The Wild Bootstrap, Tamed at Last

by

Russell Davidson

GREQAM
Centre de la Vieille Charité
2 rue de la Charité
13002 Marseille, France

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

russell@ehess.cnrs-mrs.fr

and

Emmanuel Flachaire

STICERD
London School of Economics
Houghton Street
London WC2A 2AE UK

e.flachaire@lse.ac.uk

Abstract

Various versions of the wild bootstrap are studied as applied to regression models with heteroskedastic errors. We develop formal Edgeworth expansions for the error in the rejection probability (ERP) of wild bootstrap tests based on asymptotic t statistics computed with a heteroskedasticity consistent covariance matrix estimator. Particular interest centers on the choice of the auxiliary distribution used by the wild bootstrap in order to generate bootstrap error terms. We find that the Rademacher distribution usually gives smaller ERPs, in small samples, than the version of the wild bootstrap that seems most popular in the literature, even though it does not benefit from the latter's skewness correction. This conclusion, based on Edgeworth expansions, is confirmed by a series of simulation experiments, which we also use to study some other points, such as the use of constrained or unconstrained residuals in the HCCME, about which the expansions give no definite conclusions. We conclude that a particular version of the wild bootstrap is to be preferred in almost all practical situations, and we show analytically that it, and no other version, gives perfect inference in a special case.

This research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada. We are very grateful to James MacKinnon for helpful comments on an earlier draft, to participants at the ESRC Econometrics Conference (Bristol), especially Whitney Newey, and to Joel Horowitz. Remaining errors are ours.

August, 2001

1. Introduction

Inference on the parameters of the linear regression model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u},$$

where $\boldsymbol{y}$ is an n-vector containing the values of the dependent variable, $\boldsymbol{X}$ an $n \times k$ matrix of which each column is an explanatory variable, and $\boldsymbol{\beta}$ a k-vector of parameters, requires special precautions when the error terms $\boldsymbol{u}$ are heteroskedastic, a problem that arises frequently in work on cross-section data. With heteroskedastic errors, the usual OLS estimator of the covariance of the OLS estimates $\hat{\boldsymbol{\beta}}$ is in general asymptotically biased, and so conventional t and F tests do not have their namesake distributions, even asymptotically, under the null hypotheses that they test. The problem was solved by Eicker (1963) and White (1980), who proposed a heteroskedasticity consistent covariance matrix estimator, or HCCME, that permits asymptotically correct inference on $\boldsymbol{\beta}$ in the presence of heteroskedasticity of unknown form.

MacKinnon and White (1985) considered a number of possible forms of HCCME, and showed that, in finite samples, they too, as also t or F statistics based on them, can be seriously biased; see also Chesher and Jewitt (1987), who showed that the extent of the bias is related to the structure of the regressors, and in particular to the presence of observations with high leverage. But since, unlike conventional t and F tests, HCCME-based tests are at least asymptotically correct, it makes sense to consider whether bootstrap methods might be used to alleviate their small-sample size distortion.

Bootstrap methods normally rely on simulation to approximate the finite-sample distribution of test statistics under the null hypotheses they test. In order for such methods to be reasonably accurate, it is desirable that the data-generating process (DGP) used for drawing bootstrap samples should be as close as possible to the true DGP that generated the observed data, assuming that that DGP satisfies the null hypothesis. This presents a problem if the null hypothesis admits heteroskedasticity of unknown form: If the form is unknown, it cannot be imitated in the bootstrap DGP.

In the face of this difficulty, the so-called wild bootstrap was developed by Liu (1988) following a suggestion of Wu (1986) and Beran (1986). Liu established the ability of the wild bootstrap to provide refinements for the linear regression model with heteroskedastic errors, and further evidence was provided by Mammen (1993), who showed, under a variety of regularity conditions, that the wild bootstrap is asymptotically justified, in the sense that the asymptotic distribution of various statistics is the same as the asymptotic distribution of their wild bootstrap counterparts. These authors also show that, in some circumstances, asymptotic refinements are available, which lead to agreement between the distributions of the raw and bootstrap statistics to higher than leading order asymptotically.

In this paper, we consider a number of implementations both of the Eicker-White HCCME and of the wild bootstrap applied to them. We show that, when the error terms are symmetrically distributed about the origin, the wild bootstrap applied to HCCME based statistics benefits from better asymptotic refinements than when they are asymmetrically distributed. We obtain an explicit Edgeworth expansion, valid through order n^{-1}, for the rejection probability of tests based on the wild bootstrap. The expansion yields results that do not agree very well quantitatively with results obtained by simulation, but it usually gives a fairly good idea of the order of magnitude of the error in the rejection probability (ERP). In particular, by providing explicit expressions for the coefficients in the expansion, it shows that the order of a term, as a negative power of n, is by no means always the most important determinant of its quantitative importance in not very large samples.

In section 2, we discuss a number of implementations of the wild bootstrap for asymptotic t and χ^2 tests based on various versions of the HCCME. Then, in Section 3, we study the ERPs of wild bootstrap tests by means of formal Edgeworth expansions. In Section 4, we prove a theorem that shows that one of these tests yields perfect inference, up to a small error due to discretization, for a special case. Then, in section 5, simulation experiments are described designed to measure the reliability of various tests, bootstrap and asymptotic, in various conditions, including very small samples, and to compare ERPs estimated by simulation with the predictions of Edgeworth expansions. These experiments give strong evidence in favor of the version of the wild bootstrap that gives perfect inference in the special case. A few conclusions are drawn in section 6.

2. The Wild Bootstrap

Consider the linear regression model

$$y_t = \boldsymbol{X}_{t1}\boldsymbol{\beta}_1 + \boldsymbol{X}_{t2}\boldsymbol{\beta}_2 + u_t, \quad t = 1, \ldots, n, \tag{1}$$

in which the explanatory variables are assumed to be strictly exogenous, in the sense that, for all t, $\boldsymbol{X}_{t1}$ and $\boldsymbol{X}_{t2}$ are independent of all of the error terms u_s, $s = 1, \ldots, n$. The row vectors $\boldsymbol{X}_{t1}$ and $\boldsymbol{X}_{t2}$ contain observations on k_1 and k_2 variables respectively, with $k_1 + k_2 = k$. We wish to test the null hypothesis that the parameter vector $\boldsymbol{\beta}_1$ is zero. The error terms are assumed to be mutually independent and to have a common mean of zero, but they may be heteroskedastic, with $\mathrm{E}(u_t^2) = \sigma_t^2$. We write $u_t = \sigma_t v_t$, where $\mathrm{E}(v_t^2) = 1$. We consider only *unconditional* heteroskedasticity, which means that the σ_t^2 may depend on the exogenous regressors, but not, for instance, on lagged dependent variables. The different DGPs contained in model (1) are characterized by the parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, the variances σ_t^2, and the probability distributions of the v_t. The regressors are taken as fixed and the same for all DGPs contained in the model. HCCME-based χ^2 statistics for testing whether $\boldsymbol{\beta}_1 = \boldsymbol{0}$, or t statistics if $k_1 = 1$, are then asymptotically pivotal for the restricted model in which we set $\boldsymbol{\beta}_1 = \boldsymbol{0}$, under weak regularity conditions on the regressors and the σ_t.

We write $\boldsymbol{X}_i$ for the $n \times k_i$ matrix with typical row $\boldsymbol{X}_{ti}$, $i = 1, 2$, and by $\boldsymbol{X}$ we mean the full $n \times k$ matrix $[\boldsymbol{X}_1 \quad \boldsymbol{X}_2]$. Then the basic HCCME for the OLS parameter estimates of (1) is

$$(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \hat{\boldsymbol{\Omega}} \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}, \tag{2}$$

where the $n \times n$ diagonal matrix $\hat{\boldsymbol{\Omega}}$ has typical diagonal element $\hat{u}_t^2$, where the $\hat{u}_t$ are the OLS residuals from the estimation either of the unconstrained model (1) or the constrained model in which $\boldsymbol{\beta}_1 = \boldsymbol{0}$ is imposed. We refer to the version (2) of the HCCME as HC_0. Bias is reduced by multiplying the $\hat{u}_t$ by the square root of $n/(n-k)$, thereby multiplying the elements of $\hat{\boldsymbol{\Omega}}$ by $n/(n-k)$; this procedure, analogous to the use in the homoskedastic case of the unbiased OLS estimator of the error variance, gives rise to form HC_1 of the HCCME. In the homoskedastic case, the variance of $\hat{u}_t$ is proportional to $1 - h_t$, where $h_t \equiv \boldsymbol{X}_t(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}_t^\top$, the t^{th} diagonal element of the orthogonal projection matrix on to the span of the columns of $\boldsymbol{X}$. This suggests replacing the $\hat{u}_t$ by $\hat{u}_t/(1-h_t)^{1/2}$ in order to obtain $\hat{\boldsymbol{\Omega}}$. If this is done, we obtain form HC_2 of the HCCME. Finally, arguments based on the jackknife lead MacKinnon and White to propose form HC_3, for which the $\hat{u}_t$ are replaced by $\hat{u}_t/(1 - h_t)$. MacKinnon and White (1985), and Chesher and Jewitt (1987), show that, in terms of size distortion, HC_0 is outperformed by HC_1, which is in turn outperformed by HC_2 and HC_3. The last two cannot be ranked in general, although HC_3 has been shown in a number of Monte Carlo experiments to be superior in typical cases.

As mentioned in the introduction, heteroskedasticity of unknown form cannot be mimicked in the bootstrap distribution. The wild bootstrap gets round this problem by using a bootstrap DGP of the form

$$y_t^* = \boldsymbol{X}_t \hat{\boldsymbol{\beta}} + u_t^*, \tag{3}$$

where $\hat{\boldsymbol{\beta}}$ is a vector of parameter estimates, and the bootstrap error terms are

$$u_t^* = f_t(\hat{u}_t)\varepsilon_t,$$

where $f_t(\hat{u}_t)$ is a transformation of the OLS residual $\hat{u}_t$, and the ε_t are mutually independent drawings, completely independent of the original data, from some auxiliary distribution such that

$$\mathrm{E}(\varepsilon_t) = 0 \qquad \text{and} \qquad \mathrm{E}(\varepsilon_t^2) = 1. \tag{4}$$

Thus, for each bootstrap sample, the exogenous explanatory variables are reused unchanged, as are the OLS residuals $\hat{u}_t$ from the estimation using the original observed data. The transformation $f_t(\cdot)$ can be used to modify the residuals, for instance by dividing by $1 - h_t$, just as in the different variants of the HCCME.

In the literature, the further condition that $E(\varepsilon_t^3) = 1$ is often added. Liu (1988) considers model (1) with $k = 1$, and shows that, with the extra condition, the first three moments of the bootstrap distribution of an HCCME-based statistic are in accord with those of the true distribution of the statistic up to order n^{-1}. Mammen (1993) suggested what is probably the most popular choice for the distribution of the ε_t, namely the following two-point distribution:

$$F_1: \quad \varepsilon_t = \begin{cases} -(\sqrt{5}-1)/2 & \text{with probability } p = (\sqrt{5}+1)/(2\sqrt{5}) \\ (\sqrt{5}+1)/2 & \text{with probability } 1-p. \end{cases} \tag{5}$$

Liu also mentions the possibility of Rademacher variables, defined as

$$F_2: \quad \varepsilon_t = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2, \end{cases} \tag{6}$$

which, for estimation of a mean, satisfies necessary conditions for refinements in the case of unskewed error terms. Unfortunately, she does not follow up this possibility, since (6), being a lattice distribution, does not lend itself to rigorous techniques based on Edgeworth expansion.

Conditional on the random elements $\hat{\boldsymbol{\beta}}$ and $\hat{u}_t$, the wild bootstrap DGP (3) clearly belongs to the model constituting the null hypothesis if the subvector $\hat{\boldsymbol{\beta}}_1$, corresponding to the regressors in $\boldsymbol{X}_1$, is zero, since the bootstrap error terms u_t^* have mean zero and are heteroskedastic for any distribution of the ε_t satisfying (4). Since (1) is linear, we may also set the remaining components of $\hat{\boldsymbol{\beta}}$ to zero, since the distribution of any HCCME-based statistic does not depend on the value of $\boldsymbol{\beta}_2$. Since the HCCME-based statistics we have discussed are asymptotically pivotal, inference based on the wild bootstrap using such a statistic applied to model (1) is asymptotically valid. In the case of a nonlinear regression, the distribution of the test statistic does depend on the specific value of $\boldsymbol{\beta}_2$, and so a consistent estimator of these parameters should be used in formulating the bootstrap DGP.

The arguments in Beran (1988) show that bootstrap inference benefits from asymptotic refinements when used with asymptotically pivotal statistics if the random elements in the bootstrap DGP are consistent estimators of the corresponding elements in the unknown true DGP. These arguments do not apply directly to (3), since the squared residuals are not consistent estimators of the σ_t^2. In the next section, we develop formal Edgeworth expansions for the ERP of a bootstrap test based on a t statistic in model (1), and show that asymptotic refinements are nonetheless available in certain circumstances.

3. Formal Edgeworth Expansions

The conventional asymptotic theory of the bootstrap as found, for instance, in Hall (1992), and, in particular, in the theory presented in Liu (1988) and Mammen (1993), relies on Edgeworth expansions in order to prove the existence of asymptotic refinements. In this section we present formal Edgeworth expansions up to order n^{-1} for the distribution of an HCCME-based t statistic and its wild bootstrap counterpart, and for the ERP of a wild bootstrap test based on such a statistic. It is not hard to formulate regularity conditions for the validity of the expansion for the statistic itself, but, since the wild bootstrap statistic has a discrete distribution, it would be much more complicated to find rigorous regularity conditions of sufficient generality. It is not the aim of this paper to investigate this issue, and we therefore limit ourselves to consideration of conditions for the validity of the expansion for the HCCME-based t statistic itself.

The most important of these is just that the regressors in model (1) satisfy the usual condition for the validity of OLS estimation, namely that $\text{plim}_{n\to\infty} n^{-1}\boldsymbol{X}^\top\boldsymbol{X}$ is a finite, positive definite, deterministic matrix. In addition, we assume that the regressor matrix $\boldsymbol{X}$ is independent of the error terms u_t. We also require some conditions that involve both the regressors and the error variances; these will be formulated explicitly in a moment. Basically, they require the existence of various moments, and the finiteness of the limits of these moments as the sample size tends to infinity. Conditions of this sort are enough since we are firmly in the context of the smooth function model, as set out, for instance, in Hall (1988). Precisely to avoid the problem caused by the discreteness of the wild bootstrap statistic, we assume that the error terms u_t are drawn from a continuous distribution on the real line or a subset of the real line.

We suppose that $k_1 = 1$, so that the vector $\boldsymbol{\beta}_1$ becomes a scalar β_1, and the matrix $\boldsymbol{X}_1$ becomes a vector $\boldsymbol{x}_1$. All t statistics based on the HCCME for the hypothesis that $\beta_1 = 0$ can be written as

$$\tau \equiv \boldsymbol{x}_1^\top\boldsymbol{M}_2\boldsymbol{y}/(\boldsymbol{x}_1^\top\boldsymbol{M}_2\hat{\boldsymbol{\Omega}}\boldsymbol{M}_2\boldsymbol{x}_1)^{1/2}. \tag{7}$$

Here $\boldsymbol{y}$ is the n-vector with typical element y_t, $\hat{\boldsymbol{\Omega}}$ is an $n \times n$ diagonal matrix with diagonal elements that depend on the version of the HCCME and on whether residuals from the constrained or unconstrained regression are used, and $\boldsymbol{M}_2 = \mathbf{I} - \boldsymbol{X}_2(\boldsymbol{X}_2^\top\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top$ is the orthogonal projection matrix on to the orthogonal complement of the span of the columns of $\boldsymbol{X}_2$.

Because $\hat{\boldsymbol{\Omega}}$ is diagonal, we can express the matrix product $\boldsymbol{x}_1^\top\boldsymbol{M}_2\hat{\boldsymbol{\Omega}}\boldsymbol{M}_2\boldsymbol{x}_1$ as

$$\sum_{t=1}^n a_t(\boldsymbol{M}_2\boldsymbol{x}_1)_t^2\,\ddot{u}_t^2, \tag{8}$$

where $\ddot{u}_t$ is the t^{th} residual, constrained or unconstrained, and a_t depends on the choice of the functions f_t: $a_t = 1$ for HC_0, $n/(n-k)$ for HC_1, $1/(1-h_t)$ for HC_2, and $1/(1-h_t)^2$ for HC_3. Under the assumed conditions on the regressors, $a_t = 1+O(n^{-1})$ for all t. Under the null, $\boldsymbol{M}_2\boldsymbol{y} = \boldsymbol{M}_2\boldsymbol{u}$, and so the statistic (7) is equal to

$$\tau = \frac{\sum_{t=1}^n(\boldsymbol{M}_2\boldsymbol{x}_1)_t\,u_t}{\left(\sum_{t=1}^n a_t(\boldsymbol{M}_2\boldsymbol{x}_1)_t^2\,\ddot{u}_t^2\right)^{1/2}}. \tag{9}$$

It is clear from this that the statistic depends on the regressor design only through the vector $\boldsymbol{x}_1$ and the space spanned by the columns of $\boldsymbol{X}_2$. In fact, only the component of $\boldsymbol{x}_1$ orthogonal to the columns for $\boldsymbol{X}_2$, that is, the vector $\boldsymbol{M}_2\boldsymbol{x}_1$, has any influence on τ, and then only through its direction; τ is homogeneous of degree 0 in the components of $\boldsymbol{M}_2\boldsymbol{x}_1$, and so does not depend on the norm of that vector. In addition, it is clear that τ is independent of the scale of the error terms, in the sense that multiplying all of the σ_t by the same constant leaves τ unchanged.

We may therefore, without loss of generality, choose to express the regressors in the matrix $\boldsymbol{X}_2$ in such a way that $n^{-1}\boldsymbol{X}_2^\top\boldsymbol{X}_2 = \mathbf{I}$, the $n \times n$ identity matrix. Further, we normalize the vector $\boldsymbol{M}_2\boldsymbol{x}_1$ to have squared norm of n; these operations are admissible for all n on account of our assumption about $\operatorname{plim} n^{-1}\boldsymbol{X}^\top\boldsymbol{X}$. To simplify notation, we henceforth write $\boldsymbol{x} = \boldsymbol{M}_2\boldsymbol{x}_1$, with components x_t, $t = 1, \ldots, n$, and redefine $\boldsymbol{X}$ accordingly as $\boldsymbol{X} \equiv [\boldsymbol{x} \;\; \boldsymbol{X}_2]$, with typical element X_{ti}.

Our main aim at this point is to obtain Edgeworth expansions for the ERP of a test based on applying one of several versions of the wild bootstrap to a t statistic of the form τ in (9), based on one of several versions of the HCCME. These expansions depend on a number of quantities defined in terms of the regressors and the error variances. We make the following definitions:

$$S^2 = n^{-1}\sum_t x_t^2\sigma_t^2, \qquad C = n^{-1}\sum_t x_t^3(\sigma_t/S)^3,$$
$$D = n^{-1}\sum_t x_t^4(\sigma_t/S)^4, \tag{10}$$
$$E_i = n^{-1}\sum_t x_t X_{ti}(\sigma_t/S)^2, \text{ and } F_i = n^{-1}\sum_t x_t^3 X_{ti}(\sigma_t/S)^2, \quad i = 1, \ldots, k.$$

We assume that the limits as $n \to \infty$ of all the quantities in the set of definitions (10) are finite, and that the limit of S is bounded away from zero. Because the distribution of τ is independent of the scale of the σ_t, it would be possible to specify that $S = 1$ and thereby replace the normalized σ_t/S in the definitions following that of S by just σ_t. We do not do so, because all of the above quantities have bootstrap counterparts, and the value of S is different for these, while the matrix $\boldsymbol{X}$ is the same for the original and the bootstrap statistics.

The indexed quantities E_i and F_i can be thought of as the components of vectors of k components. The invariance properties of model (1) imply that these vectors can influence the distribution of τ only through invariant scalar functions defined in terms of them. This will be clear in the results to follow.

We assume that the standardized error terms v_t share the same distribution for all t. This is not a necessary assumption, but making it saves us from further notational complexity. For this common distribution, then, we write $e_3 = E(v_t^3)$ and $e_4 = E(v_t^4)$. We also define $e_3^* = E(\varepsilon_t^3)$, and $e_4^* = E(\varepsilon_t^4)$.

We may now state the following theorem.

Theorem 1: For an HCCME-based t statistic τ of type (9), and a wild bootstrap DGP with bootstrap error terms $u_t^* = f_t(\ddot{u}_t)\varepsilon_t$, where the $\ddot{u}_t$

are the constrained or unconstrained residuals, the transformations f_t correspond to one of the HC_i, $i = 0, 1, 2, 3$, and the ε_t are independent drawings from a distribution satisfying (4), the error in the rejection probability of a one-tailed bootstrap test at nominal level α with rejection in the left-hand tail of the distribution, under a DGP of the form

$$y_t = \boldsymbol{X}_{t2}\boldsymbol{\beta}_2 + \sigma_t v_t, \quad t = 1, \ldots, n,$$

where the v_t are mutually independent, and all follow the same mean-zero, unit-variance distribution, has the following formal expansion through order n^{-1}:

$$\begin{aligned}\phi(z_\alpha)\Bigg(&\tfrac{1}{6}n^{-1/2}(1-e_3^*)e_3C(1+2z_\alpha^2) + n^{-1}\Big(e_3^2C^2(\tfrac{1}{6}z_\alpha - \tfrac{1}{9}z_\alpha^3 - \tfrac{1}{18}z_\alpha^5)\\ &+e_3^*e_3^2C^2(\tfrac{1}{6}z_\alpha + \tfrac{7}{18}z_\alpha^3 + \tfrac{1}{9}z_\alpha^5) + (e_3^*)^2e_3^2C^2(-\tfrac{1}{12}z_\alpha + \tfrac{2}{9}z_\alpha^3 - \tfrac{1}{18}z_\alpha^5)\\ &+\tfrac{1}{12}e_4D\big(z_\alpha(3(e_4^*-1) - 2e_3^*) - z_\alpha^3((e_4^*-1) + 4e_3^*)\big) + e_3^*E(\tfrac{1}{2}z_\alpha + z_\alpha^3)\Big)\Bigg).\end{aligned} \tag{11}$$

Here, $\phi(\cdot)$ is the standard normal density and z_α is the α-quantile of that distribution.

The quantities C and D are defined in (10), and the scalar E is defined in terms of the components E_i and F_i by

$$E \equiv \sum_i E_i F_i.$$

The sum over i runs from 2 to k if constrained residuals are used, and from 1 to k if unconstrained residuals are used.

Proof: In the Appendix.

Remarks and Corollaries:

The regressor design and the pattern of heteroskedasticity influence (11) through just three quantities, C, D, and E. Indeed, for given n and α, (11) depends only on these three quantities and e_3, e_4, e_3^*, and e_4^*.

For the distribution F_1 of (5), we have $e_3^* = 1$, $e_4^* = 2$. Thus, for F_1, (11) becomes

$$\phi(z_\alpha)n^{-1}\Big(e_3^2C^2(\tfrac{1}{4}z_\alpha + \tfrac{1}{2}z_\alpha^3) + \tfrac{1}{12}e_4D\big(z_\alpha - 5z_\alpha^3\big) + E(\tfrac{1}{2}z_\alpha + z_\alpha^3)\Big), \tag{12}$$

of order n^{-1} at most. For F_2 in (6), $e_3^* = 0$ and $e_4^* = 1$, and (11) becomes

$$\phi(z_\alpha)\Big(\tfrac{1}{6}n^{-1/2}e_3C(1+2z_\alpha^2) + n^{-1}e_3^2C^2(\tfrac{1}{6}z_\alpha - \tfrac{1}{9}z_\alpha^3 - \tfrac{1}{18}z_\alpha^5)\Big). \tag{13}$$

Whereas (12), like (11), depends on C, D, E, e_3, and e_4, (13) depends only on C and e_3. It can, however, have a leading-order term of order $n^{-1/2}$.

With symmetric error terms, $e_3 = 0$, and (11) simplifies to $n^{-1}\phi(z_\alpha)$ times

$$\tfrac{1}{12}e_4 D\big(z_\alpha(3(e_4^* - 1) - 2e_3^*) - z_\alpha^3((e_4^* - 1) + 4e_3^*)\big) + e_3^* E(\tfrac{1}{2}z_\alpha + z_\alpha^3). \quad (14)$$

This expression depends only on D and E for given e_3^* and e_4^*, and it is of order no higher than n^{-1}.

As can be seen from (13), if either e_3 or C is zero, the ERP for the Rademacher distribution F_2 vanishes completely to order n^{-1}, so that the ERP is at most of order $n^{-3/2}$. This is a very satisfactory degree of refinement for the wild bootstrap. It is obtained if the error terms are not skewed ($e_3 = 0$), or if the projected regressor $\boldsymbol{x}$ is not skewed in the metric of the error variances ($C = 0$). This result is the analog for heteroskedastic models of the result in Hall (1992), according to which bootstrap tests on the coefficients of homoskedastic regression models benefit from refinements unless *both* the regressors and the errors are skewed.

If the error terms are homoskedastic, then $E_i = 0$ for $i = 2, \ldots, k$. This follows from the definition in (10) of E_i and the fact that, for $i = 2, \ldots, k$, the vector $\boldsymbol{x}$ is orthogonal to the other columns of $\boldsymbol{X}$. Thus, if, but only if, constrained residuals are used in forming the HCCME, we have $E = 0$.

Although C and E can be zero for certain regression designs and patterns of heteroskedasticity, the quantity D, as is clear from (10), cannot be less than 1. Similarly, the fourth moment e_4 of the standardized error terms cannot be less than 1. Thus the *only* distribution satisfying (4) for which it is possible that the ERP of the bootstrap test vanishes through order n^{-1} is F_2. To see this, note that, for the term proportional to $e_4 D$ in (11) to vanish, it is necessary that $e_3^* = 0$ and $e_4^* = 1$. But it can readily be shown that the only distribution satisfying (4) and these two conditions is F_2. The standard normal distribution, for instance, satisfies $\mathrm{E}(\varepsilon_t) = 0$ $\mathrm{E}(\varepsilon_t^2) = 1$, $\mathrm{E}(\varepsilon_t^3) = 0$, but then has $\mathrm{E}(\varepsilon_t^4) = 3$.

The coefficients of the different terms of order n^{-1} in (11) are subject to a couple of inequalities, which indicate that, in many circumstances, the term proportional to $e_4 D$, if it is not annihilated by use of the F_2 distribution, will be the dominant term. The first of these inequalities is $C^2 \le D$, which follows from the definitions in (10) and the Cauchy-Schwartz inequality. The second is $e_3^2 \le e_4 - 1$; it holds with equality for the F_2 distribution.

The expansion (11) does not go far enough for there to be any effect associated with the choice of HCCME. Such effects will however be clear in the simulation results presented in section 5.

Theorem 2: For the setup of Theorem 1, the ERP of a two-tailed bootstrap test, for which rejection occurs if the absolute value of τ exceeds a critical value calculated as the $1-\alpha$ quantile of the distribution of the absolute value of the bootstrap statistic, has the following Edgeworth expansion:

$$2n^{-1}\phi(z_{\alpha/2})\Big(((e_3^*)^2-1)e_3^2C^2\big(\tfrac{1}{6}z_{\alpha/2}+\tfrac{1}{9}z_{\alpha/2}^3+\tfrac{1}{18}z_{\alpha/2}^5\big) \\ +\tfrac{1}{12}(e_4^*-1)e_4D\big(3z_{\alpha/2}-z_{\alpha/2}^3\big)\Big). \tag{15}$$

For the F_1 distribution, this reduces to

$$2n^{-1}\phi(z_{\alpha/2})e_4D(z_{\alpha/2}^3-3z_{\alpha/2}),$$

and for the F_2 distribution to

$$2n^{-1}\phi(z_{\alpha/2})e_3^2C^2\big(\tfrac{1}{6}z_{\alpha/2}+\tfrac{1}{9}z_{\alpha/2}^3+\tfrac{1}{18}z_{\alpha/2}^5\big).$$

If the error terms are not skewed ($e_3=0$), the ERP becomes

$$2n^{-1}\phi(z_{\alpha/2})\tfrac{1}{12}(e_4^*-1)e_4D\big(3z_{\alpha/2}-z_{\alpha/2}^3\big).$$

Proof: In the Appendix.

Remarks:

When two-tailed tests are used, the ERP of a bootstrap test is usually of lower order than for a one-tailed test. This is the case here only for the case in which the Rademacher F_2 distribution is used for the wild bootstrap, and $e_3C\neq 0$. However, this does mean that, in all cases, the ERP is of order no higher than n^{-1}. As with a one-tailed test, if $e_3C=0$, use of the F_2 distribution causes the ERP to vanish through order n^{-1}.

For tests with more than one degree of freedom, where the statistic is in asymptotically chi-squared form, the rates of convergence are the same as for the two-tailed one degree of freedom test.

In the Appendix, we give the Edgeworth expansion for statistics τ of type (9). It is straightforward to use it to obtain Edgeworth expansions of the ERP of the asymptotic test based on τ. As expected, the leading-order term is of order $n^{-1/2}$ for a one-tailed test, and n^{-1} for a two-tailed one. These are just the same orders as those we have obtained for the wild bootstrap. Remarkably, the contribution of order $n^{-1/2}$ to the ERP of the one-tailed asymptotic test is the *same* as that of the wild bootstrap test based on F_2; the two contributions therefore vanish under the same conditions. The terms of order n^{-1} are, however, more numerous and more complicated for the asymptotic test than for the bootstrap test. As simulation results in section 5 will show, even when the term of order $n^{-1/2}$ in (11) does not vanish, the ERP of the bootstrap test is usually very much smaller than that of the asymptotic test.

Chesher and Jewitt (1987) showed that the ERP of asymptotic tests based on various versions of the HCCME depend strongly on whether high leverage observations are present in the sample. This fact emerges from the Edgeworth expansion for the asymptotic test, but it is conspicuously absent from that for the bootstrap test. Of the three quantities on which the bootstrap expansion depends, C and D are not directly related to leverage at all. The third quantity, E, which does not appear in the expansion for the F_2 version of the wild bootstrap, does depend on the projection matrix on to the space spanned by the regressors, but not exclusively on the diagonal elements of that matrix, which are the conventional measures of leverage. In fact, if $\boldsymbol{P}$ denotes the orthogonal projection on to the span of the columns of $\boldsymbol{X}_2$, if constrained residuals are used, or of $\boldsymbol{X}$ if unconstrained, then, recalling the fact that $\boldsymbol{X}^\top\boldsymbol{X} = n\mathbf{I}$ by construction, it is easy to show that

$$E = n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} x_t\sigma_t^2(\boldsymbol{P})_{ts}x_s^3\sigma_s^2.$$

Our simulation results confirm that, for the F_2 bootstrap, the presence of high leverage observations has little effect on the ERP of the wild bootstrap tests.

4. A Special Case

There is an interesting special case in which the wild bootstrap using F_2 yields almost perfect inference. This case arises when the entire parameter vector $\boldsymbol{\beta}$ vanishes under the null hypothesis and constrained residuals are used for both the HCCME and the wild bootstrap DGP.

Theorem 3: Consider the linear regression model

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta} + u_t \tag{16}$$

where the $n \times k$ matrix $\boldsymbol{X}$ with typical row $\boldsymbol{X}_t$ is independent of all the symmetrically distributed and mutually independent error terms u_t, and where the regressors and error terms satisfy the same regularity conditions as for Theorem 1. Under the null hypothesis that $\boldsymbol{\beta} = \mathbf{0}$, the χ^2 statistic for a test of that null against the alternative represented by (16), based on any of the four versions HC_i, $i = 0, 1, 2, 3$ of the HCCME, constructed with constrained residuals, has the same distribution, conditional on the absolute values $|u_t|$ of the error terms, as the same statistic bootstrapped, if the ε_t of the bootstrap DGP are generated by the Rademacher distribution F_2 of (6).

The bootstrap P value is independent of the $|u_t|$, and, for sample size n, follows a discrete distribution supported by the set of points $p_i = i/2^n$, $i = 0, \ldots, 2^n - 1$, with equal probability mass 2^{-n} on each point.

Proof: We have assumed that the error terms follow a continuous distribution, which therefore can have no atom of positive probability at the origin. Thus, for each error term u_t, we can write $u_t = |u_t|s_t$, where s_t, equal to ± 1, is the sign of the error term. If the u_t are symmetrically distributed, it is easy to see that $|u_t|$ and s_t are independent, and that the s_t are independent drawings from F_2.

The OLS estimates from (16) are given by $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$, and any of the HCCMEs we consider for $\hat{\boldsymbol{\beta}}$ can be written in the form (2), with an appropriate choice of $\hat{\boldsymbol{\Omega}}$. The χ^2 statistic thus takes the form

$$\tau \equiv \boldsymbol{y}^\top\boldsymbol{X}(\boldsymbol{X}^\top\hat{\boldsymbol{\Omega}}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}. \tag{17}$$

Under the null, $\boldsymbol{y} = \boldsymbol{u}$. Define the $1\times k$ row vector $\boldsymbol{Z}_t$ as $|u_t|\boldsymbol{X}_t$, and the $n\times 1$ column vector $\boldsymbol{s}$ with typical element s_t. Then the $n \times k$ matrix $\boldsymbol{Z}$ with typical row $\boldsymbol{Z}_t$ is independent of the vector $\boldsymbol{s}$. If the constrained residuals, which are just the elements of $\boldsymbol{y}$, are used to form $\hat{\boldsymbol{\Omega}}$, the statistic (17) becomes

$$\boldsymbol{s}^\top\boldsymbol{Z}\left(\sum_{t=1}^{n} a_t\boldsymbol{Z}_t^\top\boldsymbol{Z}_t\right)^{-1}\boldsymbol{Z}^\top\boldsymbol{s}, \tag{18}$$

where the a_t are defined as in (8).

If we denote by τ^* the statistic generated by the wild bootstrap with F_2, then τ^* can be written as

$$\boldsymbol{\varepsilon}^\top\boldsymbol{Z}\left(\sum_{t=1}^{n} a_t\boldsymbol{Z}_t^\top\boldsymbol{Z}_t\right)^{-1}\boldsymbol{Z}^\top\boldsymbol{\varepsilon}, \tag{19}$$

where $\boldsymbol{\varepsilon}$ denotes the vector containing the ε_t. The matrix $\boldsymbol{Z}$ is exactly the same as in (18), because the exogenous matrix $\boldsymbol{X}$ is reused unchanged, and the wild bootstrap error terms $u_t^* = \pm u_t$, since, under F_2, $\varepsilon_t = \pm 1$. Thus, for all t, $|u_t^*| = |u_t|$. By construction, $\boldsymbol{\varepsilon}$ and $\boldsymbol{Z}$ are independent under the wild bootstrap DGP. Under the null hypothesis, $\boldsymbol{s}$ and $\boldsymbol{Z}$ are independent, and $\boldsymbol{s}$ follows exactly the same distribution as $\boldsymbol{\varepsilon}$. It follows that τ under the null and τ^* under the wild bootstrap DGP with F_2 have the same distribution conditional on the $|u_t|$. This proves the first assertion of the theorem.

This common conditional distribution of τ and τ^* is of course a discrete distribution, since $\boldsymbol{\varepsilon}$ and $\boldsymbol{s}$ can take on only 2^n different, equally probable, values, with a choice of $+1$ or -1 for each of the n components of the vector. The statistic τ must take on one of the 2^n possible values, each with the same probability of 2^{-n}. If we denote the 2^n values, arranged in increasing order, as τ_i, $i = 1, \ldots, 2^n$, with $\tau_j > \tau_i$ for $j > i$, then, if $\tau = \tau_i$, the bootstrap P value, which is the probability mass in the distribution to the right of τ_i, is just $p_i \equiv 1 - i/2^n$. As i ranges from 1 to 2^n, the P value varies over the set of points p_i, $i = 0, \ldots, 2^n - 1$, all with probability 2^{-n}. This distribution, conditional on the $|u_t|$, does not depend on the $|u_t|$, and so is also the unconditional distribution of the bootstrap P value, which is thus independent of the $|u_t|$. ∎

Remarks: For small enough n, it may be quite feasible to enumerate all the possible values of the bootstrap statistic τ^*, and thus obtain an exact bootstrap P value without simulation.

If $k = 1$, a very minor adaptation of the proof shows that the result of the Theorem applies to the t form of the one degree of freedom statistic.

Although the discrete nature of the bootstrap distribution means that it is not possible to perform exact inference for an arbitrary significance level α, the problem is no different from the problem of inference with any discrete-valued statistic. For the case with $n = 10$, which will be extensively treated in the following section, $2^n = 1024$, and so the bootstrap P value cannot be in error by more than 1 part in a thousand.

If the null hypothesis does not require all the regression parameters to be zero, or if unconstrained residuals are used, the expressions (18) and (19) for τ and τ^* continue to hold if $\boldsymbol{Z}_t$ is redefined as $|\ddot{u}_t|(\boldsymbol{M}_2\boldsymbol{X}_1)_t$, where $\boldsymbol{X}_1$ is the matrix of regressors admitted under the null. However, although $\boldsymbol{\varepsilon}$ in τ^* is by construction independent of $\boldsymbol{Z}$, $\boldsymbol{s}$ in τ is not, and its elements are not mutually independent, because the covariance matrix of the residual vector $\ddot{\boldsymbol{u}}$ is not diagonal in general, unlike that of the error terms $\boldsymbol{u}$.

5. Experimental Design and Simulation Results

Since bootstrap tests usually behave better in large samples than in small, most of our experiments are performed with a sample of size 10, in order to stress-test the wild bootstrap. All the tests we consider are of the null hypothesis that $\beta_1 = 0$ in model (1), with $k_1 = 1$ and k_2, the number of regressors in the matrix $\boldsymbol{X}_2$, varying from 0 to 5 across experiments. The regressor $\boldsymbol{x}_1$ associated with the parameter β_1 has elements which are independent drawings from $N(0, 1)$, except for the second, which is equal to 10, so as to create very substantial skewness in the regression design, and also an observation of very high leverage. For $k_2 \geq 1$, the first column of $\boldsymbol{X}_2$ is a constant. For $k_2 = 2, \ldots, 5$, additional regressors are used which are linear combinations of $\boldsymbol{x}_1$ and independent normal vectors. In Table 1, the components of all the regressors except the constant are given.

The data in all the simulation experiments discussed here are generated under the null hypothesis. Since (1) is a linear model, we set $\boldsymbol{\beta}_2 = \boldsymbol{0}$ without loss of generality. Thus our data are generated by a DGP of the form

$$y_t = \sigma_t v_t, \qquad t = 1, \ldots, n, \tag{20}$$

where n is the sample size, 10 for most experiments. For homoskedastic data, we set $\sigma_t = 1$ for all t, and for heteroskedastic data, we set $\sigma_t = |x_{t1}|$, the absolute value of the t^{th} component of $\boldsymbol{x}_1$. Because of the high leverage observation, this gives rise to very strong heteroskedasticity, which leads to serious bias of the OLS covariance matrix; see White (1980). The v_t are independent mean zero variables of unit variance, and in the experiments will be either normal or else drawings from the highly skewed $\chi^2(2)$ distribution, centred and standardized.

In Table 2, we give, for the regression designs considered and the above pattern of heteroskedasticity, the values of the quantities C, D, and E on which the approximate

ERP (11) depends. The quantity denoted E_0 is the sum over $i = 2, \ldots, k$ ($k = k_1 + k_2$), appropriate if constrained residuals are used in the HCCME, E_1 adds in the term for $i = 1$. For comparison purposes, the values are also given for homoskedastic errors.

The main object of our experiments is to compare the size distortions of wild bootstrap tests using the distributions F_1 and F_2. Although the latter gives exact inference only in a very restricted case, it always leads to less distortion than the former, in the cases we consider, for sample sizes up to 100. We are also interested in the impact on ERPs of the use of unconstrained versus constrained residuals, and the use of the different sorts of HCCME. We consider only one-tailed tests, since these typically lead to greater distortions than two-tailed tests. We present our results as P value discrepancy plots, as described in Davidson and MacKinnon (1998). These plots show ERPs as a function of the nominal level α. They also provide enough information to have a good idea of the ERP of a two-tailed test. All plots are based on experiments using 100,000 replications.

We now present our results as answers to a series of pertinent questions.

- In a representative case, with strong heteroskedasticity and regressor skewness, is the wild bootstrap capable of reducing the ERP relative to asymptotic tests?

Figure 1 shows plots for the regression design with $k = 3$, sample size $n = 10$, and normal heteroskedastic errors. The ERPs are plotted for the conventional t statistic, based on the OLS covariance matrix estimate, and for the four versions of HCCME-based statistics, HC_i, $i = 0, 1, 2, 3$, all using constrained residuals. P values for the asymptotic tests are obtained using Student's t distribution with 7 degrees of freedom. The ERP is also plotted for what will serve as a base case for the wild bootstrap: Constrained residuals are used both for the HCCME and the wild bootstrap DGP, the F_2 distribution is used for the ε_t, and the statistic that is bootstrapped is the HC_3 form. To avoid redundancy, the plots are drawn only for the range $0 \le \alpha \le 0.5$, since, as is clear from (9), all these statistics are symmetrically distributed when the errors are symmetric. In addition, the bootstrap statistics are symmetrically distributed conditional on the original data, and so the distribution of the bootstrap P value is also symmetrical about $\alpha = 0.5$. It follows that the ERP for nominal level α is the negative of that for $1 - \alpha$. Not surprisingly, the conventional t statistic, which does not have even an asymptotic justification, is the worst behaved of all, with far too much mass in the tails. But, although the HC_i statistics are less distorted, the bootstrap test is manifestly much better behaved.

- The design with $k_2 = 0$ ($k = 1$) satisfies the conditions of Theorem 3 when the errors are symmetric and the HCCME and the bootstrap DGP are based on constrained residuals. If $k_2 > 0$, bootstrap inference is no longer perfect, but, according to the Edgeworth expansion, should still be good, since the ERP is of order less than n^{-1}. To what extent is this so? Do the design-dependent quantities given in Table 2 have any predictive power for the ERP?

P value discrepancy plots are shown in Figure 2 for the designs $k = 1, \ldots, 6$ using the base-case wild bootstrap as described above. Errors are normal and heteroskedastic.

As expected, the ERP for $k = 1$ is just experimental noise, and for most other cases the ERPs are significant, but not large, even for the very small sample size. They are particularly small for $k = 2$, and, by what is presumably a coincidence induced by the specific form of the data, for $k = 6$. Such a result might perhaps be predicted on the basis of Table 2, but not in any conclusive fashion: Presumably the remainder of order less than n^{-1} in the Edgeworth expansion depends on the regression design through other things than just C, D, and E.

- How do bootstrap tests based on the F_1 and F_2 distributions compare? We expect that F_2 will lead to smaller ERPs if the errors are symmetric, but what if they are asymmetric? How effective is the skewness correction provided by F_1?

In Figure 3 plots are shown for the $k = 3$ design with heteroskedastic normal errors and skewed $\chi^2(2)$ errors. The F_1 and F_2 bootstraps give rather similar ERPs, whether or not the errors are skewed. But the F_2 bootstrap is generally better, and never worse. Very similar results, leading to same conclusion, were also obtained with the $k = 4$ design. For $k = 1$ and $k = 2$, on the other hand, the F_1 bootstrap suffers from larger ERPs than does F_2.

- What is the penalty for using the wild bootstrap when the errors are homoskedastic and inference based on the conventional t statistic is reliable, at least with normal errors? Do we get different answers for F_1 and F_2?

Again we use the $k = 3$ design. We see from Figure 4, which is like Figure 3 except that the errors are homoskedastic, that, with normal errors, the ERP is very slight with F_2, but remains significant for F_1. Thus, with unskewed, homoskedastic errors, the penalty attached to using the F_2 bootstrap is very small. With skewed errors, all three tests give substantially greater ERPs, but the F_2 version remains a good deal better than the F_1 version.

- Do the rankings of bootstrap procedures obtained so far for $n = 10$ continue to apply for larger samples? Do the ERPs become smaller rapidly as n grows?

In order to deal with larger samples, the data in Table 1 were simply repeated as needed in order to generate regressors for $n = 20, 30, \ldots$. In this way, the design-dependent quantities like C and D do not depend on n. The plots shown in Figures 3 and 4 are repeated in Figure 5 for $n = 100$. The rankings found for $n = 10$ remain unchanged, but, as suggested by the results of Section 3, the ERP for the F_2 bootstrap with skewed, heteroskedastic, errors improves less than that for the F_1 bootstrap with the increase in sample size. It is noteworthy that none of the ERPs in this diagram is very large.

In Figure 6, we plot the ERP for $\alpha = 0.05$ as a function of n, $n = 10, 20, \ldots$, with the $k = 3$ design and heteroskedastic errors, normal for F_1 and skewed for F_2, chosen because these configurations lead to comparable ERPs for n around 100, and because this is the worst setup for the F_2 bootstrap. It is interesting to observe that, at least for $\alpha = 0.05$, the ERPs are not monotonic. What seems clear is that, although the absolute magnitude of the ERPs is not disturbingly great, the rate of convergence to zero does not seem to be at all rapid. As the Edgeworth expansions suggest, it is slower for the F_2 bootstrap. Since C, D, *etc.*, do not vary with n, these results do not support the idea that a power of $n^{-1/2}$ is a good way to measure the rate of convergence.

- How do the ERPs as estimated by simulation compare with the approximations given by Edgeworth expansions?

In some cases, of course, very badly indeed, as when the Edgeworth approximation is zero, but the ERP is significant. Very badly again for the F_1 bootstrap, where, because the term of order $n^{-1/2}$ vanishes, the expression (11) is antisymmetric with respect to z_α, implying an ERP that is antisymmetric about $\alpha = 0.5$. We see from Figure 3 that, although the ERP does have this form for the F_2 bootstrap with symmetric errors, for F_1 the ERP is very far indeed from antisymmetric with skewed errors, being negative for almost all α. For another comparison with no requirement of antisymmetry, we consider the $k = 3$ design with the F_2 bootstrap and $\chi^2(2)$ errors, for which it can easily be seen that $e_3 = 2$ and $e_4 = 9$. In Figure 7a, we plot the order $n^{-1/2}$ and order n^{-1} contributions in (11) as functions of α. For $n = 10$, the order n^{-1} term is clearly quantitatively greater than the order $n^{-1/2}$ term. Then, in Figure 7b, we plot the differences between the approximate ERP (11) and the true one, as estimated by simulation, for both the F_1 and the F_2 bootstraps, for sample sizes $n = 10$ and $n = 100$, with $\chi^2(2)$ errors throughout. For $n = 10$, there is no apparent relation at all between the approximation and the true ERP. For $n = 100$, things are much better, although the discrepancy for F_1 remains quite significant. For F_2, on the other hand, the approximation is nearly perfect.

We now move on to consider some lesser questions, the answers to which justify, at least partially, the choices made in the design of our earlier experiments. We restrict attention to the F_2 bootstrap, since it is clearly the procedure of choice in practice.

- Does it matter which of the four versions of the HCCME is used?

It is clear from Figure 2 that the choice of HC_i has a substantial impact on the ERP of the asymptotic test. Since the HC_0 and HC_1 statistics differ only by a constant multiplicative factor, they yield identical bootstrap P values, as do all versions for $k = 1$ and $k = 2$. For $k = 1$ this is obvious, since the raw statistics are identical, and for $k = 2$, the only regressor other than $\boldsymbol{x}_1$ is the constant, and so h_t does not depend on t. For $k > 2$, significant differences appear, as seen in Figure 9 which treats the $k = 4$ design. HC_3 has the least distortion here, and also for the other designs with $k > 2$. This accounts for our choice of HC_3 in the base case.

- What is the best transformation $f_t(\cdot)$ to use in the definition of the bootstrap DGP? Plausible answers are either the identity transformation, or the same as that used for the HCCME.

No very clear answer to this question emerged from our numerous experiments on this point. A slight tendency in favour of using the HC_3 transformation appears, but this choice does not lead to universally smaller ERPs. However, the quantitative impact of the choice is never very large, and so the HC_3 transformation is used in our base case.

- How is performance affected if the leverage of observation 2 is reduced?

The ERPs of the asymptotic tests are greater with a high leverage observation. The Edgeworth expansions suggest that the same should be true of bootstrap tests only to a limited extent. The presence of a high leverage observation does however appear to have a considerable impact if the HC_0 statistic is used, but only for it and HC_1. In Figure 9, this is demonstrated for $k = 3$, and normal errors, and the effect of leverage is compared with that of heteroskedasticity. The latter is clearly a much more important determinant of the ERP than the former. Similar results are obtained if the null hypothesis concerns a coefficient other than β_1. In that case, the h_t differ more among themselves, since $\boldsymbol{x}_1$ is now used in their calculation, and HC_0 gives more variable results than HC_3, for which the ERPs are similar in magnitude to those for the test of $\beta_1 = 0$.

- How important is it to use constrained residuals?

For Theorem 3 to hold, it is essential, and Theorem 1 shows that an extra term is introduced into the expansion of the ERP if unconstrained residuals are used. This term can be obtained from the numerical values in the last two rows of Table 2, where the difference between E_1 and E_0 is the extra term for unconstrained residuals. Simulation results show that, except for the $k = 1$ and $k = 2$ designs, it is not very important whether one uses constrained or unconstrained residuals, although results with constrained residuals tend to be better in most cases. The simulations do however show clearly that it is a mistake to mix unconstrained residuals in the HCCME and constrained residuals for the bootstrap DGP.

6. Conclusion

The wild bootstrap is commonly applied to models with heteroskedastic error terms and an unknown pattern of heteroskedasticity, most commonly in the form that uses the asymmetric F_1 distribution in order to take account of possible skewness of the error terms. In this paper we have shown that the wild bootstrap implemented with the symmetric F_2 distribution and constrained residuals, which can give perfect inference in one very restricted case, is no worse behaved than the F_1 version, or either version with unconstrained residuals, in the rather extreme cases we investigate by simulation, and is usually markedly better. We therefore recommend that this version of the wild bootstrap should always be used in practice in preference to other versions. If it should turn out that it has a higher ERP than the F_1 version, it is very likely that, in such cases, the ERPs of both versions will be small. Our recommendation is supported by the results of simulation experiments designed to expose potential weaknesses of both versions, and by the approximate expressions of the ERP, based on Edgeworth expansions, for both versions. The approximations make clear that the leading negative power of the sample size is by no means the only useful index of a test's performance.

It is important to note that conventional confidence intervals cannot benefit from our recommended version of the wild bootstrap, since they are implicitly based on a Wald test using unconstrained residuals for the HCCME and, unless special precautions are taken, also for the bootstrap DGP. If reliable confidence intervals are essential, we recommend that they be obtained by inverting a set of tests based on the preferred wild bootstrap. Although this can be a computationally intensive procedure, it is well within the capacity of modern computers and seems to be the only way currently known to extend refinements available for tests to confidence intervals.

A final caveat seems called for: Although our experiments cover a good number of cases, some caution is still necessary on account of the fact that the extent of the ERP of wild bootstrap tests appears to be very sensitive to details of the regression design and the pattern of heteroskedasticity.

In this paper, we have tried to investigate worst case scenarios for wild bootstrap tests. This should not lead readers to conclude that the wild bootstrap is an unreliable method in practice. On the contrary, as Figure 6 makes clear, it suffers from very little distortion for samples of moderate size unless there is extreme heteroskedasticity. In most practical contexts, use of the F_2-based wild bootstrap with constrained residuals should provide satisfactory inference.

References

Abramowitz, M., and I. A. Stegun (1965). *Handbook of Mathematical Functions*, New York, Dover.

Beran, R. (1986). Discussion of "Jackknife bootstrap and other resampling methods in regression analysis" by C. F. J. Wu., *Annals of Statistics* 14, 1295–1298.

Beran, R. (1988). "Prepivoting test statistics: a bootstrap view of asymptotic refinements," *Journal of the American Statistical Association*, 83, 687–697.

Chesher A., and I. Jewitt (1987). "The bias of a heteroskedasticity consistent covariance matrix estimator," *Econometrica*, 55, 1217–1222.

Davidson, R. and J. G. MacKinnon (1998). "Graphical methods for investigating the size and power of hypothesis tests," *The Manchester School*, **66**, 1–26.

Eicker, F. (1963). "Asymptotic normality and consistency of the least squares estimators for families of linear regressions," *The Annals of Mathematical Statistics*, 34, 447–456.

Hall, P. (1988). "Theoretical Comparison of Bootstrap Confidence Intervals," *Annals of Statistics*, 16, 927–953.

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

Kendall, M., and A. Stuart (1977). *The Advanced Theory of Statistics*, 4th ed., Vol. 1, "Distribution Theory," London, Griffin.

Liu, R. Y. (1988). "Bootstrap procedures under some non-I.I.D. models," *Annals of Statistics* 16, 1696–1708.

MacKinnon, J. G., and H. White (1985). "Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties," *Journal of Econometrics*, 29, 305–325.

Mammen, E. (1993). "Bootstrap and wild bootstrap for high dimensional linear models," *Annals of Statistics* 21, 255–285.

McCullagh, P. (1987). "Tensor Methods in Statistics," London, Chapman and Hall.

White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, 48, 817–838.

Wu, C. F. J. (1986). "Jackknife bootstrap and other resampling methods in regression analysis," *Annals of Statistics* 14, 1261–1295.

Appendix

Proof of Theorems 1 and 2:

Many parts of the proof apply to both theorems, and so we distinguish only when necessary. The proof is divided into several steps. First, we develop a formal stochastic expansion through order n^{-1} for any of the HCCME-based t statistics we consider in the paper, and show how the same expansion applies as well to the bootstrapped statistic by simply redefining certain quantities. Next, approximate expressions are obtained for the low order cumulants of these statistics on the basis of the stochastic expansion. These approximate cumulants are used to obtain the (formal) Edgeworth expansion of the distributions of the statistics. In the next step, the Edgeworth expansion for a bootstrap statistic is inverted to yield the Cornish-Fisher expansion of its α-quantile. This quantile is needed in order to express the condition that the statistic τ is less than the α-quantile of the bootstrap distribution, the probability of which is the rejection probability of the bootstrap test at nominal level α. The quantile is of course different for one- and two-tailed tests; for the latter it is an absolute value that is required. Then, because the quantile of the bootstrap distribution is random, the condition for rejection is rearranged so as to put all random terms on the left-hand side. This gives rise to another random variable, the distribution of which is described by an Edgeworth expansion obtained in the next step by applying some easily computed perturbations to that for the basic statistic τ. Finally, the approximate bootstrap rejection probability, and hence also the ERP, is found by evaluating the Edgeworth expansion at the desired nominal level α.

Step 1: stochastic expansion of the statistic.

In (9), it was seen that, under the null hypothesis, any of the statistics we consider can be written as

$$\tau = \sum_{t=1}^{n} (\boldsymbol{M}_2\boldsymbol{x}_1)_t u_t \Big/ \Big(\sum_{t=1}^{n} a_t (\boldsymbol{M}_2\boldsymbol{x}_1)_t^2 \ddot{u}_t^2\Big)^{1/2}, \tag{21}$$

with appropriate choice of the a_t and the residuals (constrained or unconstrained) $\ddot{u}_t$. In the notation of the statement of the theorem, $(\boldsymbol{M}_2\boldsymbol{x}_1)_t = x_t$. If constrained residuals are used, the vector $\tilde{\boldsymbol{u}}$ with typical element $\tilde{u}_t$ is

$$\boldsymbol{M}_2\boldsymbol{u} = \boldsymbol{u} - \boldsymbol{X}_2(\boldsymbol{X}_2^\top\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{u} = \boldsymbol{u} - n^{-1}\boldsymbol{X}_2\boldsymbol{X}_2^\top\boldsymbol{u},$$

since we define $\boldsymbol{X}_2$ such that $\boldsymbol{X}_2^\top\boldsymbol{X}_2 = n\mathbf{I}$. In terms of the IID variables v_t, we find that

$$\tilde{u}_t = \sigma_t v_t - n^{-1}\sum_{i=2}^{k} X_{ti}\sum_{s=1}^{n} X_{si}\sigma_s v_s. \tag{22}$$

If unconstrained residuals $\hat{u}_t$ are used, (22) is modified so that the sum over i runs from 1 to k, rather than from 2 to k. To avoid having to distinguish the two cases, we write just $\sum_i$ in subsequent expressions, and maintain the ambiguous notation $\ddot{u}_t$ for the residuals.

Let us make the definition

$$w_i \equiv n^{-1/2} S^{-1} \sum_{t=1}^{n} X_{ti} \sigma_t v_t, \quad i = 1, \ldots, k; \tag{23}$$

recall the definition of S in (10). By the central limit theorem, the w_i are asymptotically normal, mean zero, and $O(1)$. The numerator of the statistic (21), divided by $n^{1/2}$, is just Sw_0. For the denominator, we need the stochastic expansion of the residuals, which from (22) and (23) is

$$\ddot{u}_t = \sigma_t v_t - n^{-1/2} S \sum_i X_{ti} w_i.$$

Make the following definitions, for $i, j = 1, \ldots, k$:

$$w_{ai} \equiv n^{-1/2} S^{-1} \sum_{t=1}^{n} a_t x_t^2 X_{ti} \sigma_t v_t, \quad q \equiv n^{-1/2} S^{-2} \sum_{t=1}^{n} a_t x_t^2 \sigma_t^2 (v_t^2 - 1),$$

$$A_{ij} \equiv n^{-1} \sum_{t=1}^{n} a_t x_t^2 X_{ti} X_{tj}, \text{ and } H \equiv \sum_{t=1}^{n} (a_t - 1) x_t^2 \sigma_t^2 / (n^{-1} \sum_t x_t^2 \sigma_t^2).$$

Clearly the w_{ai} and q are asymptotically normal, mean zero, and $O(1)$, while the A_{ij} and H are deterministic and $O(1)$. (Recall that $a_t - 1 = O(n^{-1})$.) Then the denominator of (21), also divided by $n^{1/2}$, is S times

$$\big(1 + n^{-1/2} q + n^{-1} H - 2n^{-1} \sum_i w_i w_{ai} + n^{-1} \sum_i \sum_j A_{ij} w_i w_j\big)^{1/2}.$$

With this, we can formulate the stochastic expansion of τ through order n^{-1}:

$$\tau_1 \overset{a}{=} w_0 \Big(1 - \tfrac{1}{2} n^{-1/2} q + n^{-1} \sum_i w_i w_{ai} - \tfrac{1}{2} n^{-1} H - \tfrac{1}{2} n^{-1} \sum_i \sum_j A_{ij} w_i w_j + \tfrac{3}{8} n^{-1} q^2 \Big). \tag{24}$$

A wild bootstrap statistic is defined by the same formula (21) as τ itself, but the error terms u_t are replaced by the wild bootstrap error terms u_t^*. If we write $u_t^* = s_t \varepsilon_t$, then

$$s_t = a_t^{1/2} (\sigma_t v_t - n^{-1/2} S \sum_i X_{ti} w_i). \tag{25}$$

Since the bootstrap DGP generates data conditional on the realised v_t, the only random elements in a bootstrap sample are the ε_t, just as the random elements in a drawing from the true DGP are the v_t. Thus the factorisation $u_t^* = s_t \varepsilon_t$ plays exactly the same role for the bootstrap DGP as the factorisation $u_t = \sigma_t v_t$ does for the true DGP. It follows that the stochastic expansion of the bootstrap statistic (conditional on the v_t) is given by (24), with all the variables redefined with s_t and ε_t in place of σ_t and v_t respectively.

Step 2: Formal Edgeworth expansion based on cumulants.

The Edgeworth expansion of the CDF F of an asymptotically $N(0,1)$ statistic τ can be written as

$$F(x) = \Phi(x) - n^{-1/2}\phi(x)\sum_{i=1}^{\infty}\lambda_i He_{i-1}(x). \tag{26}$$

Here $\Phi(\cdot)$ and $\phi(\cdot)$ are respectively the CDF and the density of the $N(0,1)$ distribution, and $He_i(\cdot)$ is the Hermite polynomial of degree i (see for instance Abramowitz and Stegun (1965), Chapter 22 for details of these polynomials). The expansion as written in (26) is more properly referred to as the Gram-Charlier series, but, unless truncated, the Edgeworth and Gram-Charlier series are equivalent. In this proof, we truncate everything of order lower than n^{-1}, and so we obtain true Edgeworth series. The λ_i in (26) are coefficients that are at most of order unity, defined by the relations

$$\lambda_j = \frac{n^{1/2}}{j!}\mathrm{E}\big(He_j(\tau)\big), \tag{27}$$

so that, for the first few values of j, $\lambda_1 = n^{1/2}\mu_1$, $\lambda_2 = n^{1/2}(\mu_2 - 1)/2$, $\lambda_3 = n^{1/2}(\mu_3 - 3\mu_1)/6$, $\lambda_4 = n^{1/2}(\mu_4 - 6\mu_2 + 3)/24$, *etc*, where μ_i is the uncentred moment of τ of order i.

The leading-order term of the stochastic expansion τ_1 in (24) is w_0, which is a normalized sum of mean-zero variables that converges to the $N(0,1)$ distribution as $n \to \infty$ by the central limit theorem. Under the regularity conditions of all the theorems in this paper, the cumulant of w_0 of order j, for $j > 2$, is of order $n^{-(j-2)/2}$ – see Chapter 5 of McCullagh (1987) for many more details on cumulants as applied to Edgeworth expansions. It also follows from the theory of that chapter that $\mathrm{E}(He_j(w_0))$ is equal to the "formal moment" of order j corresponding to a sequence of "formal cumulants," κ_j, say, where κ_j is the j^{th} cumulant of w_0, except for $j = 2$, for which κ_2 is the second cumulant of w_0 minus 1.

The first-order cumulant of w_0 is its expectation, which is zero. The second order cumulant is the variance, which is unity, and so the formal cumulant of second order is also zero. With zero mean and unit variance, the third and fourth cumulants, which are also the formal cumulants, are respectively the third moment and the fourth moment minus 3, that is, $n^{-1/2}e_3C$ and $n^{-1}(e_4D - 3)$. It can be seen to follow from this (see McCullagh (1987) again) that the formal moments associated with these formal cumulants of order higher than 4 are all of order lower than n^{-1}, except the sixth, which is 10 times the square of the third formal cumulant. Further discussion of these points can also be found in Kendall and Stuart (1977), Chapter 6.

For ease of notation, write $\tau_1 = w_0 + n^{-1/2}\xi$, where

$$\xi \equiv -\tfrac{1}{2}w_0 q + n^{-1/2}w_0\Big(\sum_i w_i w_{ai} - \tfrac{1}{2}\sum_i\sum_j A_{ij}w_i w_j - \tfrac{1}{2}H + \tfrac{3}{8}q^2\Big). \tag{28}$$

For $j = 1$, we find from (27) that $\lambda_1 = n^{1/2}\mathrm{E}\big(w_0 + n^{-1/2}\xi\big) = \mathrm{E}(\xi)$. To compute $\mathrm{E}(\xi)$, note that

$$\mathrm{E}(w_0 q) = e_3 (nS^3)^{-1} \sum_{t=1}^{n} a_t x_t^3 \sigma_t^3 = e_3 C + O(n^{-1})$$

The expectations of the terms in ξ of order $n^{-1/2}$ all involve a product of three random variables, and hence implicitly a triple sum over the observations. They are thus of order $n^{-1/2}$, because only the terms for which all three observation indices coincide have a nonzero expectation, and the resulting sum over n terms is multiplied by a factor of $n^{-3/2}$. Thus $\lambda_1 = -\frac{1}{2} e_3 C + O(n^{-1})$.

In order to compute λ_2, we note that

$$\begin{aligned}\tau_1^2 &= w_0^2 + 2n^{-1/2} w_0 \xi + n^{-1}\xi^2 \\ &= w_0^2 - n^{-1/2} w_0^2 q + n^{-1} w_0^2 \Big(q^2 + 2\sum_i w_i w_{ai} - \sum_i \sum_j A_{ij} w_i w_j - H\Big)\end{aligned}$$

Now we have

$$\begin{gathered}\mathrm{E}(w_0^2 q) = n^{-1/2} D(e_4 - 1), \quad \mathrm{E}(w_0^2 w_i w_j) = B_{ij} + 2E_i E_j, \\ \mathrm{E}(w_0^2 q^2) = D(e_4 - 1) + 2e_3^2 C^2, \quad \mathrm{E}(w_0^2 w_i w_{ai}) = G_i + 2E_i F_i,\end{gathered}$$

where we have implicitly defined the following deterministic, order 1, quantities

$$B_{ij} = S^{-2} n^{-1} \sum_{t=1}^{n} X_{ti} X_{tj} \sigma_t^2 \quad \text{and} \quad G_i = S^{-2} n^{-1} \sum_{t=1}^{n} a_t x_t^2 X_{ti}^2 \sigma_t^2.$$

Hence, to order n^{-1}, we obtain

$$\begin{aligned}\lambda_2 = \tfrac{1}{2} n^{1/2} \mathrm{E}(\tau_1^2 - 1) = n^{-1/2} \Big(& e_3^2 C^2 + \sum_i (G_i + 2E_i F_i) \\ & - \tfrac{1}{2} \sum_i \sum_j A_{ij} (B_{ij} + 2E_i E_j) - \tfrac{1}{2} H\Big) + O(n^{-1}).\end{aligned}$$

Similar calculations, of which we skip the details, show that $\lambda_3 = -\frac{1}{3} e_3 C + O(n^{-1})$, and

$$\lambda_4 = n^{-1/2} \Big(-\tfrac{1}{12} e_4 D + \tfrac{2}{3} e_3^2 C^2 + \sum_i E_i F_i - \tfrac{1}{2} \sum_i \sum_j A_{ij} E_i E_j\Big) + O(n^{-1}).$$

Finally, λ_6 is $n^{1/2}/720$ times the sixth formal moment, which was seen to be 10 times the square of the third formal cumulant. The third formal cumulant is $6n^{-1/2}\lambda_3 = -2n^{-1/2} e_3 C$, and so $\lambda_6 = \frac{1}{18} n^{-1/2} e_3^2 C^2$.

For the wild bootstrap statistic, we define coefficients λ_j^* by the same formulas as those for the λ_j, but with s_t instead of σ_t and ε_t instead of v_t in the definitions. As with the λ_j, we need work only through order $n^{-1/2}$. Using a star systematically to denote a quantity defined for the bootstrap distribution, we find, using the definition (25) of s_t, that

$$
\begin{aligned}
(S^*)^2 \equiv n^{-1}\sum_{t=1}^n x_t^2 s_t^2 &= n^{-1}\sum_{t=1}^n a_t x_t^2(\sigma_t^2 v_t^2 - 2n^{-1/2} S\sigma_t v_t \sum_i X_{ti} w_i) + O(n^{-1}) \\
&= S^2(1 + n^{-1/2} q) + O(n^{-1}),
\end{aligned}
$$

since $n^{-1}\sum_i w_i n^{-1/2}\sum_t x_t^2 X_{ti}\sigma_t v_t = O(n^{-1})$. Then through order $n^{-1/2}$,

$$
C^* = (S^*)^{-3} n^{-1}\sum_{t=1}^n x_t^3 s_t^3 = (S^*)^{-3} n^{-1}\sum_{t=1}^n x_t^3(\sigma_t^3 v_t^3 - 3n^{-1/2} S\sigma_t^2 v_t^2 \sum_i X_{ti} w_i).
$$

Define the asymptotically normal, variable $c \equiv n^{-1/2} S^{-3}\sum_{t=1}^n x_t^3\sigma_t^3(v_t^3 - e_3)$, of order 1 and mean 0. Then

$$
\begin{aligned}
C^* &= (S^*)^{-3} S^3(e_3 C + n^{-1/2} c - 3n^{-1/2}\sum_i w_i F_i) + O(n^{-1}) \\
&= e_3 C + n^{-1/2}(-\tfrac{3}{2} q e_3 C + c - 3\sum_i w_i F_i) + O(n^{-1}).
\end{aligned}
$$

From this, we see that, through order $n^{-1/2}$

$$
\begin{aligned}
&\lambda_1^* = -\tfrac{1}{2} e_3^* C^* = -\tfrac{1}{2} e_3^* e_3 C + n^{-1/2} e_3^*(\tfrac{3}{4} q e_3 C - \tfrac{1}{2} c + \tfrac{3}{2}\sum_i w_i F_i), \\
&\lambda_3^* = \tfrac{2}{3}\lambda_1^*, \quad \text{and} \quad \lambda_6^* = \tfrac{1}{18} n^{-1/2}(e_3^*)^2 e_3^2 C^2.
\end{aligned} \tag{29}
$$

Since λ_2^* and λ_4^* are of order only $n^{-1/2}$, we do not need the quantities on which they depend past leading order. Thus, to order 1,

$$
D^* = (S^*)^{-4} n^{-1}\sum_{t=1}^n x_t^4 s_t^4 = S^{-4} n^{-1}\sum_{t=1}^n x_t^4\sigma_t^4 v_t^4 = e_4 D,
$$

and one easily checks that to that order, and for $i, j = 1, \ldots, k$, $B_{ij}^* = B_{ij}$, $E_i^* = E_i$, $F_i^* = F_i$, and $G_i^* = G_i$. Since A_{ij} and H depend neither on the σ_t nor on the v_t, they are the same for the true and the bootstrap DGP. We thus see that, to order $n^{-1/2}$,

$$
\begin{aligned}
\lambda_2^* &= n^{-1/2}\Big((e_3^*)^2 e_3^2 C^2 + \sum_i (G_i + 2E_i F_i) - \tfrac{1}{2}\sum_i\sum_j A_{ij}(B_{ij} + 2E_i E_j) - \tfrac{1}{2} H\Big), \\
\lambda_4^* &= n^{-1/2}\Big(-\tfrac{1}{12} e_4^* e_4 D + \tfrac{2}{3}(e_3^*)^2 e_3^2 C^2 + \sum_i E_i F_i - \tfrac{1}{2}\sum_i\sum_j A_{ij} E_i E_j\Big).
\end{aligned}
$$

Step 3: Determination of the quantile of the bootstrap distribution.

For the CDF (26), the α-quantile is defined implicitly by the equation $F(x_\alpha) = \alpha$. An expansion of x_α in powers of $n^{-1/2}$, usually called the Cornish-Fisher expansion, can be obtained by inverting the definition (26) of F. The result is

$$x_\alpha = z_\alpha + n^{-1/2} \sum_i \lambda_i He_{i-1}(z_\alpha) + \tfrac{1}{2} n^{-1} \sum_i \sum_j \lambda_i \lambda_j h_{ij}(z_\alpha) + O(n^{-3/2}), \qquad (30)$$

where the polynomials $h_{ij}(z)$ can be defined in terms of the Hermite polynomials. See Kendall and Stuart (1977), Chapter 6, for details of this sort of expansion.

Let us denote by $Q_\alpha(\{\lambda_i\})$ the quantile (30) for the sequence $\{\lambda_i\} \equiv \{\lambda_1, \lambda_2, \ldots\}$. Let $l_i \equiv \lambda_i^* - \lambda_i$. For a one-tailed bootstrap test at nominal level α, with rejection in the left-hand tail of the distribution, the event that corresponds to rejection is $\tau < Q_\alpha(\{\lambda_i^*\}) = Q_\alpha(\{\lambda_i + l_i\})$. If we write $Q_\alpha(\{\lambda_i + l_i\}) = Q_\alpha(\{\lambda_i\}) + n^{-1/2} q_\alpha^*$, then

$$q_\alpha^* = \sum_i l_i He_{i-1}(z_\alpha) + \tfrac{1}{2} n^{-1/2} \sum_i \sum_j (l_i \lambda_j + \lambda_i l_j) h_{ij}(z_\alpha) + O(n^{-1}), \qquad (31)$$

Write $Q_\alpha = Q_\alpha(\{\lambda_i\})$, $\nu_i = \mathrm{E}(l_i)$, and let $q_\alpha = \mathrm{E}(q_\alpha^*)$. We have

$$q_\alpha = \sum_i \nu_i He_{i-1}(z_\alpha) + \tfrac{1}{2} n^{-1/2} \sum_i \sum_j (\nu_i \lambda_j + \lambda_i \nu_j) h_{ij}(z_\alpha) + O(n^{-1}). \qquad (32)$$

Comparison with (31) shows that, through order n^{-1}, $Q_\alpha + n^{-1/2} q_\alpha$ is the α-quantile of the distribution characterised by the sequence $\{\lambda_i + \nu_i\}$. Finally, let $\gamma_\alpha = q_\alpha^* - q_\alpha$. Rejection by the bootstrap test is the event $\tau < Q_\alpha + n^{-1/2} q_\alpha^*$, or, equivalently, $\tau - n^{-1/2}\gamma_\alpha < Q_\alpha + n^{-1/2} q_\alpha$, in which all random terms are on the left-hand side of the inequality.

Suppose that the distribution of the random variable $\tau - n^{-1/2}\gamma_\alpha$ is given by an expansion of the form (26) with a sequence of coefficients $\{\lambda_i + \eta_i\}$. Then it follows that the rejection probability of the bootstrap test is given by the expansion

$$\Phi(Q_\alpha + n^{-1/2} q_\alpha) - n^{-1/2} \phi(Q_\alpha + n^{-1/2} q_\alpha) \sum_i (\lambda_i + \eta_i) He_{i-1}(Q_\alpha + n^{-1/2} q_\alpha).$$

Since $Q_\alpha + n^{-1/2} q_\alpha$ is the α-quantile of the distribution characterised by $\{\lambda_i + \nu_i\}$, we have

$$\Phi(Q_\alpha + n^{-1/2} q_\alpha) - n^{-1/2} \phi(Q_\alpha + n^{-1/2} q_\alpha) \sum_i (\lambda_i + \nu_i) He_{i-1}(Q_\alpha + n^{-1/2} q_\alpha) = \alpha,$$

and so, on subtraction, we find that the RPE of the bootstrap test is

$$n^{-1/2} \phi(Q_\alpha + n^{-1/2} q_\alpha) \sum_i (\nu_i - \eta_i) He_{i-1}(Q_\alpha + n^{-1/2} q_\alpha). \qquad (33)$$

For a two-tailed test, the critical value c_α for a test at nominal level α is defined implicitly in terms of the CDF F of the statistic by the equation $\alpha = F(-c_\alpha) + 1 - F(c_\alpha)$. If F is given by the expansion (26), this equation becomes

$$\alpha/2 = \Phi(-c_\alpha) - n^{-1/2}\phi(c_\alpha)\sum_i{}'\lambda_i He_{i-1}(-c_\alpha), \tag{34}$$

where the notation $\sum_i'$ means that the sum is over even values of i only. The result (34) follows from the fact that the Hermite polynomials of even degree are even, and of odd degree odd. The right-hand side of (34) is of the form (26), and so the Cornish-Fisher expansion of $-c_\alpha$ for the wild bootstrap distribution can be written as $Q_{\alpha/2}(\{\lambda_i + l_i\}')$, where the prime means that the elements for odd i are zero.

Through order n^{-1}, it can be seen that $Q_{\alpha/2}(\{\lambda_i + l_i\}')$ is nonstochastic, since for i even, the λ_i^* are nonstochastic through order $n^{-1/2}$. This implies that for i even, $l_i = \nu_i$. With this simplification, it follows that the rejection probability of the two-tailed bootstrap test can be computed directly as the sum of the probabilities that $\tau < Q_{\alpha/2}(\{\lambda_i + \nu_i\}')$ and $\tau > -Q_{\alpha/2}(\{\lambda_i + \nu_i\}')$. A calculation just like that leading to (33), but simpler, shows that the ERP of the two-tailed bootstrap test is, through order n^{-1},

$$2n^{-1/2}\phi(z_{\alpha/2})\sum_i{}'\nu_i He_{i-1}(z_{\alpha/2}), \tag{35}$$

independently of the η_i.

Step 4: Computation of the bootstrap ERP.

We begin with the computation of the ν_i. From (29), we see that, through order $n^{-1/2}$,

$$\nu_1 = \mathrm{E}(\lambda_1^* - \lambda_1) = -\tfrac{1}{2}(e_3^* - 1)e_3 C,$$

since $\mathrm{E}(q) = \mathrm{E}(c) = \mathrm{E}(w_i) = 0$. Similarly, through order $n^{-1/2}$,

$$\begin{aligned} \nu_3 &= -\tfrac{1}{3}(e_3^* - 1)e_3 C, & \nu_6 &= \tfrac{1}{18}n^{-1/2}((e_3^*)^2 - 1)e_3^2 C^2, \\ \nu_2 &= n^{-1/2}((e_3^*)^2 - 1)e_3^2 C^2, & \nu_4 &= n^{-1/2}\big(-\tfrac{1}{12}(e_4^* - 1)e_4 D + \tfrac{2}{3}((e_3^*)^2 - 1)e_3^2 C^2\big). \end{aligned}$$

We saw above that, through order $n^{-1/2}$, $l_2 = \nu_2$, $l_4 = \nu_4$, and $l_6 = \nu_6$. Thus, from (31) and (32), we have that

$$\begin{aligned} \gamma_\alpha = &q_\alpha^* - q_\alpha = l_1 - \nu_1 + (l_3 - \nu_3)(z_\alpha^2 - 1) \\ &+ n^{-1/2}(l_1 - \nu_1)\sum_j \lambda_j h_{1j}(z_\alpha) + n^{-1/2}(l_3 - \nu_3)\sum_j \lambda_j h_{3j}(z_\alpha) + O(n^{-1}). \end{aligned} \tag{36}$$

Define the zero mean random variable

$$\zeta \equiv n^{1/2}\tfrac{1}{3}(l_1 - \nu_1) = n^{1/2}\tfrac{1}{2}(l_3 - \nu_3) = e_3^*(\tfrac{1}{4}qe_3C - \tfrac{1}{6}c + \tfrac{1}{2}\sum_i w_i F_i).$$

We see from this that we need only the first two terms in the expression (36) for γ_α, since the last two are $O(n^{-1})$. To the desired order,

$$\gamma_\alpha = n^{-1/2}\zeta\big(3 + 2(z_\alpha^2 - 1)\big) = n^{-1/2}\zeta(1 + 2z_\alpha^2).$$

For the η_i, we note first that $\lambda_1 + \eta_1 = n^{1/2}\mathrm{E}(\tau - n^{-1/2}\gamma_\alpha) = n^{1/2}\mathrm{E}(\tau) = \lambda_1$, so that $\eta_1 = 0$. For η_2, since $\gamma_\alpha = O(n^{-1/2})$, we have

$$\lambda_2 + \eta_2 = \tfrac{1}{2}n^{1/2}\mathrm{E}\big((\tau - n^{-1/2}\gamma_\alpha)^2 - 1\big) = \lambda_2 - \mathrm{E}(\tau\gamma_\alpha) + O(n^{-3/2}).$$

Now, since $\tau = w_0 + O(n^{-1/2})$,

$$\begin{aligned}\eta_2 &= -\mathrm{E}(\tau\gamma_\alpha) = -\mathrm{E}(w_0\gamma_\alpha) + O(n^{-1}) = -n^{-1/2}(1 + 2z_\alpha^2)\mathrm{E}(w_0\zeta) + O(n^{-1})\\ &= -n^{-1/2}(1 + 2z_\alpha^2)e_3^*\big(\tfrac{1}{4}e_3^2C^2 - \tfrac{1}{6}e_4D + \tfrac{1}{2}\sum_i E_iF_i\big) + O(n^{-1}),\end{aligned}$$

since, as we have already seen, $\mathrm{E}(w_0q) = e_3C$, and, as can easily be checked, $\mathrm{E}(w_0c) = e_4D$ and $\mathrm{E}(w_0w_i) = E_i$. For η_3, we compute

$$\begin{aligned}\lambda_3 + \eta_3 &= \tfrac{1}{6}n^{1/2}\mathrm{E}\big((\tau - n^{-1/2}\gamma_\alpha)^3 - 3(\tau - n^{-1/2}\gamma_\alpha)\big)\\ &= \lambda_3 - \tfrac{1}{2}n^{-1/2}(1 + 2z_\alpha^2)\mathrm{E}(w_0^2\zeta) + O(n^{-1}).\end{aligned}$$

But $\mathrm{E}(w_0^2q)$, $\mathrm{E}(w_0^2c)$, and $\mathrm{E}(w_0^2w_i)$ are all $O(n^{-1/2})$, and so $\eta_3 = O(n^{-1})$. For η_4, we find that

$$\begin{aligned}\lambda_4 + \eta_4 &= \tfrac{1}{24}n^{1/2}\mathrm{E}\big((\tau - n^{-1/2}\gamma_\alpha)^4 - 6(\tau - n^{-1/2}\gamma_\alpha)^2 + 3\big)\\ &= \lambda_4 - \tfrac{1}{24}n^{-1/2}(1 + 2z_\alpha^2)\big(4\mathrm{E}(w_0^3\zeta) - 12\mathrm{E}(w_0\zeta)\big) + O(n^{-1})\end{aligned}$$

Now it can be checked that $\mathrm{E}(w_0^3\zeta) = 3\mathrm{E}(w_0\zeta)$, since $\mathrm{E}(w_0^3q) = 3e_3C$, with similar results for $\mathrm{E}(w_0^3c)$ and $\mathrm{E}(w_0^3w_i)$. Thus $\eta_4 = O(n^{-1})$. Since in general λ_6 is through order n^{-1} a function of λ_3, and since $\eta_3 = O(n^{-1})$, it follows that $\eta_6 = O(n^{-1})$.

We now return to the evaluation of (33). Since to relevant order $Q_\alpha + n^{-1/2}q_\alpha$ is the α-quantile of the distribution characterised by $\{\lambda_i + \nu_i\}$, we obtain from (30) that

$$Q_\alpha + n^{-1/2}q_\alpha = z_\alpha + n^{-1/2}\sum_i(\lambda_i + \nu_i)He_{i-1}(z_\alpha) + O(n^{-1}).$$

Performing a Taylor expansion of (33) about z_α gives for the ERP of the bootstrap test

$$\begin{aligned}&n^{-1/2}\phi(z_\alpha)\sum_i(\nu_i - \eta_i)He_{i-1}(z_\alpha)\\ &-n^{-1}\phi(z_\alpha)\sum_j(\lambda_j + \nu_j)He_{j-1}(z_\alpha)\sum_i(\nu_i - \eta_i)He_i(z_\alpha) + o(n^{-1}),\end{aligned} \tag{37}$$

since the derivative of $\phi(z)He_i(z)$ is $-\phi(z)He_{i+1}(z)$. Of the $\lambda_i + \nu_i$, only those for $i = 1$ and $i = 3$ are $O(1)$. Further, to leading order, $2(\lambda_1+\nu_1) = 3(\lambda_3+\nu_3) = -e_3^*e_3C$. Thus, to leading order,

$$\sum_j (\lambda_j + \nu_j)He_{j-1}(z_\alpha) = -e_3^*e_3C\bigl(\tfrac{1}{2} + \tfrac{1}{3}(z_\alpha^2 - 1)\bigr) = -\tfrac{1}{6}e_3^*e_3C(1 + 2z_\alpha^2).$$

With this, the ERP (37) is

$$n^{-1/2}\phi(z_\alpha)\sum_i (\nu_i - \eta_i)\bigl(He_{i-1}(z_\alpha) + \tfrac{1}{6}n^{-1/2}e_3^*e_3C(1 + 2z_\alpha^2)He_i(z_\alpha)\bigr). \qquad (38)$$

If we first concentrate on the contribution of order $n^{-1/2}$ to this ERP, we see that this contribution comes only from the terms with $i = 1, 3$, and it is

$$-\tfrac{1}{6}n^{-1/2}\phi(z_\alpha)(e_3^* - 1)e_3C(1 + 2z_\alpha^2),$$

in accord with (11). The contribution of order n^{-1} from these same two terms is

$$\tfrac{1}{36}n^{-1}\phi(z_\alpha)((e_3^*)^2 - e_3^*)e_3^2C^2(3z_\alpha + 4z_\alpha^3 - 4z_\alpha^5). \qquad (39)$$

For $i = 2$, the leading-order contribution to (38) is $n^{-1}\phi(z_\alpha)$ times

$$z_\alpha\bigl(((e_3^*)^2 - 1)e_3^2C^2 + (1 + 2z_\alpha^2)e_3^*(\tfrac{1}{4}e_3^2C^2 - \tfrac{1}{6}e_4D + \tfrac{1}{2}\sum_i E_iF_i)\bigr). \qquad (40)$$

For $i = 4$, we get a contribution of $n^{-1}\phi(z_\alpha)$ times

$$(z_\alpha^3 - 3z_\alpha)\bigl(-\tfrac{1}{12}(e_4^* - 1)e_4D + \tfrac{2}{3}((e_3^*)^2 - 1)e_3^2C^2\bigr), \qquad (41)$$

and, for $i = 6$, $n^{-1}\phi(z_\alpha)$ times

$$(z_\alpha^5 - 10z_\alpha^3 + 15z_\alpha)\tfrac{1}{18}((e_3^*)^2 - 1)e_3^2C^2. \qquad (42)$$

Adding up the contributions (39), (40), (41), and (42) yields the term of order n^{-1} in (11). This completes the proof of Theorem 1.

For Theorem 2, we must evaluate (35), a relatively simple matter, since it depends only on the ν_i for even i. A straightforward calculation yields (15).

The rejection probability of the asymptotic test based on τ (one-tailed with rejection in the left-hand tail) is the probability mass to the left of z_α in the distribution with expansion (26). Thus, with the values of the λ_i computed here, the ERP is $n^{-1/2}\phi(z_\alpha)$ times

$$\begin{aligned}&\tfrac{1}{6}e_3C(1 + 2z_\alpha^2) + n^{-1/2}\Bigl(\tfrac{1}{2}Hz_\alpha + e_3^2C^2(\tfrac{1}{6}z_\alpha - \tfrac{1}{9}z_\alpha^3 - \tfrac{1}{18}z_\alpha^5) + \tfrac{1}{12}e_4D(z_\alpha^3 - 3z_\alpha)\\ &+(\tfrac{1}{2}\sum_i\sum_j A_{ij}E_iE_j - \sum_i E_iF_i)(z_\alpha^3 - z_\alpha) + (\tfrac{1}{2}\sum_i\sum_j A_{ij}B_{ij} - \sum_i G_i)z_\alpha\Bigr).\end{aligned}$$

As stated in the text, the leading-order term of this equal to that of (11) with $e_3^* = 0$.

Table 1. Regressors

Obs	$\boldsymbol{x}_1$	$\boldsymbol{x}_3$	$\boldsymbol{x}_4$	$\boldsymbol{x}_5$	$\boldsymbol{x}_6$
1	0.616572	0.511730	0.210851	-0.651571	0.509960
2	10.000000	5.179612	4.749082	6.441719	1.212823
3	-0.600679	0.255896	-0.150372	-0.530344	0.318283
4	-0.613076	0.705476	0.447747	-1.599614	-0.601335
5	-1.972106	-0.673980	-1.513501	0.533987	0.654767
6	0.409741	0.922026	1.162060	-1.328799	1.607007
7	-0.676614	0.515275	-0.241203	-1.424305	-0.360405
8	0.400136	0.459530	0.166282	0.040292	-0.018642
9	1.106144	2.509302	0.899661	-0.188744	1.031873
10	0.671560	0.454057	-0.584329	1.451838	0.665312

Note: For $k = 1$, the only regressor is $\boldsymbol{x}_1$, for $k = 2$ there is also the constant, for $k = 3$ there are the constant, $\boldsymbol{x}_1$, and $\boldsymbol{x}_3$, and so forth.

Table 2. Influence of the design on the Edgeworth expansion

	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$
	Heteroskedastic errors					
C	3.15	3.14	2.87	2.86	-1.81	-1.57
D	9.97	9.91	8.92	8.78	5.27	4.08
E_0	0.00	0.98	6.67	6.98	1.71	1.02
E_1	9.29	9.27	8.31	8.17	3.94	2.90
	Homoskedastic errors					
C	2.82	2.28	-1.13	-0.20	0.44	0.51
D	8.67	7.02	3.69	1.78	2.29	2.54
E_0	0.00	0.00	0.00	0.00	0.00	0.00
E_1	8.67	7.02	3.69	1.78	2.29	2.54

Note: E_0 is $\sum_i E_i F_i$ for the case of constrained residuals; E_1 is for unconstrained residuals.

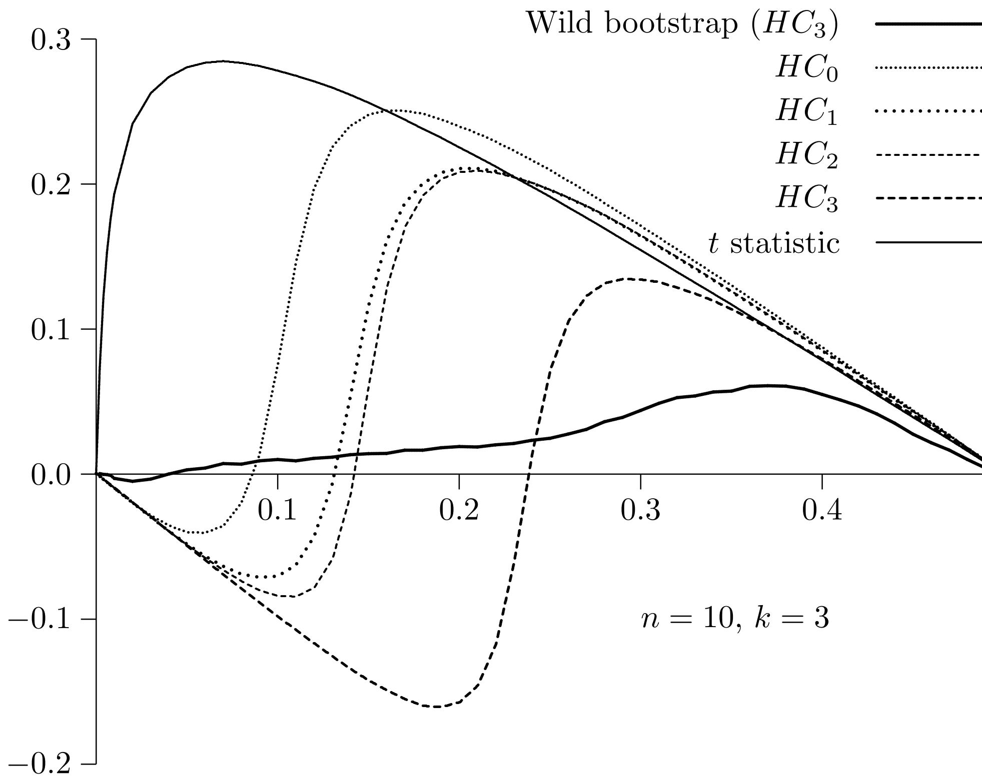

Figure 1. ERPs of asymptotic and bootstrap tests

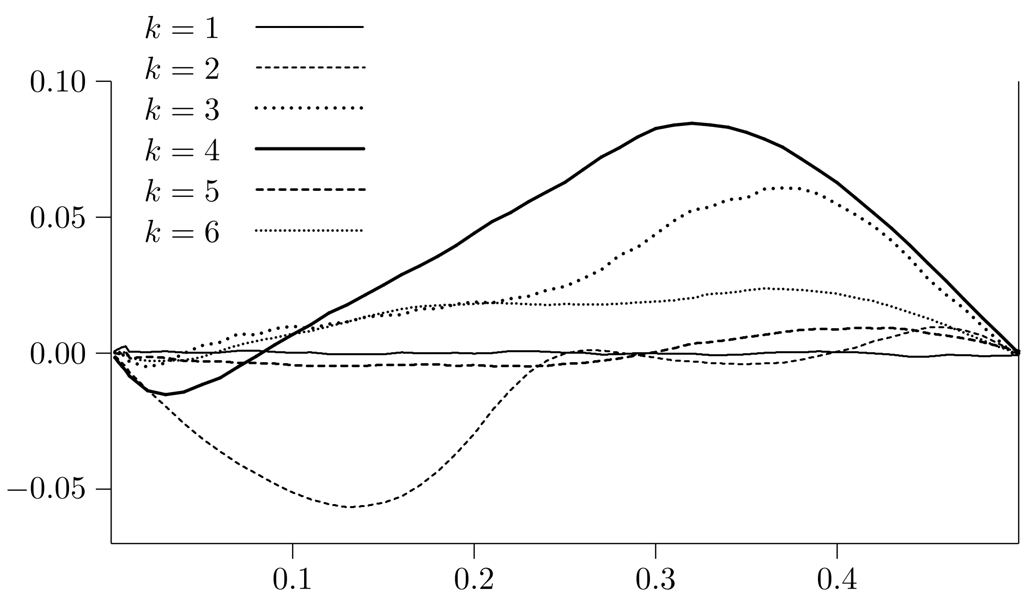

Figure 2. Base case with different designs

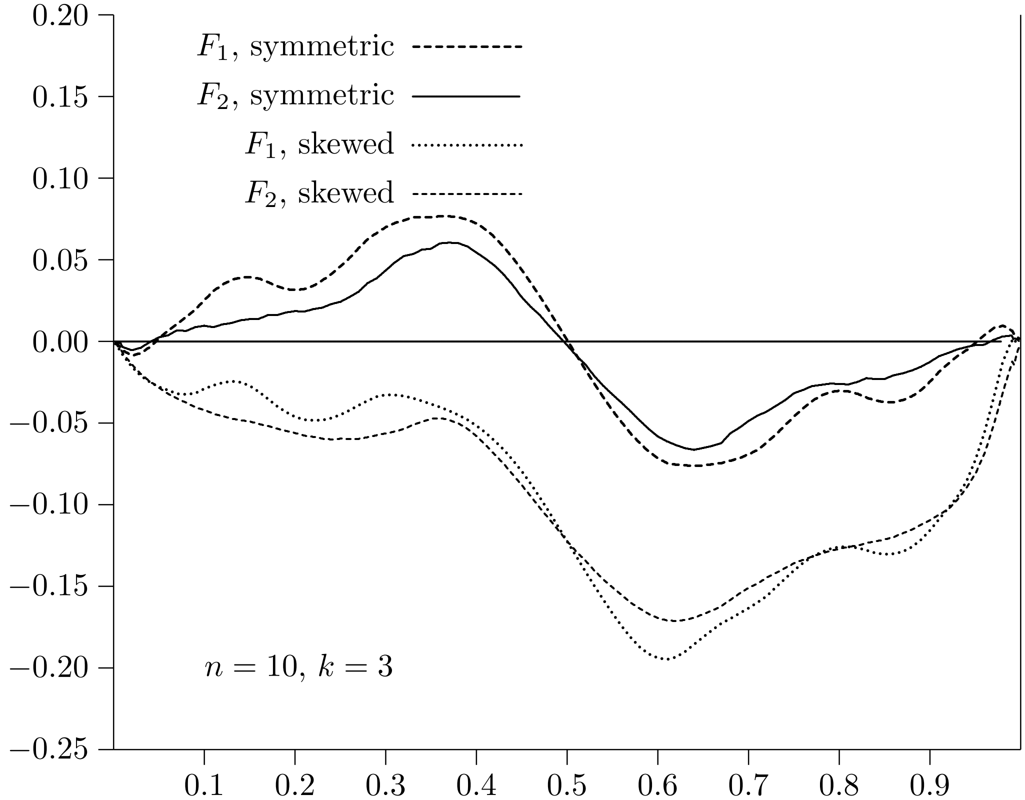

Figure 3. Symmetric and skewed errors, F_1 and F_2

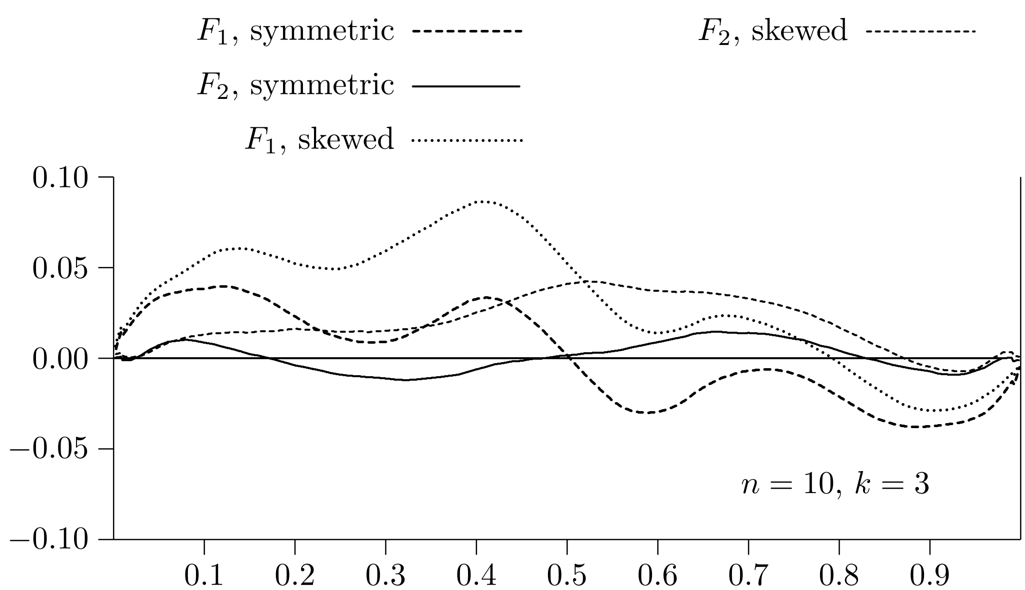

Figure 4. Homoskedastic errors

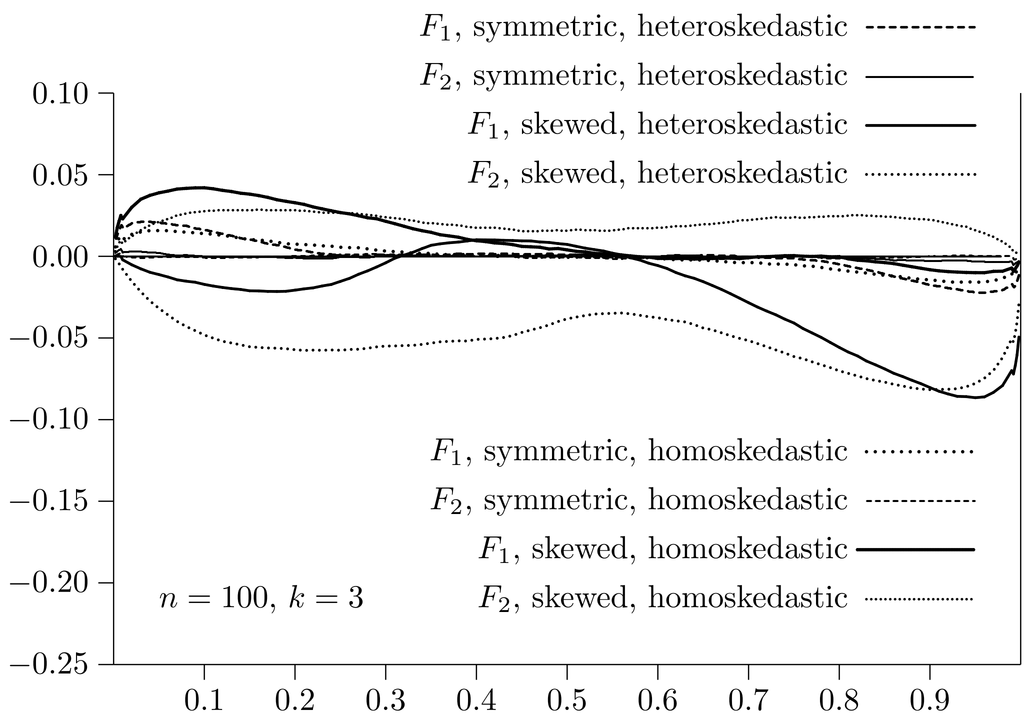

Figure 5. ERPs for $n = 100$

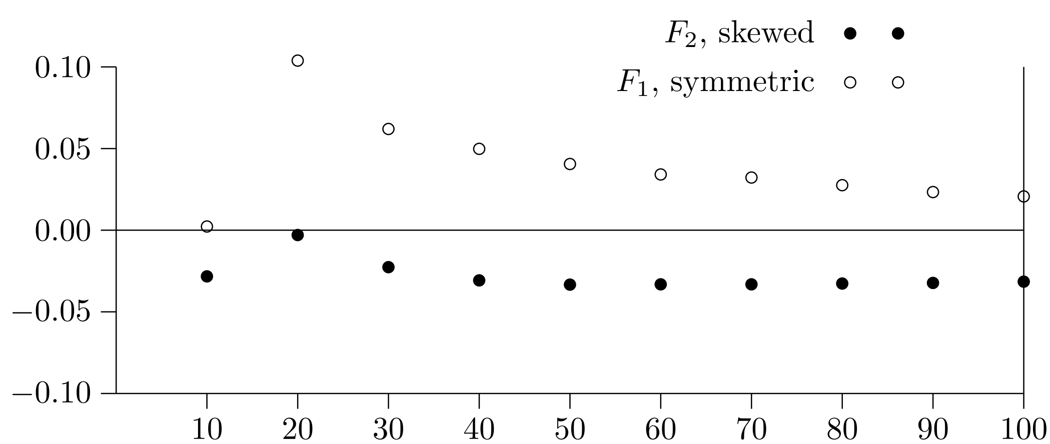

Figure 6. ERP for nominal level 0.05 as function of sample size

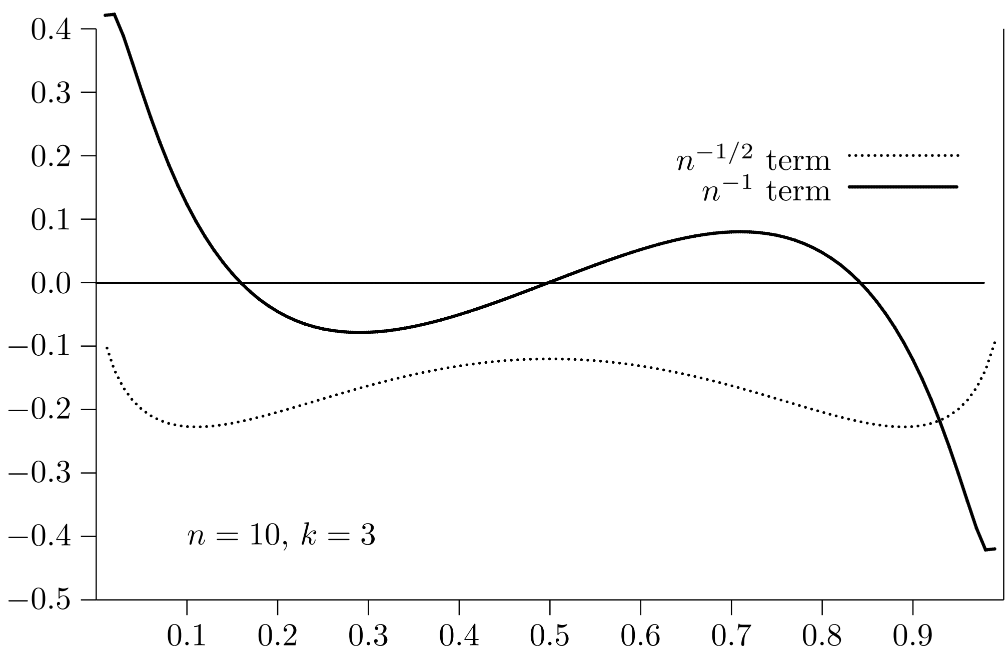

Figure 7a. The contributions of order $n^{-1/2}$ and n^{-1} to the ERP

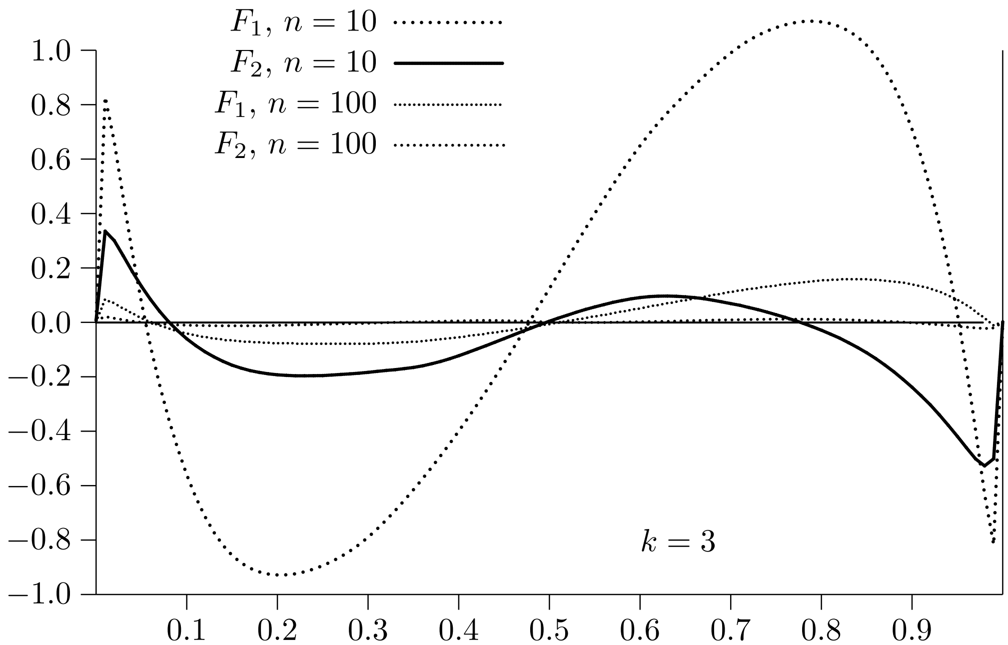

Figure 7b. Error in Edgeworth ERP, F_1, F_2, skewed errors

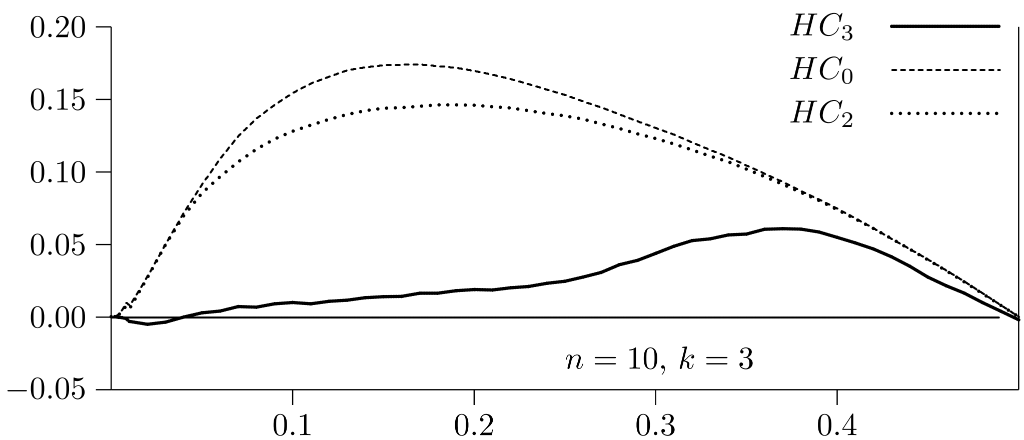

Figure 8. HC_3 **compared with** HC_0 **and** HC_2

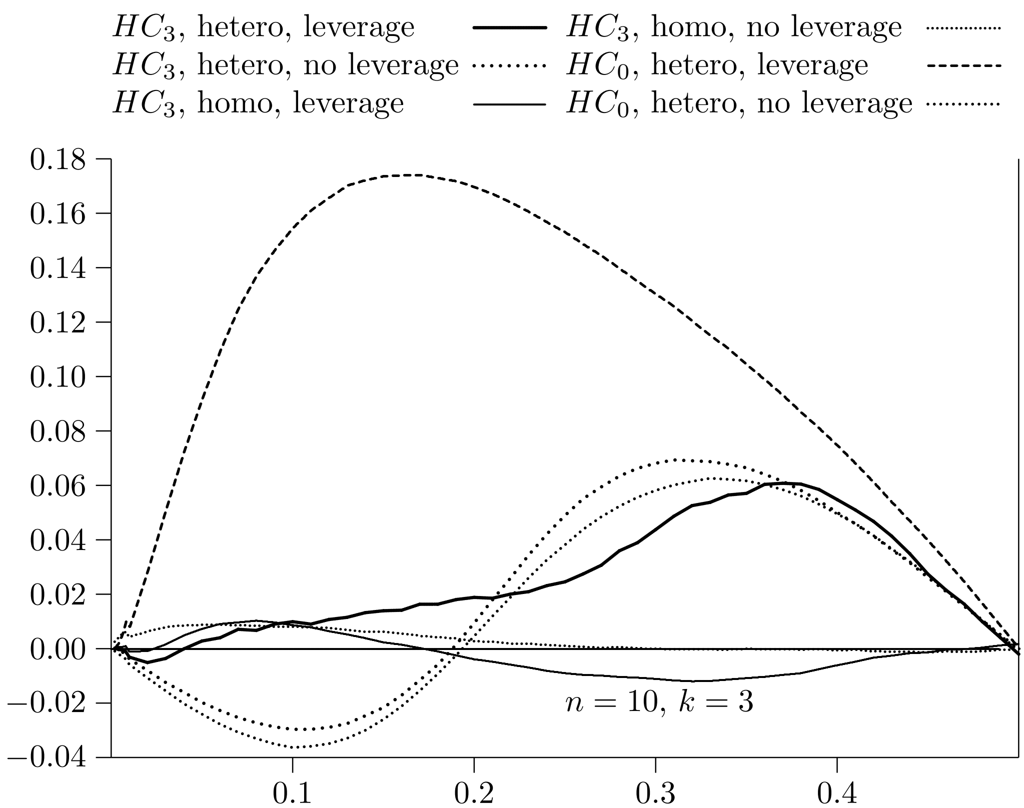

Figure 9. Relative importance of leverage and heteroskedasticity