Available online at www.sciencedirect.com

SCIENCE DIRECT®

Statistics & Probability Letters 74 (2005) 356–365

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Bootstrap hypothesis testing in regression models

Efstathios Paparoditis[a], Dimitris N. Politis[b,*]

[a]*Department of Mathematics and Statistics, University of Cyprus, P.O.Box 20537, CY 1678 Nicosia, Cyprus*
[b]*Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112, USA*

Received 14 May 2004; accepted 1 April 2005
Available online 27 June 2005

Abstract

The paper investigates how the particular choice of residuals used in a bootstrap-based testing procedure affects the properties of the test. The properties of the tests are investigated both under the null and under the alternative. It is shown that for non-pivotal test statistics, the method used to obtain residuals largely affects the power behavior of the tests. For instance, imposing the null hypothesis in the residual estimation step—although it does not affect the behavior of the test if the null is true—it leads to a loss of power under the alternative as compared to tests based on resampling unrestricted residuals. Residuals obtained using a parameter estimator which minimizes their variance maximizes the power of the corresponding bootstrap-based tests. In this context, studentizing makes the tests more robust to such residual effects.

Keywords: Hypothesis testing; Parametric models; Resampling; Residuals

1. Introduction

Consider data $Y_1, \ldots, Y_n$ arising from the simple linear regression

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n$$

with $\varepsilon_i \sim \text{IID}(0, \sigma_\varepsilon^2)$ and x_i some fixed design points. To construct a bootstrap test of the hypothesis $\text{H}_0 : \beta_1 = 0$ vs $\text{H}_1 : \beta_1 \neq 0$ based on the least-squares (LS) estimator $\hat{\beta}_1$ and the test statistic $\sqrt{n}\hat{\beta}_1$,

*Corresponding author.
E-mail address: politis@math.ucsd.edu (D.N. Politis).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2005.04.057

we can entertain the following two possibilities:

I (a) Impose the null hypothesis and estimate the residuals by $\hat{\varepsilon}_i = Y_i - \hat{\beta}_{0,\mathrm{H}_0}$ where $\hat{\beta}_{0,\mathrm{H}_0} = \bar{Y}$ is the restricted (by H_0) least-squares estimator of β_0; (b) generate pseudo-data $Y_1^*, \ldots, Y_n^*$ by $Y_i^* = \hat{\beta}_{0,\mathrm{H}_0} + \hat{\varepsilon}_i^*$ for $i = 1, \ldots, n$, where the $\hat{\varepsilon}_i^*$ are an i.i.d. sample from the empirical distribution of $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$; (c) (re)compute the LS statistic $\hat{\beta}_1$ on the pseudo-data to get a pseudo-replication $\hat{\beta}_1^*$; (d) repeating the above two steps B times we form a set of B pseudo-replications of our statistic from which a pseudo-empirical bootstrap null distribution of $\hat{\beta}_1$ is constructed; based on the latter, the test's critical region can be identified.

II (a) Do *not* impose the null hypothesis from the start; instead, estimate the residuals by $\tilde{\varepsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$ where $\hat{\beta}_0, \hat{\beta}_1$ are the usual (unrestricted) LS estimates; (b) generate pseudo-data $Y_1^*, \ldots, Y_n^*$ by $Y_i^* = \hat{\beta}_0 + \hat{\beta}_1 x_i + \tilde{\varepsilon}_i^*$ for $i = 1, \ldots, n$, where the $\tilde{\varepsilon}_i^*$ are an i.i.d. sample from the empirical distribution of $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$; (c) (re)compute the statistic $\hat{\beta}_1$ on the pseudo-data to get a pseudo-replication $\hat{\beta}_1^*$; (d) same as above.

Which of the above two proposals is preferable? It is easy to see that, under H_0, they are both approximately equivalent as they both give asymptotically valid estimators of the true null distribution of $\hat{\beta}_1$; as a matter of fact, method (I) may have a slight edge since it directly uses the information that $\beta_1 = 0$ whereas method (II) employs an estimator of β_1. Nevertheless, the situation is radically different when the given data do *not* satisfy H_0. In that case method (II) still 'works', i.e., manages to give asymptotically valid estimators of the true distribution of $\hat{\beta}_1$ under the null, but method (I) can behave erratically. The reason is that if H_1 is true, then the (empirical) variance of the restricted residuals $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$ is larger than that of the unrestricted residuals based on the least squares estimator which is generally consistent. This can result under the alternative to a shrinkage of the rejection region of the bootstrap test based on restricted residuals and therefore to a loss of power. Thus whereas both methods seem to asymptotically achieve the desired level of the test, method (I) is expected to have power problems as it may well fail to identify the optimal critical region for the test under H_1; thus, method (II) seems generally preferable.

The above intuitive arguments are indeed true for non-pivotal test statistics and can be rigorously proven under the more general set-up of (possibly) nonlinear regression with fixed and/or random design and with independent or dependent explanatory variables. However, the situation is different for pivotal test statistics such as studentized statistics based on $\hat{\beta}_1$. Here a robustifying effect occurs which make the corresponding test less sensitive with respect to the set of residuals used in the resampling step.

In the statistical literature, hypothesis testing based on bootstrap critical values has received less attention compared to the construction of confidence intervals or approximating the distribution of estimators. Beran (1986) discussed problems related to the power and the size of bootstrap tests. Bootstrap tests of significance have also been discussed in Hinkley (1989). Hall and Wilson (1991) and Hall (1992) provided some general guidelines for bootstrap-based hypothesis testing while Horowitz (1994) pointed out the importance of using pivotal statistics in hypothesis testing.

The use of residuals obtained by imposing the null hypothesis in bootstrap-based testing has been advocated in the literature by many authors in different contexts; see among others Li and Maddala (1996), Nankervis and Savin (1996), Park (2003), Swensen (2003). The idea underlying