

# Bootstrap hypothesis testing in regression models

Efstathios Paparoditis  
Department of Mathematics and Statistics  
University of Cyprus  
P.O.Box 20537  
CY 1678 Nicosia, CYPRUS

Dimitris N. Politis  
Department of Mathematics  
University of California, San Diego  
La Jolla, CA 92093-0112  
USA

## 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.

**Key words:** Hypothesis Testing, Parametric Models, Resampling, Residuals.

# 1 Introduction

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

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

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  $H_0 : \beta_1 = 0$  vs.  $H_1 : \beta_1 \neq 0$  based on the least-squares (LS) estimator  $\hat{\beta}_1$  and the test statistic  $\sqrt{n}\hat{\beta}_1$ , we can entertain the following two possibilities:

I: (a) Impose the null hypothesis and estimate the residuals by  $\hat{\varepsilon}_i = Y_i - \hat{\beta}_{0,H_0}$  where  $\hat{\beta}_{0,H_0} = \bar{Y}$  is the restricted (by  $H_0$ ) least-squares estimator of  $\beta_0$ ; (b) generate pseudo-data  $Y_1^*, \dots, Y_n^*$  by  $Y_i^* = \hat{\beta}_{0,H_0} + \hat{\varepsilon}_i^*$  for  $i = 1, \dots, n$ , where the  $\hat{\varepsilon}_i^*$  are an i.i.d. sample from the empirical distribution of  $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ ; (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  $Y_1^*, \dots, Y_n^*$  by  $Y_i^* = \hat{\beta}_0 + \hat{\beta}_1 x_i + \tilde{\varepsilon}_i^*$  for  $i = 1, \dots, n$ , where the  $\tilde{\varepsilon}_i^*$  are an i.i.d. sample from the empirical distribution of  $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n$ ; (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  $\beta_1$ . 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  $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$  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 (2000), Swensen (2003). The idea underlying the use of such restricted residuals, is that if the estimate of the parameter of interest is far from its true value, then the empirical distribution of the residuals will suffer from a poor approximation of the true distribution of the errors. The aim of this paper is to investigate closely and to give an analytical answer to the questions related to the effects that the residuals used in the resampling step have on the properties of bootstrap-based tests. It is shown, that for non pivotal test statistics and for a huge and commonly used class of parameter estimators, the use of restricted residuals leads to a loss in power, i.e., tests based on resampling residuals obtained by imposing the null are less powerful than tests based on resampling unrestricted residuals.

Furthermore, bootstrap tests based on sets of residuals corresponding to parameter estimators which minimize the residual variance, maximize the power of the bootstrap tests. This is asymptotically true despite which particular estimator is used to carry out the test within a large class of parameter estimators. However, for pivotal test statistics a robustifying effect occurs which is due to the fact that the characteristics of the distribution of the residuals used in the resampling step and in particular their variance do not affect the limiting distribution of the test statistic under the null. Thus using asymptotically pivotal test statistics make bootstrap tests less sensitive to the effects of the residuals used in the resampling step.

The paper is organized as follows. In the next section, the general regression framework is described and the general bootstrap testing algorithm is rigorously stated. Section 3 states the assumptions needed and establishes our main results. A brief Conclusions section is finally added.

## 2 A General Hypothesis Testing Set-up and Bootstrap Testing Algorithm

Assume that observations  $Y_1, Y_2, \dots, Y_n$  from the model

$$Y_i = G(\mathbf{X}_i, \beta) + \varepsilon_i, \tag{2.1}$$

are available, where  $\beta = (\theta, \nu) \in \Delta = \Theta \times N \subset \mathbb{R}^p \times \mathbb{R}^m$ ,  $\varepsilon_i \sim IID(0, \sigma_\varepsilon^2)$ ,  $\varepsilon_i$  and  $\mathbf{X}_j$  are independent for all  $i, j \in \{1, 2, \dots, n\}$  and  $G(\cdot)$  is some known function. The design variables

$\mathbf{X}_i = (X_{i,1}, X_{i,2}, \dots, X_{i,p+m})$ ,  $i = 1, 2, \dots, n$  may consist of fixed and/or random variables which are allowed to dependent on each other. Let  $\theta$  be the parameter of interest and consider the testing problem

$$H_0 : \mathbf{r}'\theta = \mathbf{r}'\theta_0 \quad \text{vs} \quad H_1 : \mathbf{r}'\theta < \mathbf{r}'\theta_0,$$

where  $\mathbf{r}$  is a  $p$ -dimensional vector and  $\nu$  is a set of  $m$  nuisance parameters. Other choices of the null and of the alternative hypothesis can be considered as well. Let  $\hat{\beta} = (\hat{\theta}, \hat{\nu})$  be an estimator of  $\beta = (\theta, \nu)$  satisfying certain regularity conditions to be specified later. Statistics useful for testing such a null hypothesis are commonly given by:

$$T_n = \sqrt{n}\mathbf{r}'(\hat{\theta} - \theta_0), \quad \text{or by the studentized version} \quad t_n = \sqrt{n}\mathbf{r}'(\hat{\theta} - \theta_0)/\hat{S}_\theta,$$

where  $\hat{S}_\theta$  denotes an estimator of the standard deviation of the limiting distribution of  $\sqrt{n}\mathbf{r}'(\hat{\theta} - \theta)$  to be specified later; see (3.2). Some examples that fit into this framework are the following:

**Example 1:** (Linear regression model). Let  $\mathbf{X}_i$  be a set of  $(p + m)$  explanatory variables and  $Y_i$  a response variable following model (2.1). For  $G(\mathbf{X}_i, \beta) = \mathbf{X}_i'\beta$  we are in the context of the classical linear regression model. Testing hypothesis of the form  $H_0 : \mathbf{r}'\theta = 0$  is a very common problem in statistical applications.

**Example 2:** (Regression with dependent variables). Let  $\{X_t, t \in \mathbf{Z}\}$  be a stationary stochastic process and  $\{\varepsilon_t, t \in \mathbf{Z}\}$  a sequence of independent, identically distributed random variables with mean zero and variance  $\sigma_\varepsilon^2$ . Let  $Y_i$  be defined by (2.1) and  $\mathbf{X}_t = (X_t, X_{t-1}, \dots, X_{t-p+1})$ . For  $G(\mathbf{X}_t, \beta) = \mathbf{X}_t'\beta$  and  $\beta = (\beta_1, \dots, \beta_p)'$  we are in the context of model (2.1) involving dependent random variables. Several hypothesis regarding the parameters of this model can be tested.

Performing the test based on the statistics  $T_n$  or  $t_n$  requires knowledge of their distribution under the null hypothesis, i.e., under the hypothesis that  $\mathbf{r}'\theta = \mathbf{r}'\theta_0$ . One way to estimate this distribution is by using the bootstrap. Typically, such a bootstrap algorithm generates pseudo observations  $Y_1^*, Y_2^*, \dots, Y_n^*$  using the parametric model under the null and resampling from estimated residuals. Resampling residuals imitates the randomness in (2.1) due to the true errors  $\varepsilon_i$ . One issue raised in this context is which set of residuals should be used for such a purpose. For instance, should the residuals used be estimated by imposing the null hypothesis or not. Putting it in a general context, the following residual-based bootstrap algorithm can be used to estimate the distribution of the test statistics  $T_n$  or  $t_n$  under the null hypothesis.

**Step 1:** Let  $\{\tilde{\beta}_n\}_{n \in \mathbf{N}} = \{(\tilde{\theta}_n, \tilde{\nu}_n)\}_{n \in \mathbf{N}} \in \Delta$  be a sequence of parameters. Usually,  $\tilde{\beta}_n$  is a sequence of estimators of  $\beta$  but this is not necessary. Define residuals

$$\tilde{\varepsilon}_i = Y_i - G(\mathbf{X}_i, \tilde{\beta}_n), \quad i = 1, 2, \dots, n$$

and let  $\tilde{F}_n$  be the empirical distribution function of the centered  $\tilde{\varepsilon}_i$ 's, i.e.,

$$\tilde{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, x]}(\tilde{\varepsilon}_i - \bar{e}), \quad \text{where} \quad \bar{e} = \frac{1}{n} \sum_{i=1}^n \tilde{\varepsilon}_i.$$

**Step 2:** Generate pseudo-observations

$$Y_i^* = G(\mathbf{X}_i^*, \tilde{\beta}_0) + \varepsilon_i^*,$$

where  $\varepsilon_i^* \sim \tilde{F}_n$  are independent identically distributed,  $\tilde{\beta}_0 = (\theta_0, \tilde{\nu}_n)$  and  $\tilde{\nu}_n$  as in Step 1.

**Step 3:** Let  $\hat{\beta}^* = (\hat{\theta}^*, \hat{\nu}^*)$  be the estimator of  $\tilde{\beta}_0$  obtained using the same estimation method as the one used to obtain  $\hat{\beta}$  and define the bootstrap analogue of  $T_n$  by

$$T_n^* = \sqrt{n} \mathbf{r}' (\hat{\theta}^* - \theta_0),$$

and of  $t_n$  by

$$t_n^* = \sqrt{n} \mathbf{r}' (\hat{\theta}^* - \theta_0) / \hat{S}_\theta^*$$

where  $\hat{S}_\theta^*$  denotes the bootstrap analogue of  $\hat{S}_\theta$ .

**Step 4:** Repeat Step1 to Step 3 a large number (say  $B$ ) of times, and let  $C_{\alpha, B}^*(\tilde{\beta}_n, T_n)$  be the lower empirical  $\alpha$ -percentage point of the distribution of  $T_n^*$  and by  $C_{\alpha, B}^*(\tilde{\beta}_n, t_n)$  the corresponding percentage point of the distribution of  $t_n^*$ . The bootstrap test based on  $T_n$  rejects the null hypothesis if  $T_n \leq C_{\alpha, B}^*(\tilde{\beta}_n, T_n)$  while that based on  $t_n$  rejects if  $t_n \leq C_{\alpha, B}^*(\tilde{\beta}_n, t_n)$ .

In the above algorithm we do not elaborate on the particular procedure used to generate the bootstrap design variables  $\mathbf{X}_i^*$  because this procedure is case dependent. For instance, if  $\mathbf{X}_i = \mathbf{x}_i$ , i.e., if the  $\mathbf{X}_i$  are fixed design points, then  $\mathbf{X}_i^* = \mathbf{x}_i$  is an obvious choice. If the  $\mathbf{X}_i$ 's are random variables then so should be the  $\mathbf{X}_i^*$  imitating correctly the second order structure of the observed  $\mathbf{X}_i$ 's. For our purposes, the only requirement on the bootstrap design variables  $\mathbf{X}_i^*$  needed are stated in assumption (B2) in the next section.

Notice that in the above algorithm two different notations for  $\beta$  are applied:  $\tilde{\beta}_n$  refers to the sequence of parameters used to obtain the residuals  $\tilde{\varepsilon}_i$  while  $\hat{\beta} = (\hat{\theta}, \hat{\nu})$  refers to the sequence of estimators used to perform the test about  $\theta$ . These two sequences need not be identical. Moreover, different choices of  $\{\tilde{\beta}_n\}$  in Step 1 lead to different set of 'residuals' and consequently to different bootstrap critical values. For instance, for  $\tilde{\beta}_n = (\theta_0, \tilde{\nu}_n)$  we are in the case where the residuals are obtained by imposing the null hypothesis. For  $\tilde{\beta}_n = \hat{\beta} = (\hat{\theta}, \hat{\nu})$  we are in the case where the residuals are obtained using the (unrestricted) estimator of  $\beta$ . Our formulation of the bootstrap algorithm allows for other choices of  $\tilde{\beta}_n$  as well leading to different resampling sets of residuals.

### 3 Assumptions and Main Results

In order to investigate to what extent different choices of  $\tilde{\beta}_n$  affect the properties of the bootstrap tests, we state some assumptions needed in the sequel. The first assumption imposes an

elementary smoothness condition on the regression function  $G(\cdot)$ .

**(A1)**  $|G(\mathbf{x}, \beta_1) - G(\mathbf{x}, \beta_2)| \leq L(\mathbf{x})\|\beta_1 - \beta_2\|$ , where  $\sup_{\mathbf{x}} L(\mathbf{x}) < C$  and  $\|\cdot\|$  denotes  $L_2$  norm.

The second assumption deals with the asymptotic behavior of the sequence  $\tilde{\beta}_n$  used to obtain the residuals in the first step of the bootstrap algorithm.

**(A2)** There exists some  $\tilde{\beta} \in \Delta$  such that  $\tilde{\beta}_n \rightarrow \tilde{\beta}$  in some appropriate sense as  $n \rightarrow \infty$ .

Assumption (A2) is satisfied if  $\tilde{\beta}_n$  is any weak or strong consistent estimator of  $\beta$ . In this case  $\tilde{\beta} = \beta$ . (A2) is however also satisfied if  $\tilde{\beta}_n = (\theta_0, \tilde{\nu})$  or if  $\tilde{\beta}_n$  is some deterministic sequence in  $\Delta$  converging to some  $\tilde{\beta} \in \Delta$ .

We assume that the sequence  $\hat{\beta}$  of estimators used to perform the test satisfies

**(A3)**  $\left\| \sqrt{n}(\hat{\beta} - \beta) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\mathbf{X}_i)(Y_i - G(\mathbf{X}_i, \beta)) \right\| \rightarrow 0$  in probability, where  $\psi$  is a measurable function.

Assumption (A3) is satisfied by several choices of the sequence of estimators  $\hat{\beta}$ . For instance, the least squares estimator in the linear regression model of Example 1, is given by  $\hat{\beta}_{LS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ , where  $\mathbf{X} = (X_{i,j})_{i=1,2,\dots,n}^{j=1,2,\dots,p+m}$  and  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)'$ . It is easily seen that in this case (A3) is satisfied with  $\psi(\mathbf{X}_i) = [(n^{-1} \sum_{r=1}^n X_{l,r} X_{r,s})^{-1} X_{i,k}]_{k=1,2,\dots,p+m}$ . The fact that the deviation  $(Y_i - G(\mathbf{X}_i, \beta))$  appears in the approximating sequence of  $\sqrt{n}(\hat{\beta} - \beta)$  displayed in assumption (A3) above implicitly restricts our focus to particular classes of estimators  $\hat{\beta}$  of  $\beta$ . They are essentially characterized by the property that they minimize an (asymptotic)  $L_2$ -type distance between  $Y_i$  and its predictor under the regression model. Based on this approximation it is shown below that the second moment of  $(Y_i - G(\mathbf{X}_i, \beta))$  enters the asymptotic distribution of  $\sqrt{n}(\hat{\beta} - \beta)$  as a multiplicative factor of the variance of the limiting distribution of this statistic. This is an important point because it implies that the distribution of the errors  $Y_i - G(\mathbf{X}_i, \beta)$  affects the asymptotic distribution of the statistic  $\sqrt{n}(\hat{\beta} - \beta)$  essentially through the variance of the errors  $Y_i - G(\mathbf{X}_i, \beta)$ .

In the following we assume that the  $\mathbf{X}_i$ 's are random variables. If the  $\mathbf{X}_i$ 's are fixed design points then all results remain valid with obvious modifications. The next assumption requires that the approximating sequence in (A3) obeys a central limit theorem.

**(A4)** For triangular arrays  $\{e_{i,n}, i = 1, 2, \dots, n\}$  of independent and identically distributed random variables with mean zero,  $E(e_{i,n}^2) = \tau_n^2$ ,  $\lim_{n \rightarrow \infty} \tau_{i,n}^2 = \tau^2 > 0$  and  $E(e_{i,n} \mathbf{X}_i) = 0$ , we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\mathbf{X}_i) e_{i,n} \Rightarrow N\left(0, \tau^2 \text{Var}[\psi(\mathbf{X}_i)]\right),$$

as  $n \rightarrow \infty$ , where  $\text{Var}[\psi(\mathbf{X}_i)]$  is a positive definite  $(p+m) \times (p+m)$  matrix.

Although  $\varepsilon_i = (Y_i - G(\mathbf{X}_i, \beta))$  affects the asymptotic distribution of  $\sqrt{n}(\hat{\beta} - \beta)$  by its second order moment  $E(Y_i - G(\mathbf{X}_i, \beta))^2$ , it could be that the variance of the limiting Gaussian

distribution does not depend on  $E(Y_i - G(\mathbf{X}_i, \beta))^2$ . For instance, if  $Y_i = X_t$  is the  $p$ -th order causal autoregressive process  $X_t = \sum_{i=1}^p \beta_i X_{t-i} + \varepsilon_t$  and  $\hat{\beta}$  the least squares estimator of  $\beta$ , then  $\sqrt{n}(\hat{\beta} - \beta) \rightarrow N(0, \Sigma)$  with  $\Sigma = \sigma_\varepsilon^2 \Gamma_p^{-1}$ ,  $\Gamma_p = (\gamma(i-j))_{i,j=1,2,\dots,p}$  and  $\gamma(h) = E(X_t X_{t+h})$  the autocovariance at lag  $h$  of  $X_t$ . Note that because  $\gamma(h) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}$  with  $\psi_j$  the coefficients of the power series  $(1 - \sum_{j=1}^p \beta_j z^j)^{-1} = \sum_{j=0}^{\infty} \psi_j z^j$ , the asymptotic covariance matrix  $\Sigma$  does not depend in this case on the variance  $\sigma_\varepsilon^2$  of the errors  $\varepsilon_t$ .

To avoid inconsistency of the bootstrap in approximating correctly the distribution of  $\sqrt{n}(\hat{\beta} - \beta)$ , we require that an analogous condition to (A3) is fulfilled for the bootstrap estimator  $\hat{\beta}^*$  as well. More precisely we assume that

$$(B1) \quad \left\| \sqrt{n}(\hat{\beta}^* - \tilde{\beta}_0) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\mathbf{X}_i^*)(Y_i^* - G(\mathbf{X}_i^*, \tilde{\beta}_0)) \right\| \rightarrow 0, \text{ in probability.}$$

Our final assumption deals with the properties of the bootstrap quantities  $\mathbf{X}_i^*$  used to mimic the observed  $\mathbf{X}_i$ . More specifically we assume that

$$(B2) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^n E[\psi(\mathbf{X}_i) - \psi(\mathbf{X}_i^*)][\psi(\mathbf{X}_j) - \psi(\mathbf{X}_j^*)]' \rightarrow 0 \text{ in probability, as } n \rightarrow \infty.$$

Assumptions (A3) and (A4) imply the asymptotic normality of the estimator  $\sqrt{n}(\hat{\beta} - \beta)$ , i.e.,

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow N(0, \sigma_\varepsilon^2 \text{Var}[\psi(\mathbf{X}_i)]).$$

We define  $\hat{S}_\theta = \sqrt{\hat{S}_\theta^2}$  where

$$\hat{S}_\theta^2 = \hat{\sigma}_\varepsilon^2 R' \text{Var}[\psi(\mathbf{X}_i)] R, \quad (3.2)$$

$\hat{\sigma}_\varepsilon^2 = n^{-1} \sum_{i=1}^n (Y_i - G(\mathbf{X}_i, \hat{\beta}))^2$ ,  $R = (r', 0'_m)$  and  $0'_m = (0, 0, \dots, 0)$  the  $m$  dimensional zero vector.

By (B1), (B2) and (A4) we have

$$\begin{aligned} \sqrt{n}(\hat{\beta}^* - \tilde{\beta}_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\mathbf{X}_i^*)(Y_i - G(\mathbf{X}_i^*, \tilde{\beta}_0)) + o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\mathbf{X}_i^*) \varepsilon_i^* + o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\mathbf{X}_i) \varepsilon_i^* + o_P(1) \\ &\rightarrow N(0, \tilde{\tau}^2 \text{Var}[\psi(\mathbf{X}_i)]), \end{aligned}$$

as  $n \rightarrow \infty$ . Here

$$\tilde{\tau}^2 = E(Y_i - G(\mathbf{X}_i, \tilde{\beta}))^2$$

denotes the limit of  $\text{Var}^*(\varepsilon_i^*)$  as  $n \rightarrow \infty$ . Notice that using (A1) we get

$$\text{Var}^*(\varepsilon_i^*) = \frac{1}{n} \sum_{i=1}^n (Y_i - G(\mathbf{X}_i, \tilde{\beta}_n))^2 + o_P(1)$$

$$\begin{aligned}
&= \frac{1}{n} \sum_{i=1}^n (Y_i - G(\mathbf{X}_i, \tilde{\beta}))^2 + O_P(\|\tilde{\beta}_n - \tilde{\beta}\|) + o_P(1) \\
&\rightarrow \tilde{\tau}^2 = E(Y_i - G(\mathbf{X}_i, \tilde{\beta}))^2
\end{aligned}$$

in probability as  $n \rightarrow \infty$  by (A2) and the weak law of large numbers.

A comment concerning assumption (B2) is in order. This assumption is obviously satisfied if we set  $\mathbf{X}_i^* = \mathbf{X}_i$  which is a correct choice if the  $\mathbf{X}_i$ 's are fixed design points. However, as the above discussion shows, such a simple choice of the bootstrap design variables  $\mathbf{X}_i^*$  can be also appropriate for the purpose of approximating correctly the desired distribution of  $\sqrt{n}(\hat{\beta} - \beta)$  even if the  $\mathbf{X}_i$  are random variables. In this case the bootstrap variables  $Y_i^*$  are simple generated by  $Y_i^* = G(\mathbf{X}_i, \tilde{\beta}_0) + \varepsilon_i^*$ .

Let us now return to the problem of testing the null hypothesis  $H_0$  and consider first the test statistic  $T_n$ . Let  $\pi_{T_n}^*(\theta; \tilde{\beta}_n)$  be the power function of the bootstrap test based on a resampling scheme using the residuals  $\tilde{\varepsilon}_i = Y_i - G(\mathbf{X}_i, \tilde{\beta}_n)$ , i.e.,

$$\pi_{T_n}^*(\theta; \tilde{\beta}_n) = P(T_n \leq C_{\alpha, B}^*(\tilde{\beta}_n; T_n) \mid \theta \text{ is the true parameter}) \equiv P_\theta(T_n \leq C_{\alpha, B}^*(\tilde{\beta}_n; T_n)),$$

and recall that  $\tilde{\beta} \in \Delta$  denotes the limit of  $\tilde{\beta}_n$ ; see (A2). The following theorem is one of our main results.

**Theorem 3.1** *Suppose that (A1)-(A4) and (B1)-(B2) are true. Suppose further that the distribution of  $\sqrt{n}r'(\hat{\theta} - \theta)$  is absolutely continuous with respect to Lebesgue measure on  $\mathbb{R}$  and has a bounded density.*

(i) *If  $\tilde{\beta} = (\theta_0, \nu)$  then*

$$\lim_{n \rightarrow \infty} \pi_{T_n}^*(\theta_0, \tilde{\beta}_n) = \alpha, \quad \text{in probability.}$$

(ii) *For every  $\tilde{\beta} \in \Delta$  and every  $\theta \in \Theta \setminus \{\theta_0\}$ ,*

$$\lim_{n \rightarrow \infty} \pi_{T_n}^*(\theta, \tilde{\beta}_n) = 1, \quad \text{in probability.}$$

(iii) *If  $\tilde{\beta}_n \rightarrow \tilde{\beta} \in \Delta \setminus \{\beta\}$  and  $\alpha < 0.5$ , then, on an event whose probability tends to one as  $n \rightarrow \infty$ , we have:*

$$\pi_{T_n}^*(\theta, \tilde{\beta}_n) \leq \pi_{T_n}^*(\theta, \beta_n^+) \quad \text{for every } \theta \in \Theta \setminus \{\theta_0\}.$$

*Here  $\beta_n^+ \in \Delta$  denotes any parameter sequence satisfying  $\beta_n^+ \rightarrow \beta$  in probability as  $n \rightarrow \infty$ .*

**Proof:** Denote by

$$\omega^2 = \text{Var}[R' \psi(\mathbf{X}_i)] = R' \text{Var}[\psi(\mathbf{X}_i)]R.$$

We verify first that

$$\lim_{n \rightarrow \infty} \pi_{T_n}^*(\theta, \tilde{\beta}_n) = \begin{cases} \lim_{n \rightarrow \infty} P_\theta(\sqrt{n}r'(\hat{\theta} - \theta_0) \leq C_{\alpha, B}^*(\tilde{\beta}_n; T_n)) & \text{if } H_0 \text{ true} \\ \lim_{n \rightarrow \infty} P_\theta(\sqrt{n}r'(\hat{\theta} - \theta) \leq -\sqrt{n}r'(\theta - \theta_0) + C_{\alpha, B}^*(\tilde{\beta}_n; T_n)) & \text{if } H_1 \text{ true.} \end{cases}$$

Since under  $H_0$ ,  $\sqrt{nr}'(\hat{\theta} - \theta_0) \rightarrow N(0, \tau^2 \omega^2)$ , to establish assertions (i) we have to show that  $C_{\alpha, B}^*(\tilde{\beta}_n; T_n) \rightarrow z_\alpha \sigma \omega$  in probability as  $n$  and  $B$  goes to infinity. Now, if  $H_0$  is true and  $\tilde{\beta}_n \rightarrow (\theta_0, \nu)$  then  $Var(\varepsilon_i^*) \rightarrow \sigma_\varepsilon^2 = E(Y_i - G(\mathbf{X}_i, \beta))^2$ . This implies that  $\sqrt{nr}'(\hat{\beta}^* - \tilde{\beta}_0) \rightarrow N(0, \sigma_\varepsilon^2 \omega^2)$  in probability, from which assertion (i) follows. Assertion (ii) is established if we show that under  $H_1$ ,  $C_{\alpha, B}^*(\tilde{\beta}_n; T_n) \rightarrow c < \infty$  in probability as  $n$  and  $B$  goes to infinity. Actually to show assertion (ii) it suffices to show that  $C_{\alpha, B}^*(\tilde{\beta}_n; T_n) = O_P(n^{1/2-\delta})$  for some  $\delta > 0$ . Note that if  $\tilde{\beta}_n \rightarrow \tilde{\beta}$  then  $\sqrt{nr}'(\hat{\beta}^* - \tilde{\beta}_0) \rightarrow N(0, \tilde{\tau}^2 \omega^2)$  in probability, where  $\tilde{\tau}^2 = E(Y_i - G(\mathbf{X}_i, \tilde{\beta}))^2$ . Thus  $C_{\alpha, B}(\tilde{\beta}_n; T_n) \rightarrow z_\alpha \tilde{\tau} \omega < \infty$  in probability, as  $n, B \rightarrow \infty$ . This implies that  $\lim_{n \rightarrow \infty} \pi_{T_n}^*(\theta, \tilde{\beta}_n) = 1$  for every  $\theta \in \Theta \setminus \{\theta_0\}$ .

Consider next assertion (iii) and notice that

$$\begin{aligned} \frac{\pi_{T_n}^*(\theta, \beta_n^+)}{\pi_{T_n}^*(\theta, \tilde{\beta}_n)} &= 1 + \frac{\pi_{T_n}^*(\theta, \beta_n^+) - \pi_{T_n}^*(\theta, \tilde{\beta}_n)}{\pi_{T_n}^*(\theta, \tilde{\beta}_n)} \\ &= 1 + \frac{W_n(-\sqrt{nr}'(\theta - \theta_0) + C_{\alpha, B}(\beta_n^+; T_n)) - W_n(-\sqrt{nr}'(\theta - \theta_0) + C_{\alpha, B}(\tilde{\beta}_n; T_n))}{W_n(-\sqrt{nr}'(\theta - \theta_0) + C_{\alpha, B}(\tilde{\beta}_n; T_n))}, \end{aligned}$$

where  $W_n(\cdot) = P(\sqrt{nr}'(\hat{\theta} - \theta) \leq \cdot)$  is the distribution function of  $\sqrt{nr}'(\hat{\theta} - \theta)$ . For the residuals  $\varepsilon_i^+ = Y_i - G(\mathbf{X}_i, \beta_n^+)$  obtained using  $\beta_n^+$ , we have because  $\beta_n^+ \rightarrow \beta$  in probability, that  $Var(\varepsilon_i^+) \rightarrow \sigma_\varepsilon^2$  in probability as  $n \rightarrow \infty$ . Now, for  $w_n(x) = W_n'(x)$  the density of  $\sqrt{nr}'(\hat{\theta} - \theta)$ , we get by the mean value theorem and the fact that  $C_{\alpha, B}(\beta_n^+; T_n) \rightarrow z_\alpha \sigma_\varepsilon \omega$  and  $C_{\alpha, B}(\beta_n^+; T_n) \rightarrow z_\alpha \tilde{\tau} \omega$  in probability, as  $n, B \rightarrow \infty$ , that

$$\pi_{T_n}^*(\theta, \beta_n^+) = \pi_{T_n}^*(\theta, \tilde{\beta}_n) \left( 1 + (\sigma_\varepsilon - \tilde{\tau}) K_n + o_P(1) \right),$$

where  $K_n = z_\alpha \omega w_n(\xi_n)$  and  $\xi_n$  is some value between  $-\sqrt{nr}'(\theta - \theta_0) + C_{\alpha, B}(\beta_n^+; T_n)$  and  $-\sqrt{nr}'(\theta - \theta_0) + C_{\alpha, B}(\tilde{\beta}_n; T_n)$ . Now  $w_n(\xi_n) > 0$  and  $\tilde{\tau}^2 \equiv E(Y_i - G(\mathbf{X}_i, \tilde{\beta}))^2 > E(Y_i - G(\mathbf{X}_i, \beta))^2 \equiv \sigma_\varepsilon^2 = \inf_{\delta \in \Delta} E(Y_i - G(\mathbf{X}_i, \delta))^2$  because  $\beta \neq \tilde{\beta}$ . Taking into account that  $z_\alpha < 0$  we get  $K_n \leq 0$ . This implies

$$(\sigma_\varepsilon - \tilde{\tau}) K_n \geq 0$$

which concludes the proof of assertion (iii).  $\square$

**Remark 1:** By assertion (i) of the theorem any bootstrap test achieves (asymptotically) the desired level  $\alpha$  provided the parameter sequence  $\tilde{\beta}_n$  used in the first step of the bootstrap algorithm converges to  $(\theta_0, \nu)$ . This is for instance true, if  $\tilde{\beta}_n = (\theta_0, \hat{\nu})$ , i.e., if the null hypothesis is imposed in order to obtain the residuals  $\tilde{\varepsilon}_i$ . Note that  $\hat{\nu}$  is a consistent estimator of the nuisance parameter  $\nu$ .

**Remark 2:** Assertions (ii) and (iii) deal with the behavior of the bootstrap test under the alternative. According to assertion (ii) if  $H_0$  is false, then whatever the limit  $\tilde{\beta}$  of the sequence  $\{\tilde{\beta}_n\}$  is, the corresponding bootstrap test is consistent, i.e., its power approaches unity as  $n \rightarrow \infty$ . However, by assertion (iii) and if the limit  $\tilde{\beta}$  of  $\tilde{\beta}_n$  under the alternative is different from the true

parameter vector  $\beta$ , then the corresponding test has less power than a test based on resampling residuals obtained by a parameter estimator converging to the true value  $\beta$  under the alternative. In fact, in order for a bootstrap procedure to achieve the desired level *and* to have good power properties, it should be able to mimic correctly the distribution of the test statistic used under the null even if the null hypothesis is false. A bootstrap approach based on resampling residuals obtained using an estimator  $\tilde{\beta}_n$  such that  $\tilde{\beta}_n \rightarrow \tilde{\beta} \neq \beta$  under the alternative, does not succeed in this respect.

As the proof of Theorem 3.1 suggests, in practice we have good chances to optimize the power of the bootstrap test based on  $T_n$  if the estimator  $\tilde{\beta}_n$  used to obtain the residuals  $\tilde{\varepsilon}_i$  is chosen so that the variance of the estimated residuals  $\tilde{\varepsilon}_i$  becomes minimal. This can be for instance achieved if we set  $\tilde{\beta}_n = \hat{\beta}_{LS}$  with  $\hat{\beta}_{LS}$  the least squares estimator of  $\beta$ . We state this result in the following corollary.

**Corollary 3.1** *Let  $\tilde{\beta}_n = \hat{\beta}_{LS}$  in Step 1 of the bootstrap testing algorithm. Then for all  $\theta \in \Theta \setminus \{\theta_0\}$ ,*

$$\lim_{n \rightarrow \infty} \frac{\pi_{T_n}^*(\theta; \hat{\beta}_{LS})}{\sup_{\beta^+ \in \Delta} \pi_{T_n}^*(\theta; \beta^+)} = 1,$$

*in probability. As before,  $\beta_n^+ \in \Delta$  denotes any parameter sequence satisfying  $\beta_n^+ \rightarrow \beta$  in probability as  $n \rightarrow \infty$ .*

We conclude this section by a discussion of the effects that the different residuals have on the properties of the studentized test  $t_n$ . Let  $\pi_{t_n}^*(\theta; \tilde{\beta}_n)$  be the power function of the test based on the studentized statistic  $t_n$ , i.e.,

$$\pi_{t_n}^*(\theta; \tilde{\beta}_n) = P(t_n \leq C_{\alpha, B}^*(\tilde{\beta}_n; t_n) | \theta \text{ is the true parameter}) \equiv P_\theta(t_n \leq C_{\alpha, B}^*(\tilde{\beta}_n; t_n)).$$

**Theorem 3.2** *Suppose that (A1)-(A4) and (B1)-(B2) are satisfied. For every sequence  $\{\tilde{\beta}_n; n \in \mathbb{N}\}$  the following is true:*

$$\lim_{n \rightarrow \infty} \pi_{t_n}^*(\theta; \tilde{\beta}_n) = \begin{cases} \alpha & \text{if } \theta = \theta_0 \\ 1 & \text{if } \theta \neq \theta_0, \end{cases}$$

*in probability.*

**Proof:** Since under  $H_0$ ,  $t_n = \sqrt{nr}'(\hat{\theta} - \theta_0)/\hat{S}_\theta \rightarrow N(0, 1)$  in distribution, it suffices to show that  $C_{\alpha, B}^*(\tilde{\beta}_n; t_n) \rightarrow z_\alpha$  in probability as  $n, B \rightarrow \infty$ . Now,

$$\sqrt{nr}'(\hat{\theta}^* - \theta_0) = \sqrt{n}R'(\hat{\beta}^* - \tilde{\beta}_0) \rightarrow N(0, \tilde{\tau}^2 R' \text{Var}[\psi(\mathbf{X}_i)]R),$$

in probability as  $n \rightarrow \infty$ . Furthermore, since  $\hat{\sigma}^{*2} = n^{-1} \sum_{i=1}^n (\varepsilon_i^*)^2 \rightarrow \tilde{\tau}^2$ , we get that

$$\hat{S}_\theta^* = \hat{\sigma}^{*2} R' \text{Var}[\psi(\mathbf{X}_i^*)]R \rightarrow \tilde{\tau}^2 R' \text{Var}[\psi(\mathbf{X}_i)]R,$$

in probability, from which the desired result follows.

**Remark 3:** Note that in contrast to the non-studentized statistic  $T_n^*$ , the bootstrap studentized statistic  $t_n^*$  converges to the same Gaussian limit regardless of the choice of residuals used in the resampling step. This implies that the bootstrap test based on  $t_n^*$  is (at least asymptotically) less sensitive with respect to the set of residuals used. Furthermore, as the theorem shows, apart from being consistent, the corresponding bootstrap test achieves asymptotically the correct level  $\alpha$  regardless to whether the limit  $\tilde{\beta}$  of the parameter sequence  $\tilde{\beta}_n$  is identical to the true value  $\beta$ . Notice that this is not true in the non-studentized case.

## 4 Conclusions

In this paper the problem of how to choose the residuals in a bootstrap based testing procedure has been investigated. It has been shown that for non-pivotal test statistics, estimating these residuals by imposing the null hypothesis, although asymptotically valid if the null hypothesis is correct, it leads to a loss of power if this hypothesis is false. For a bootstrap-based testing procedure to be optimal in the sense of retaining the desired level and maximizing power, it should be based on residuals which are obtained using a sequence of parameter estimators that converge to the true parameter value both, under the null *and* under the alternative. The situation is different if a studentized test statistic is considered. In this case the tests considered achieve the desired level and are consistent against any alternative independent of which set of residuals is used in the resampling step. The reason for this is that the limiting distribution of studentized statistics under the null does not depend on characteristics of the distribution of the errors. Although it seems preferable in this respect to use a studentized test statistic, this may not be true if one compares both classes of tests in term of power. Some limited simulation results performed by the authors confirm that non-studentized tests are more powerful than their studentized counterparts.

## References

- [1] Beran, R. (1986), Simulated Power Functions. *The Annals of Statistics*, 14, 151-173.
- [2] Hall, P. (1992) *The Bootstrap and Edgeworth Expansions*. Springer: Berlin.
- [3] Hall, P. and Wilson, S. R. (1991), Two Guidelines for Bootstrap Hypothesis Testing. *Biometrics*, 47, 757-762.
- [4] Hinkley, D. V. (1989), Bootstrap Significance Tests. Proceedings of the 47th Session of the International Statistical Institute, Vol. 3, pp. 65-74.
- [5] Horowitz, J. L. (1994), Bootstrap-based Critical Values for the Information-matrix test. *Journal of Econometrics*, 61, 395-411.

- [6] Li, H. and Maddala, G. S. (1996), Bootstrapping Time Series Models. *Econometric Reviews*, 15, 115-158.
- [7] Nankervis, J. C. and N. E. Savin (1996), The Level and Power of the Bootstrap  $t$  Test in the AR(1) Model with Trend. *Journal of Business and Economic Statistics*, 14, 161-168.
- [8] Park, Y. J. (2000), Bootstrap Unit Root Tests. *Econometrica* (to appear).
- [9] Swensen, A. R. (2003), Bootstrapping Unit Root Tests for Integrated Processes. *Journal of Time Series Analysis*, 24, 99-126.