Working Paper 93-05
Statistics and Econometrics Series 05
February 1993

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

BOOTSTRAPPING THE GENERAL LINEAR HYPOTHESIS TEST

Pedro Delicado and Manuel del Río*

Abstract

We discuss the use of bootstrap methodology in hypothesis testing, focusing on the classical F-test for linear hypotheses in the linear model. A modification of the F-statistics which allows for resampling under the null hypothesis is proposed. This approach is specifically considered in the one-way analysis of variance model. A simulation study illustrating the behaviour of our proposal is presented.

Key Words

Bootstrap, F-test; General Linear Hypothesis; Hypothesis Testing; Linear Model; One-way Model; Resampling.

*Delicado, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid; Del Río, Departamento de Estadística e Investigación Operativa, Universidad Complutense de Madrid. Correspondence to: Manuel del Río, Departamento Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad Complutense de Madrid, 28040 Madrid, Spain.

1 Introduction.

The use of bootstrap methodology in hypothesis testing —essentially based on approximating the critical values by bootstrap distributions— has received less attention that its application in other problems like, for instance, the construction of confidence regions. Here are some recent references. Beran (1988) studies the asymptotic error in level of bootstrap tests and the improvement given by prepivoting the test statistic. In Hinkley (1988) some general ideas about bootstrap testing are briefly discussed. Romano (1988) uses the bootstrap to aproximate critical values of nonparametric tests based on measures defined on the empirical distribution. Hall and Wilson (1991) and Hall (1992) insist in two guidelines that we analyze in Section 2: usage of pivotal functions and resampling reflecting the null hypothesis. In Mammen (1992) the convergence of bootstrapped F-test in linear models is proved.

This paper is concerned with the use of bootstrap idea in hypothesis testing. Section 3 is dedicated to present our work for testing a general linear hypothesis in a linear model. Following the second guideline cited above, we propose a natural modification of the classical F-statistics that permits resampling under the null hypothesis, though the data fail to comply with it. In Section 4, this proposal —and the basic idea of resampling taking into account the model in consideration— is analyzed with some detail in the framework of the one-way analysis of variance. In Section 5, the corresponding behaviour is illustrated by reporting the results of a simulation study.

2 Two general guidelines.

Naturally, the duality between hypothesis testing and confidence regions is maintained under bootstrap methodology. That leads to consider the first guideline cited: use of (asymptotically) pivotal quantities, that could be extended to the use of methods with good behaviour in the problem of confidence interval construction. In general, acting in this way will imply an improvement in the test level.

However, there is an important conceptual difference when both problems are considered under the bootstrap approach. In hypothesis testing accurate estimates of the critical values are needed, even if the data had their origin in the alternative hypothesis. This demand could invalidate some direct approximations to the distribution of interest. Consider, as in Hall and Wilson (1991), a one-dimensional parameter situation and the simple testing problem $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. Given that $T = T(X_1, \ldots, X_n)$ is a good estimator of the unknown θ, a reasonable test could be based on the difference $T - \theta_0$. If our goal is in relation to its level or p-value, our interest lies in the distribution of $T - \theta_0$ under H_0. A direct —and naïve— application of the bootstrap method would lead to estimate this distribution using the statistic $T^* - \theta_0$, where T^* is the value of T in the resampling, that is, under the empirical distribution of $(X_1, \ldots, X_n)$. In Hall (1992, Sec.3.12) it is shown the bad behaviour of the power function of the corresponding test. The trouble here is that $T^* - \theta_0$ does not approximate the null hypothesis when the sample comes from a parameter far away from θ_0 .

Hall and Wilson (1991) propose to estimate the $T - \theta_0$ distribution by means of the statistic $T^* - T$. Note that this corresponds to resampling the quantity $T_0 = T - \theta$ instead of $T - \theta_0$. In the next section we extend this idea to testing a general linear hypothesis in a linear model.

3 Bootstrap hypothesis testing in the linear model.

It will be convenient to consider the coordinate-free version of the linear model. Let $Y = \mu + e$, where Y denotes the n-dimensional vector of observations, μ is the vector of means belonging to the subspace $V \subset \mathbf{R}^n$ with dimension $p < n$, and e is the vector of independent errors $e_i \sim F$ such that $E(e) = 0$. We want to test $H_0 : \mu \in V_0$ versus $\mu \notin V_0$. being $V_0 \subset V$ a subspace with dimension $p_0 < p$. The usual F-statistic can be written as

$$T = T(Y) = \frac{\|\mathbf{P}_{V|V_0} Y\|^2/(p-p_0)}{\|\mathbf{P}_{V^\perp} Y\|^2/(n-p)}, \tag{3.1}$$

where $V^\perp$, $V|V_0$, and $\mathbf{P}_W$ denote, respectively, the subspace orthogonal to V, the orthogonal complement of V_0 in V, and the orthogonal projection onto W.

Here is an schematic outline of the bootstrap procedure for this version of the linear model. We start from the model (μ, F); after observing the response vector Y, we adjust $(\hat{\mu}, \hat{F}_n)$, being $\hat{\mu} = \mathbf{P}_V Y$ and $\hat{F}_n$ the empirical distribution of the residual vector $\hat{e} = Y - \hat{\mu} = \mathbf{P}_{V^\perp} Y$, that we suppose centered at 0. The bootstrap estimation of an arbitrary function $R = R(Y, F)$ is the (conditional) distribution of $R^* = R(Y^*, \hat{F}_n)$ where $Y^* = \hat{\mu} + e^*$, and the vector e^* has independent components $e_i^* \sim \hat{F}_n$.

The bootstrap methodology proposes to consider as the critical region of nominal α-level $T > \hat{t}_\alpha$, where $\hat{t}_\alpha$ is the $1-\alpha$ quantile of the bootstrap distribution of T under samples coming from H_0 , i.e., $P(T(Y^*) \leq \hat{t}_\alpha \mid Y \in H_0) = 1 - \alpha$.

Hence, we need to know —or, in the practice, to approximate by Monte Carlo trials— this null bootstrap distribution. For this, we propose to consider

$$T_0 = T_0(Y, \mu) = \frac{\|\mathbf{P}_{V|V_0}(Y-\mu)\|^2/(p-p_0)}{\|\mathbf{P}_{V^\perp}(Y-\mu)\|^2/(n-p)} = \frac{\|\mathbf{P}_{V|V_0} e\|^2/(p-p_0)}{\|\mathbf{P}_{V^\perp} e\|^2/(n-p)}. \tag{3.2}$$

Comments:

i) Since it is based on $Y - \mu$, T_0 does not relies on the hypothesis under which the data are obtained; in other words, it is invariant against the hypothesis generating the data. Besides, T_0 agrees with T under H_0.

ii) A direct reasoning leading to T_0 is the following. Since we attempt to approximate the null distribution of T, let us take an arbitrary $\mu_0 \in V$ and transform Y to $Y_0 = Y - \mu + \mu_0$. The vector Y_0 verifies H_0 and, we are naturally conducted to

$$T(Y_0) = \frac{\|\mathbf{P}_{V|V_0} Y_0\|^2/(p-p_0)}{\|\mathbf{P}_{V^\perp} Y_0\|^2/(n-p)} = \frac{\|\mathbf{P}_{V|V_0}(Y-\mu)\|^2/(p-p_0)}{\|\mathbf{P}_{V^\perp}(Y-\mu)\|^2/(n-p)} = T_0(Y, \mu).$$

Note that the above expression is independent of the initial vector $\mu_0 \in V$.

Therefore, the bootstrap distribution of $T_0(Y,\mu)$ is the null bootstrap distribution of T; consequently, the critical value $\hat{t}_\alpha$ should be taken verifying

$$P(T_0^* = T_0(Y^*,\mu) \leq \hat{t}_\alpha \mid Y) = 1-\alpha. \tag{3.3}$$

We see that the modification of the F-statistics arises in a natural and direct way. In some sense, this presentation completes the one given in Mammen (1992). Moreover, his results assure the convergence of our proposal.

4 Bootstrapping the one-way ANOVA model.

We now consider the previous proposal in the particular case of testing equality of means in the one-way model.

4.1 Raising the problem.

We assume a set of n_i observations, $Y_{ij}, j = 1,\ldots,n_i$, coming from a population P_i with mean μ_i and variance σ^2, $i = 1,\ldots,p$. We want to test $H_0 : \mu_1 = \cdots = \mu_p$. This problem fits into the framework of last section by defining $Y = (Y_{11},\ldots,Y_{1n_1},\ldots,Y_{pn_p})'$, $n = \sum_{i=1}^p n_i$, $\mu = (\mu_1,\ldots,\mu_p)'$; V_0 and V are now the spaces spanned, respectively, by the vector $1_n = (1,\overset{n)}{\ldots},1)'$ and by the p vectors in $\mathbb{R}^n$: $(1'_{n_1},0)', (0,1'_{n_2},0)', \ldots, (0,1'_{n_p})'$. It is well know that

$$T(Y) = \frac{\sum_i n_i(\overline{Y}_i - \overline{Y})^2/(p-1)}{\sum_i\sum_j (Y_{ij}-\overline{Y}_i)^2/(n-p)}. \tag{4.1}$$

The statistic T_0 in (3.2) is now

$$T_0(Y,\mu) = \frac{\sum_i n_i(\overline{X}_i - \overline{X})^2/(p-1)}{\sum_i\sum_j (X_{ij}-\overline{X}_i)^2/(n-p)} = \frac{\sum_i n_i(\overline{e}_i - \overline{e})^2/(p-1)}{\sum_i\sum_j (e_{ij}-\overline{e}_i)^2/(n-p)}, \tag{4.2}$$

where $X_{ij} = Y_{ij} - \mu_i = e_{ij}$, $\overline{X}_i = \sum_j X_{ij}/n_i$, $\overline{X} = \sum_i\sum_j X_{ij}/n$.

4.2 Resampling schemes.

To obtain a good approximation through the bootstrap approach in a general arbitrary situation, the resampling distribution must reflect the model distribution in an adequate way. In our concrete situation this means that the bootstrap distribution of T_0 must be obtained "mimicking" the considered model. The one-way model will allow us to illustrate this comment by considering two slightly different sets of assumptions on the underlying populations.

Resampling R_1: Identical populations.

This is the standard case where the distributions associated to the populations differ only in the means, that is, $e_{ij} \sim F_i = F$, $i = 1, \ldots, p$. Since errors are interchangeable through the populations, we should take $\hat{F}_n$ as the empirical distribution of the n residuals $\hat{e}_{ij} = Y_{ij} - \overline{Y}_i$, $i = 1, \ldots, p$, $j = 1, \ldots, n_i$. Recalling the schematic outline presented after (3.1), we have

$$T_0^* = \frac{\sum_i n_i (\overline{X}_i^* - \overline{X}^*)^2/(p-1)}{\sum_i \sum_j (X_{ij}^* - \overline{X}_i^*)^2/(n-p)} = \frac{\sum_i n_i (\overline{e}_i^* - \overline{e}^*)^2/(p-1)}{\sum_i \sum_j (e_{ij}^* - \overline{e}_i^*)^2/(n-p)}, \tag{4.3}$$

where $X_{ij}^* = Y_{ij}^* - \hat{\mu}_i = e_{ij}^* + \overline{Y}_i - \overline{Y}_i = e_{ij}^* \sim \hat{F}_n$.

In practice, the simulation of the T_0^* distribution will be done generating B independent samples from the empirical distribution of the n residuals $\hat{e}_{ij}$, that is, considering B sets with n elements chosen with replacement from $R = \{\hat{e}_{ij}, i = 1, \ldots, p, j = 1, \ldots, n_i\}$.

Resampling R_2: Different populations.

Assume now the distributions associated with every population are different not only in the means but in other aspects; that is, e_{ij} are independent and $e_{ij} \sim F_i$, $j = 1, \ldots, n_i$, where F_i are different with mean 0 and variance σ^2, $i = 1, \ldots, p$. To reflect the new joint distribution of the vector of errors, we will take now $\hat{F}_n$ as the joint distribution of the p empirical independent distributions, $\hat{F}_{in_i}$, corresponding to the sets $R_i = \{\hat{e}_{ij}, j = 1, \ldots, n_i\}$, $i = 1, \ldots, p$. The statistic T_0^* still has the expresion

(4.3), with $X_{ij}^* = e_{ij}^* \sim \hat{F}_{in_i}$.

In practice, T_0^* will be simulated from B sets of n elements, n_i of them will be selected with replacement from R_i, $i = 1, \ldots, p$. Note that resampling from R_i is equivalent to resampling from $O_i = \{Y_{ij}, j = 1, \ldots, n_i\}$, then obtaining $O_i^* = \{Y_{ij}^*, j = 1, \ldots, n_i\}$, and, finally, defining $X_{ij}^* = Y_{ij}^* - \overline{Y}_i$, $j = 1, \ldots, n_i$.

4.3 Normalizing the residuals.

The idea of reflecting the model hypothesis in the bootstrap scheme, leads us to consider the resampling in a population of residuals where the empirical variances in each subpopulation are equal; without loss of generality we take this common value equal to 1. This leads us to replace (4.3) by

$$T_N^* = \frac{\sum_i n_i(\overline{e}_{Ni}^* - \overline{e}_N^*)^2/(p-1)}{\sum_i \sum_j (e_{Nij}^* - \overline{e}_{Ni}^*)^2/(n-p)}, \tag{4.4}$$

where $e_{Nij}^* \sim \hat{F}_{Nn}$, and $\hat{F}_{Nn}^*$ is the empirical (joint) distribution of the standardized residuals: $\hat{e}_{Nij} = (Y_{ij} - \overline{Y}_i)/\hat{\sigma}_i$, $\hat{\sigma}_i^2 = \sum_j (Y_{ij} - \overline{Y}_i)^2/n_i$.

Note that the statistic T_N^* can be also obtained in the following way. Assume initially heteroscedasticity and consider, as alternative to (4.2),

$$T_{0N}(Y, \mu, \sigma_1, \ldots, \sigma_p) = \frac{\sum_i n_i(\overline{W}_i - \overline{W})^2/(p-1)}{\sum_i \sum_j (W_{ij} - \overline{W}_i)^2/(n-p)},$$

where $W_{ij} = (Y_{ij} - \mu_i)/\sigma_i$, and σ_i^2 is the populational variance of P_i. If there exists homoscedasticity, T_{0N} becomes T_0 and its distribution is the same as the distribution of T under H_0. It follows inmediately that the bootstrap distribution of T_{0N}^* coincides with the distribution of T_N^*. Heuristically, we should expect that this statistic will improve the approximation to the null distribution.

The simulation of T_N^* will be done similarly as in Subsection 4.2., but starting with the normalized residuals $\hat{e}_{Nij}$. Both schemes above cited are still valid using now the normalized residuals. In the next section, we will denote them resampling R_3 and R_4.

5 A simulation study.

In this section we illustrate the behaviour of our proposal by means of a simulation study . The results, obtained using only 199 bootstrap resamples, generally indicate a satisfactory performance.

Through our simulation study, we have maintained three populations, i.e., $p = 3$. We considered combinations of two nominal levels, $\alpha = .05,\ .1$, two total sample sizes, $n = 30,\ 60$, and three groups of error distributions associated to the populations, $G_k,\ k = 1, 2, 3$. The composition of these groups is the following. In G_1 the errors have normal distributions; in G_2 they have shifted exponential distributions; and, in G_3 the errors have different distributions: $N(0, 1)$, shifted exponential and normalized t_3.

The computations were performed using Fortran routines running on a DECstation 5000/200 under Ultrix-32. The routines GGUBS, GGNML and GGAMR of the IMSL library are used to generate the pseudo-random variables and to make the resamplings.

5.1 Results under the null hypothesis.

Without loss of generality, we took $\mu_1 = \mu_2 = \mu_3 = 0$. The specific distributions for the above groups were: $N(0, 1)$, shifted exponential with $\lambda = 1$, and t_3 normalized to have variance unity.

We conducted 1000 trials under each of the 28 combinations of 4 particular population sizes and 7 types of resampling listed in Table 1. In each trial, $B = 199$ resamples were drawn from the empirical distribution associated to the resampling. By means of (3.3). these resamples provided the approximate $\hat{t}_{\alpha i}$ values and, in each trial, H_0 was rejected if $T(Y) > \hat{t}_{\alpha i}, \alpha = .05, .1, i = 1, \ldots, 1000$. (Specifically, $\hat{t}_{\alpha i}$ was the 9th largest of the 199 values of T_0^* when $\alpha = .05$, and the 19th largest when $\alpha = .1$.)

The entries of Table 1 are the bootstrap levels, α^*, that is, the proportion of trials rejecting H_0. The symbol # denotes cases with α^* levels differing significantly of α.

Error distributions		$G_1 : N(0,1)$		$G_2 : E(1)$		$G_3 : N(0,1), E(1), t_3$		
Resamplings		R_1	R_3	R_1	R_3	R_1	R_2	R_4
Population sizes								
$(10,10,10)$	$\alpha = .05$	0.0530	0.0550	0.0510	0.0480	0.0490	0.0320#	0.0420
	$\alpha = .1$	0.0990	0.0990	0.0950	0.0960	0.0980	0.0800#	0.0960
$(5,10,15)$	$\alpha = .05$	0.0380	0.0420	0.0440	0.0410	0.0600	0.0430	0.0520
	$\alpha = .1$	0.0910	0.0900	0.1090	0.1110	0.1150	0.0900	0.1100
$(20,20,20)$	$\alpha = .05$	0.0430	0.0470	0.0460	0.0470	0.0500	0.0430	0.0500
	$\alpha = .1$	0.0960	0.0930	0.0970	0.0970	0.1120	0.0940	0.1050
$(10,20,30)$	$\alpha = .05$	0.0460	0.0430	0.0570	0.0560	0.0630	0.0430	0.0520
	$\alpha = .1$	0.1010	0.0970	0.1100	0.1070	0.1060	0.0900	0.0970

Table 1: Bootstrap levels for 28 combinations of error distributions, resamplings, and population sizes.

Figure 1 depicts the values of α^* for the last row of Table 1 as a function of the number of trials. The 40 points in the trajectories correspond to the values computed from each consecutive 25 trials.

Some comments are in order. Most of the bootstrap levels are very close to the nominal levels, specially when the appropriate resamplings are used, and even when the inadequate resamplig R_1 is used in G_3. The performances under normal or exponential distributed errors are very similar, pointing out that asymmetric errors are not troublesome. The stability of the trajectories is reached soon. For the group G_3 (different populations), the results with resampling R_2 are lightly better than with R_1 and both are dominated by the results using R_4.

(Insert Figure 1 about here)

5.2 Results under heteroscedasticity.

Although our development does not deal with heteroscedastic situations, we planned to check how our proposal would perform under unequal variances. With this aim, we considered the groups G_1 and G_3. In G_1 the normal distributions had standard deviations $\sigma_i = i,\ i = 1, 2, 3$; in G_3 we included standard normal, shifted exponential with $\sigma = 2$, and t_3 ($\sigma = 3$) distributions. The total sizes ($n = 30, 60$) were assigned to the populations in three ways: balanced, higher sizes to higher variances, and viceversa. For each assignation, resamplings R_1 and R_2 were considered.

Both groups and sizes provided similar results. Figure 2 shows the 6 bootstrap level trajectories for G_1 and $n = 30$. As a summary of empirical conclusions we have: the resampling R_1 was very unappropriated with unbalanced sizes, the resampling R_1 was beaten by R_2 in all the situations; therefore, if we suspect of possible heteroscedasticity, the resampling starting from different populations will provide more accurate levels. The best results occur when there is correspondence between sizes and variability in populations.

(Insert Figure 2 about here)

5.3 Results under the alternative hypothesis.

We have also verified the good approximation to the true critical values and the high power provided by the bootstrap test when the data come from populations with unequal means. Our illustration will use some of the combinations considered in Section 5.1, being now $\mu_1 = -1,\ \mu_2 = 0,\ \mu_3 = 1$.

(Insert Figure 3 about here)

Figure 3 depicts box-plots (with whiskers ending at the extreme values) of 1000 bootstrap estimated critical values, $\hat{t}_{\alpha i}$, obtained from (3.3). The errors belong to group G_1 with $N(0,1)$ and G_2 with shifted exponential, $\lambda = 1$. (The true values are $F_{2,27,.05} = 3.35$ in G_1 and were obtained by simulation using 5000 trials in G_2.)

Error distributions		$G_1 : N(0,1)$		$G_2 : E(1)$		$G_3 : N(0,1), E(1), t_3$		
Resamplings		R_1	R_3	R_1	R_3	R_1	R_2	R_4
Population sizes								
$(10,10,10)$	$\alpha = .05$	0.9720	0.9700	0.9550	0.9530	0.9680	0.9220	0.9520
	$\alpha = .1$	0.9870	0.9860	0.9790	0.9760	0.9770	0.9560	0.9750
$(5,10,15)$	$\alpha = .05$	0.9370	0.9320	0.9250	0.9230	0.9160	0.8660	0.8920
	$\alpha = .1$	0.9680	0.9680	0.9670	0.9640	0.9510	0.9230	0.9330
$(20,20,20)$	$\alpha = .05$	1.0000	1.0000	1.0000	1.0000	0.9980	0.9930	0.9980
	$\alpha = .1$	1.0000	1.0000	1.0000	1.0000	0.9980	0.9950	0.9980
$(10,20,30)$	$\alpha = .05$	1.0000	1.0000	0.9930	0.9950	0.9980	0.9910	0.9950
	$\alpha = .1$	1.0000	1.0000	0.9970	0.9970	0.9980	0.9940	0.9970

Table 2: Power in H_A: $\mu_1 = -1$, $\mu_2 = 0$, $\mu_3 = 1$, for the 28 combinations of Table 1.

Note the concentration around the true critical values, and the similar results given by R_1 and R_3.

Table 2 gives the proportion of trials rejecting the equality of means, that is, the power, under the above alternative for the combinations considered in Section 5.1. It is worthwhile noting the high values attained.

5.4 Bootstrapping the F-statistic directly

Finally, we will show the inaccuracy of the naïve bootstrap directly based on the F-statistic of (4.1), that is, when the resampling uses

$$T_D^* = \frac{\sum_i n_i(\overline{Y}_i^* - \overline{Y}^*)^2/(p-1)}{\sum_i \sum_j (Y_{ij}^* - \overline{Y}_i^*)^2/(n-p)},$$

where $Y_{ij}^* = e_{ij}^* + \bar{Y}_i$, and e_{ij}^* follow one of the empirical distributions presented in Sections 4.2 and 4.3. We will refer this procedure as *direct* bootstrap.

We began conducting 1000 trials for the situation and combinations considered in Section 5.1. The direct bootstrap levels, α_D^*, turned out to be much lower than the

nominal levels —in fact, only 5 of 56 cases considered were different of zero with a maximum value of .004. We also studied the distribution of the direct critical values, $\hat{t}^D_{\alpha i}$, $i = 1, \ldots, 1000$, $\alpha = .05, .1$, corresponding to 4 combinations: groups G_1 and G_2 with resampling R_1, and G_3 with R_1 and R_3. The population sizes were $n_i = 10$. This study was done with data from H_0 and from the alternative $H_A : \mu_1 = -1, \mu_2 = 0, \mu_3 = 1$.

The distributions of the direct critical values were very similar in the four combinations. Figure 4 shows box-plots summarizing the distributions under H_0 and H_A, $\alpha = .05$, $n = 30$, in the combination G_1 and R_1. For comparison, we have also included box-plots for the distributions of the critical values, $\hat{t}^*_{\alpha i}$, derived from our proposal.

(Insert Figure 4 about here)

Two points clearly stand out. Whether the data came from H_0 or from H_A, the distribution of critical values $\hat{t}^D_{\alpha i}$ is useless to approximate the true critical value ($F_{2,28,.05} = 3.35$). On the contrary, both from H_0 and H_A, the distributions of proposed critical values $\hat{t}_{\alpha i}$ are extremely similar and very concentrated around the true critical value.

References

Beran, R., Prepivoting tests statistics: A bootstrap view of asymptotic refinements, *J. Amer. Statist. Assoc.*, **83** (1988) 687–697.

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

Hall, P. and Wilson, S.R., Two guidelines for bootstrap hypothesis testing, *Biometrics*, **47** (1991) 757–762.

Hinkley, D. V., Bootstrap methods, *J. Roy. Statist. Soc. Ser. B*,**50** (1988) 321–337.

Mammen, E., *When does bootstrap work?* (Lecture Notes in Statistics, Vol. 77, Springer-Verlag, New York, 1992)

Romano, J. H., A bootstrap revival of some nonparametric distance tests, *J. Amer. Statist. Assoc.*, **83** (1988) 698–708.

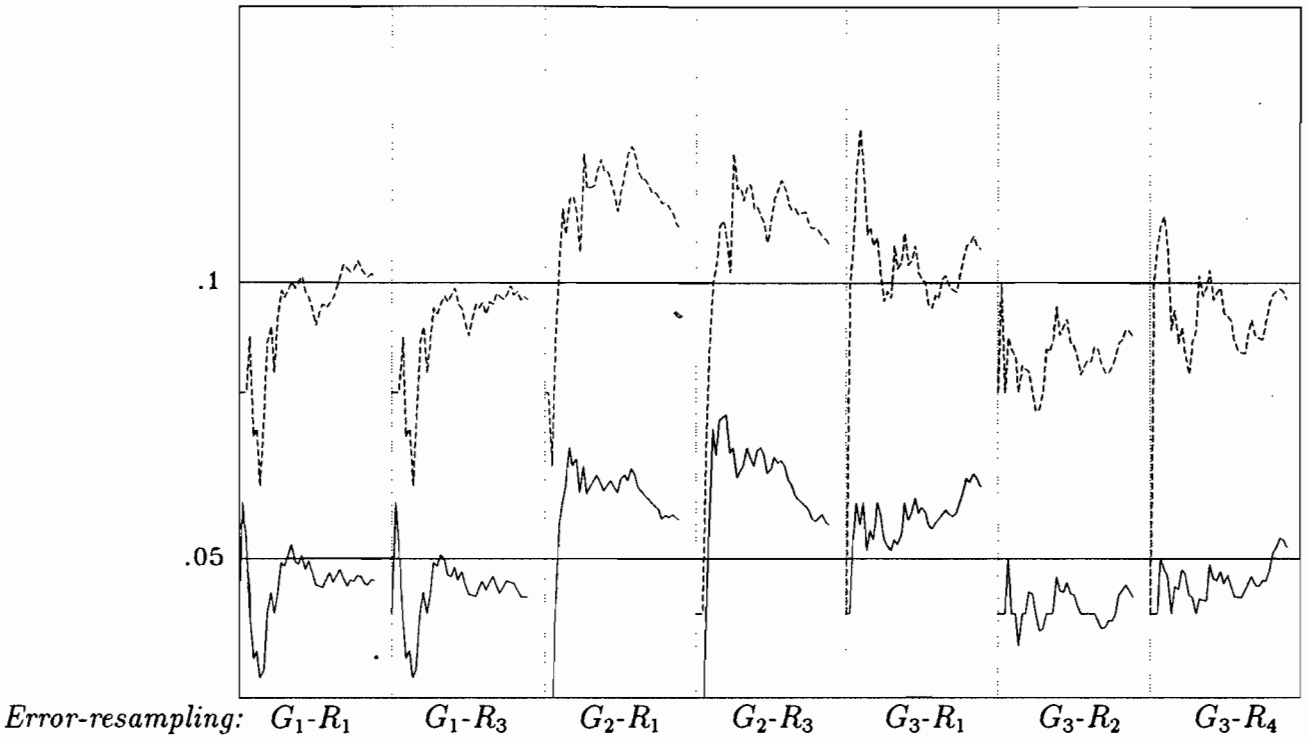

Fig. 1: Bootstrap levels as function of the number of trials (N=1000). The 40 points correspond to computations from each consecutive 25 trials.

—— $\alpha = .05$, - - - $\alpha = .1$.

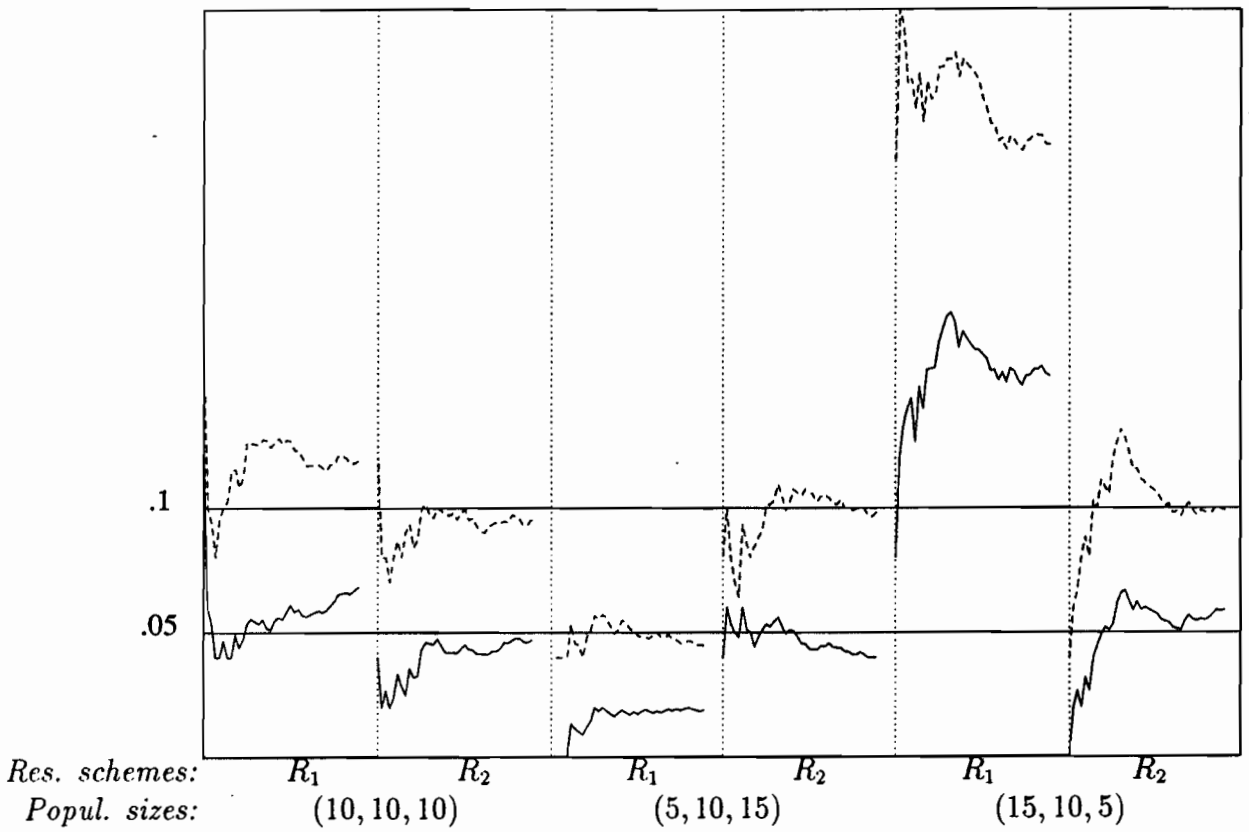

Fig. 2: Bootstrap levels with heteroscedastic data.

Error distributions: $N(0, \sigma_i = i)$, $i = 1, 2, 3$. —— $\alpha = .05$, - - - $\alpha = .1$.

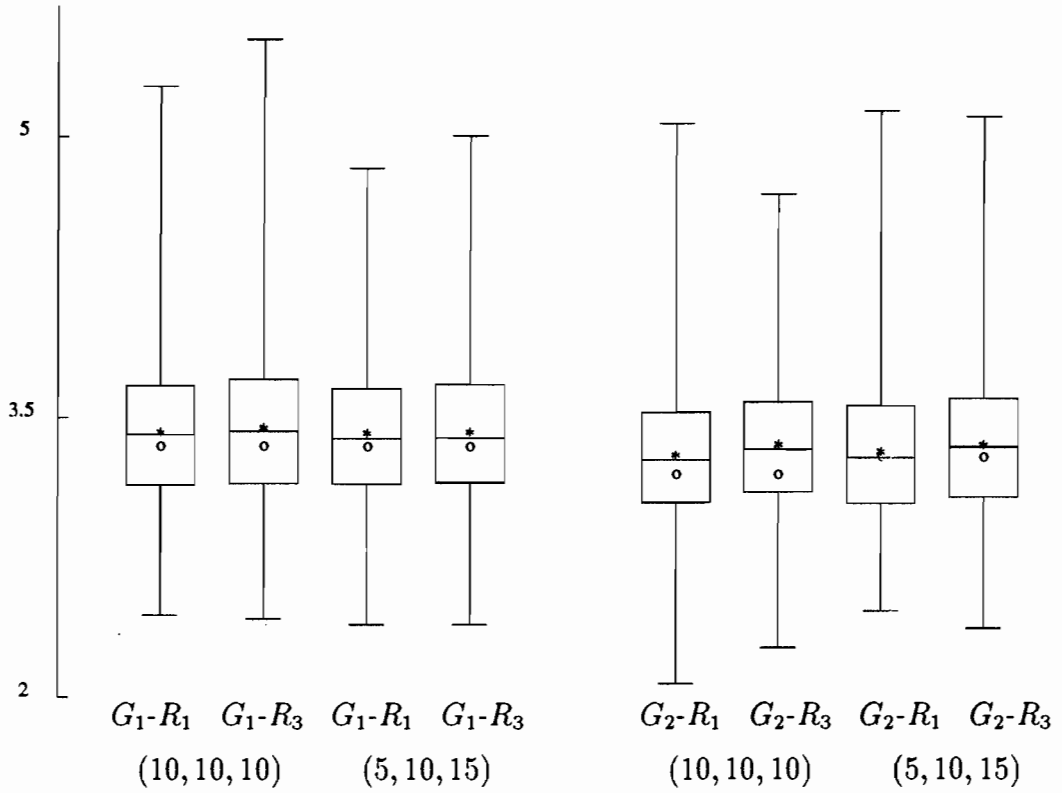

Fig. 3: Box-plots of bootstrap critical values (N=1000) with data from H_A: $\mu_1 = -1,\ \mu_2 = 0,\ \mu_3 = 1$.

$*$ Mean of bootstrap critical values. $\circ$ True critical value.

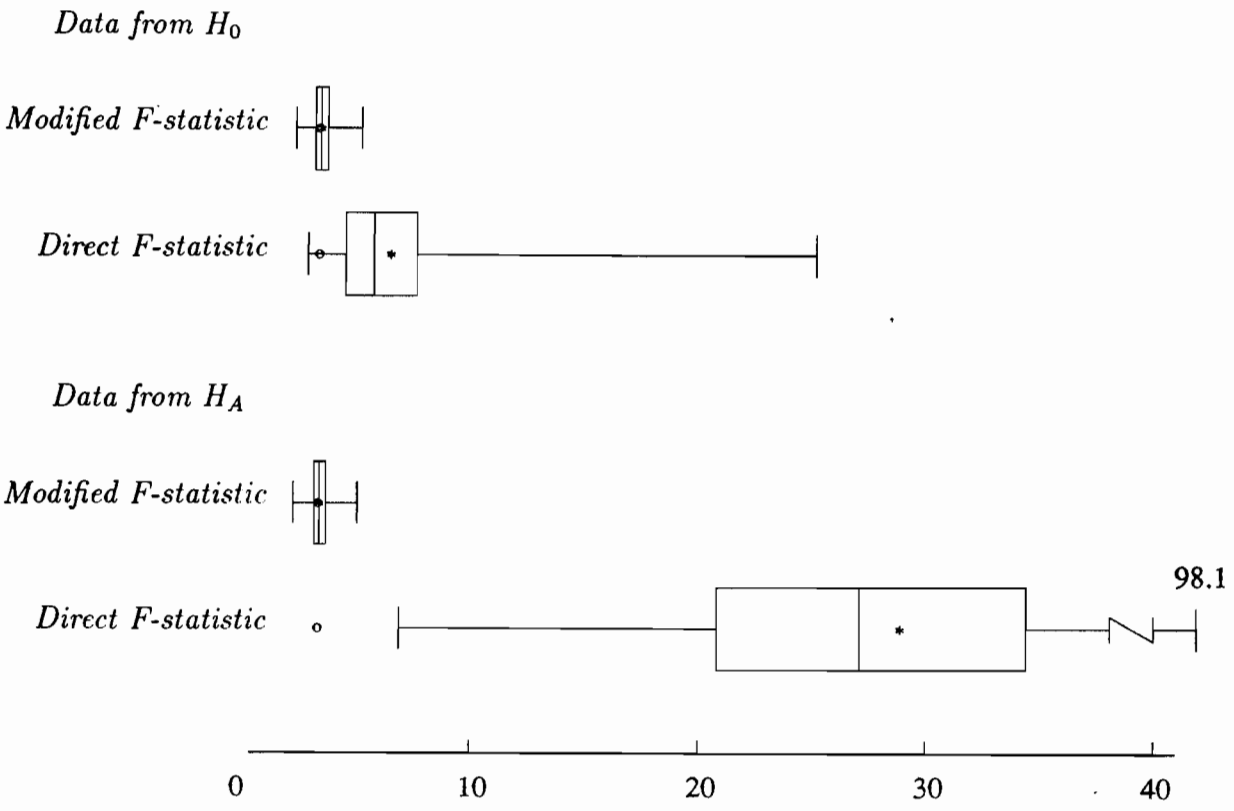

Fig. 4: Box-plots of bootstrap critical values (N=1000) using modified and direct F-statistics.

∗ Mean of bootstrap critical values. o True critical value