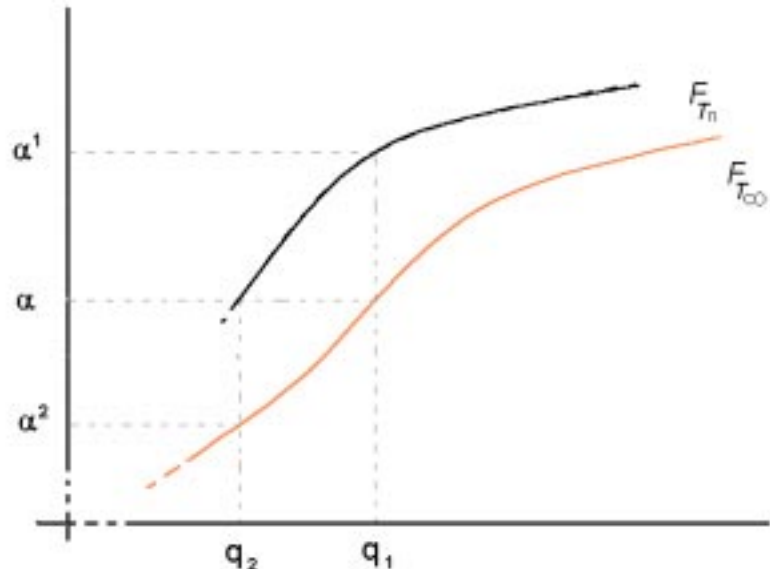

Figure 1:

where $\widehat{F}_{T_n}$ is a Monte Carlo estimate of F_{T_n}. Since $\widehat{F}_{T_n}$ can be made as accurate as necessary, in Figure 1 we have plotted F_{T_n} and $\widehat{F}_{T_n}$ identical. In fact, the same practice has been followed for all figures in this note. The power of the test (2) is called *size corrected power*. Notice that since $F_{T_n}^{-1}$ and $\widehat{F}_{T_n}^{-1}$ are case dependent, in practice the test (2) cannot be computed.

An alternative way of looking at *size corrected power* is the following. Let $\alpha^2 = F_{T_\infty}(\widehat{F}_{T_n}^{-1}(\alpha))$. In Figure 1 notice that $F_{T_\infty}^{-1}(\alpha^2) = F_{T_n}^{-1}(\alpha) = q_2$, therefore the test (2) is equivalent to

$$\text{reject the null hypothesis if } T_n > F_{T_\infty}^{-1}(\alpha^2) = q_2.$$

3 *Size corrected power* for a bootstrap test

Recently, bootstrap tests have been employed extensively in Econometrics. A bootstrap test is based on the construction of a bootstrap test statistic, T_n^*. This statistic is constructed by applying the same procedure to an artificial sample obtained by drawing observations from the original sample with replacement. In addition, the role played in the asymptotic test by F_{T_∞} is now assumed by $F_{T_n^*}$. Hence, an ideal bootstrap test would be the following:

$$\text{reject the null hypothesis if } T_n > F_{T_n^*}^{-1}(\alpha),$$

where $F_{T_n^*}$ denotes the exact distribution of the bootstrap test statistic T_n^*. However, frequently $F_{T_n^*}$ is difficult to obtain and is estimated by Monte Carlo simulations. Denote this estimate by $\widehat{F}_{T_n^*}$. Hence, the typical bootstrap test is the following:

$$\text{reject the null hypothesis if } T_n > \widehat{F}_{T_n^*}^{-1}(\alpha). \tag{3}$$

Here again we have the problem that $\widehat{F}_{T_n^*}$ is not exactly the same as F_{T_n}, so the RP of the test (3) is not exactly $1-\alpha$ but $1-\alpha^b$. How can the researcher perform *size corrected power* in this framework? As we have seen in the previous section, the idea is to continue using the $\widehat{F}_{T_n^*}$ distribution but with the critical value that corresponds to the desired level. For this purpose define (see Figure 2)

$$\alpha^c = F_{T_n^*}(F_{T_n}^{-1}(\alpha)).$$

The test would be

$$\text{reject the null hypothesis if } T_n > \widehat{F}_{T_n^*}^{-1}(\alpha^c). \tag{4}$$

However, since α^c is unknown, the test (4) can not be carried out. The solution is to estimate α^c by

$$\widehat{\alpha}^c = \widehat{F}_{T_n^*}(\widehat{F}_{T_n}^{-1}(\alpha)). \tag{5}$$

Hence, the *size corrected power* bootstrap test is

$$\text{reject the null hypothesis if } T_n > \widehat{F}_{T_n^*}^{-1}(\widehat{\alpha}^c).$$

Therefore, the practical procedure for performing size-corrected power for a bootstrap test consists in the following steps:

1) Calculate $\widehat{\alpha}^c$ using (5).

2) Generate a different data set and calculate T_n. Then, generate B bootstrap estimators of T_n. Call them $\{T_{nb}^* : b = 1, ...B\}$. Let $T_{n(d)}^*$ be the d-th order statistic of the sample $\{T_{nb}^* : b = 1, ...B\}$.

Then, reject the null if

$$T_n > T_{n([B\widehat{\alpha}^c])}^*$$

3) Repeat step 2 R times where R is the number of replications chosen by the researcher in the Monte Carlo study.

4) Finally, the size corrected power of the bootstrap test is

$$\frac{\#(T_n > T_{n([B\widehat{\alpha}^c])}^*)}{R}.$$

Remark. Note that in the case of the classical size corrected power, the critical value is fixed but in the bootstrap context it changes from one sample to another in a Monte Carlo experiment, keeping fixed the value $\widehat{\alpha}^c$.

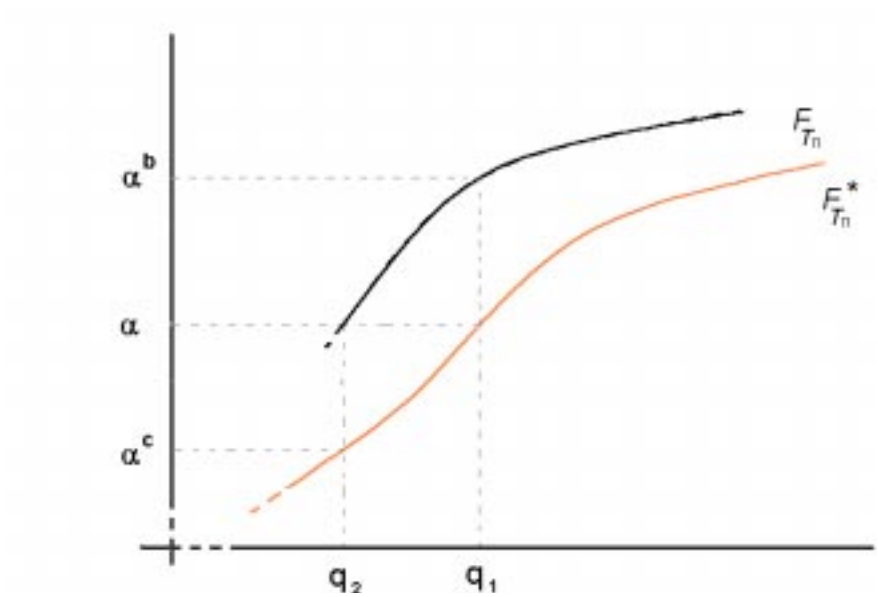

Figure 2:

4 Size-corrected testing

The exposition so far is related to carrying out a finite sample comparison among several tests. However, it is also of interest to perform *size-corrected* tests in applications. We propose to do that by iterating the bootstrap as we explain next.

Notice that in practice, the data generating process is not available and so, $\widehat{\alpha}^c$ in (5) can not be computed since $\widehat{F}_{T_n}$ is, in general, case dependent. In this situation we suggest to approximate $\widehat{\alpha}^c$ by using the iterated bootstrap. Let T_n^{**} be the bootstrap statistic applied to a resample drawn with replacement from a previous bootstrap sample of the original data set. The idea is to estimate $\widehat{\alpha}^c$ by

$$\alpha^{c*} = \widehat{F}_{T_n^{**}}(\widehat{F}^{-1}_{T_n^*}(\alpha)),$$

where $\widehat{F}_{T_n^{**}}$ denotes a Monte Carlo estimate of the exact distribution of the iterated bootstrap statistic T_n^{**}, see Figure 3.

Hence, in practice a *size-corrected* test would be

$$\text{reject the null hypothesis if } T_n > \widehat{F}^{-1}_{T_n^*}(\alpha^{c*}). \tag{6}$$

Notice that size distortion is a finite sample problem. The rationale behind (6) is that the iterated bootstrap converges faster than the one-step bootstrap for asymptotically pivotal statistics, and hence, some improvement in finite sample behavior can be expected. Also,

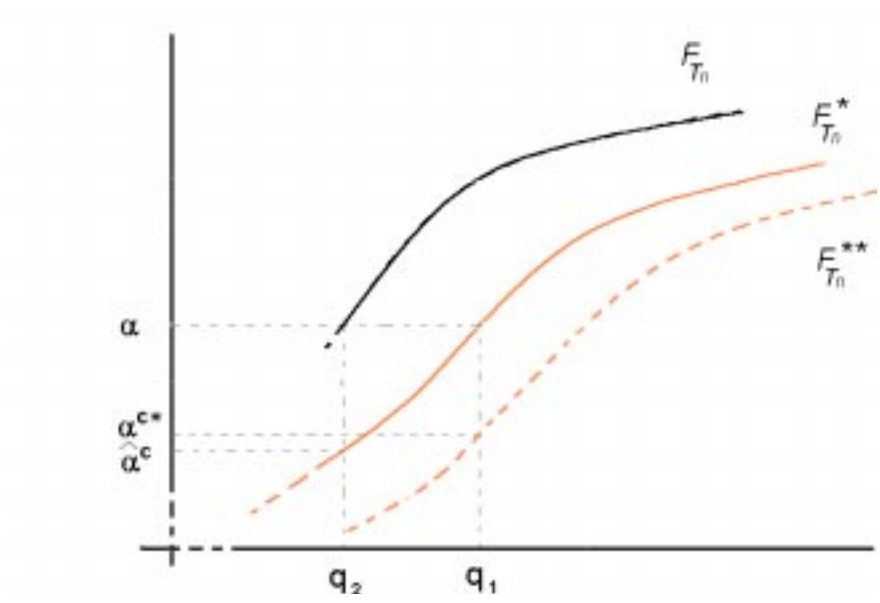

Figure 3:

note that here we are employing *size-correction* as the analogous of *bias-correction* in the bootstrap literature (see Hall (1992)).

Finally, the discussion in this note refers to one-sided test that is the most common situation found in practice. The analysis for two-sided tests is completely analogous.

References:

Andrews, D.W.K. (1997), A conditional Kolmogorov test, Econometrica, 65, 1097-1128.

Dufour, J.-M. and J.F. Kiviet (1998), Exact inference methods for first-order autoregressive distributed lag models, Econometrica, 66, 79-104.

Hall, P. (1992), The Bootstrap and Edgeworth expansion, Springer-Verlag, New York.

Hall, P. y J. Horowitz (1996), Bootstrap critical values for tests based on Generalized-Method-of-Moments Estimators, Econometrica, 64, 891-916.

Hansen, B. (1996), Inference when a nuisance parameter is not identified under the null hypothesis, Econometrica, 64, 413-430.

Hong, Y. y H. White (1995), Consistent specification testing via nonparametric series regressions, Econometrica, 63, 1133-1160.

Stock, J. (1994), Unit roots, structural breaks and trends, in : R.F. Engle and D.L. McFadden, eds., Handbook of Econometrics, Vol. IV (North-Holland, Amsterdam) 2739-2841.