

**ISSN 1352-2469**

**Distributional Analysis Research Programme**

# **The Wild Bootstrap, Tamed at Last**

**by**

**Russell Davidson**

**GREQAM**

**and**

**Department of Economics**

**Queen's University**

**Kingston, Ontario, Canada**

**Emmanuel Flachaire**

**STICERD**

**London School of Economics**

**Discussion Paper**

**No. DARP 58**

**February 2001**

**Distributional Analysis Research Programme**

**The Toyota Centre**

**Suntory and Toyota International**

**Centres for Economics and**

**Related Disciplines**

**London School of Economics**

**Houghton Street**

**London WC2A 2AE**

## **Distributional Analysis Research Programme**

The Distributional Analysis Research Programme was established in 1993 with funding from the Economic and Social Research Council. It is located within the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) at the London School of Economics and Political Science. The programme is directed by Frank Cowell. The Discussion Paper series is available free of charge and most papers are downloadable from the website. To subscribe to the DARP paper series, or for further information on the work of the Programme, please contact our Research Secretary, Sue Coles on:

Telephone:	UK+20 7955 6678
Fax:	UK+20 7955 6951
Email:	s.coles@lse.ac.uk
Web site:	<a href="http://sticerd.lse.ac.uk/DARP">http://sticerd.lse.ac.uk/DARP</a>

© Authors: **Russell Davidson and Emmanuel Flachaire**

All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

## ABSTRACT

Various versions of the wild bootstrap are studied as applied to regression models with heteroskedastic errors. It is shown that some versions can be qualified as “tamed,” in the sense that the statistic bootstrapped is asymptotically independent of the distribution of the wild bootstrap DGP. This can, in one very specific case, lead to perfect bootstrap inference, and leads to substantial reduction in the error in the rejection probability of a bootstrap test much more generally. However, the version of the wild bootstrap with this desirable property does not benefit from the skewness correction afforded by the most popular version of the wild bootstrap in the literature. Edgeworth expansions and simulation experiments are used to show why this defect does not prevent the preferred version from having the smallest error in rejection probability in small and medium-sized samples. It is concluded that this preferred version should always be used in practice.

## ACKNOWLEDGEMENT

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. Remaining errors are ours.

**Keywords:** Wild bootstrap, heteroskedasticity consistent covariance matrix estimators, size distortion.

**JEL classification:** C1

---

© Emmanuel Flachaire. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

## 1. Introduction

Inference on the parameters of the linear regression model

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

where  $\mathbf{y}$  is an  $n$ -vector containing the values of the dependent variable,  $\mathbf{X}$  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  $\mathbf{u}$  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, especially in the presence of observations with high leverage; see also Chesher and Jewitt (1987), who show that the extent of the bias is related to the structure of the regressors. 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.

Freedman (1981) proposed a bootstrap technique which is asymptotically valid in the presence of heteroskedasticity of unknown form. This technique, often called the  $(y, X)$  bootstrap, involves resampling, for each observation in a bootstrap sample, both the dependent variable ( $y$ ) and the regressors ( $X$ ) of an observation in the original sample. In this way, the heteroskedasticity associated with a particular set of values of the regressors is retained in the bootstrap samples. This is achieved at the cost of generating bootstrap error terms that do not have mean zero conditional on the bootstrap regressors for a given observation, but this does not prevent its being applicable in certain circumstances. However, as we will see later, simulation evidence reveals that the  $(y, X)$  bootstrap does not give as satisfactory results as another procedure.

This better procedure is the so-called wild bootstrap. It 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, like the  $(y, X)$  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, HCCME based statistics are asymptotically independent of the random elements that determine some versions of the wild bootstrap DGP. Davidson and MacKinnon (1999) have shown that such asymptotic independence can lead to asymptotic refinements of bootstrap inference, and this seems also to be the case with the wild bootstrap. We are able to go further when the hypothesis under test is that all the regression parameters are zero. In that event, we show that one version of the wild bootstrap gives essentially perfect inference.

Much of the asymptotic theory of the bootstrap is based on Edgeworth expansions; see Hall (1992). In its application to testing, this theory is most often used to determine the order of the difference between actual and nominal rejection probabilities of a bootstrap test as a negative power of the sample size  $n$ . At a certain algebraic cost, approximate expressions for this difference can also be obtained. However, these are rarely good approximations, especially in the tails of distributions. In this paper, we obtain such an approximation, valid through order  $n^{-1}$ , for the wild bootstrap. It gives approximate confirmation of our other results, and is useful for developing intuition. In particular, it shows that the order of a term in an expansion as a negative power of  $n$  is not the only, or even the most important, determinant of its quantitative importance in not very large samples. We conclude that using Edgeworth expansions merely to determine powers of  $n$  may yield misleading conclusions.

In general, the version of the wild bootstrap that gives perfect inference in very restricted circumstances suffers from some size distortion, but, it would appear, never more than any other version, as we demonstrate in a series of simulation experiments. For these experiments, our policy is to concentrate on cases in which the asymptotic tests based on the HCCME are very badly behaved, and to try to identify bootstrap procedures that go furthest in correcting this bad behaviour. Thus, except for the purposes of obtaining benchmarks, we look at small samples of size 10, with an observation of very high leverage, and a great deal of heteroskedasticity closely correlated with the regressors.

It is of course important to study what happens when the error terms are not symmetrically distributed. The asymptotic refinements found by Wu and Mammen for certain versions of the wild bootstrap are due to taking account of such skewness. We show the extent of the degradation in performance with asymmetric error terms, but show that our preferred version of the wild bootstrap continues

to work at least as well as any other, including the popular version of Liu and Mammen which takes explicit account of skewness.

In section 2, we review the properties of bootstrap  $P$  values, and the circumstances in which they may benefit from refinements of various sorts. In section 3, we discuss a number of ways in which the wild bootstrap may be implemented, and show that, with symmetrically distributed error terms, a property of asymptotic independence holds that gives rise to an asymptotic refinement of bootstrap  $P$  values. In some special cases, the refinement can give rise to essentially exact inference. In Section 4, we present results on the difference between actual and nominal rejection probabilities based on Edgeworth expansions. 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 rejection probabilities estimated by simulation with the predictions of Edgeworth expansions. These experiments give strong evidence in favour of our preferred version of the wild bootstrap. A few conclusions are drawn in section 6.

## 2. Bootstrap $P$ Values

Beran (1988) showed that bootstrap inference is refined when the quantity bootstrapped is asymptotically pivotal. It is convenient to formalise the idea of pivotalness by means of a few formal definitions. A *data-generating process*, or *DGP*, is any rule sufficiently specific to allow artificial samples of arbitrary size to be simulated on the computer. Thus all parameter values and all probability distributions must be provided in the specification of a DGP. A *model* is a set of DGPs. Models are usually generated by allowing parameters and probability distributions to vary over admissible sets. A test statistic is a random variable that is a deterministic function of the data generated by a DGP and, possibly, other exogenous variables. A test statistic  $\tau$  is a *pivot* for a model  $\mathbb{M}$  if, for each sample size  $n$ , its distribution is independent of the DGP  $\mu \in \mathbb{M}$  which generates the data from which  $\tau$  is calculated. The *asymptotic distribution* of a test statistic  $\tau$  for a DGP  $\mu$  is the limit, if it exists, of the distribution of  $\tau$  under  $\mu$  as the sample size tends to infinity. The statistic  $\tau$  is *asymptotically pivotal* for  $\mathbb{M}$  if its asymptotic distribution exists for all  $\mu \in \mathbb{M}$  and is independent of  $\mu$ .

In hypothesis testing, the null hypothesis under test is represented by a model, as defined above. A test statistic is said to be pivotal or asymptotically pivotal under the null hypothesis if it is a pivot or an asymptotic pivot for the model that represents the hypothesis. Most test statistics commonly used in econometric practice are asymptotically pivotal under the null hypotheses they test, since asymptotically they have distributions, like standard normal, or chi-squared, that do not depend on unknown parameters. Conventional asymptotic inference is based on these known asymptotic distributions.

If an asymptotic pivot  $\tau$  is not an exact pivot, its distribution depends on which particular DGP  $\mu \in \mathbb{M}$  generates the data used to compute it. In this case, bootstrap inference is no longer exact in general. The bootstrap samples used

to estimate the finite-sample distribution of  $\tau$  are generated by a *bootstrap DGP*, which, although it usually belongs to  $\mathbb{M}$ , is in general different from the DGP that generated the original data.

In order to perform a bootstrap test, the simplest approach is to calculate the marginal significance level, or  $P$  value, corresponding to the realised statistic. Suppose that data are generated by a DGP  $\mu_0$  belonging to  $\mathbb{M}$ , and used to compute a realisation  $\hat{\tau}$  of the random variable  $\tau$ . Then, for a test that rejects for large values of the statistic, the  $P$  value we would ideally like to compute is

$$p(\hat{\tau}) \equiv \Pr_{\mu_0}(\tau > \hat{\tau}). \quad (1)$$

In practice, (1) cannot be computed, or estimated by simulation, because the DGP  $\mu_0$  that generates observed data is unknown. If  $\tau$  is an exact pivot, this does not matter, since (1) can be computed using any DGP in  $\mathbb{M}$ . In this case,  $p(\hat{\tau})$  is a drawing from the  $U(0, 1)$  distribution. If  $\tau$  is only an asymptotic pivot, the *bootstrap  $P$  value* is defined by

$$p^*(\hat{\tau}, \hat{\mu}) \equiv \Pr_{\hat{\mu}}(\tau > \hat{\tau}), \quad (2)$$

where  $\hat{\mu}$  is a (random) bootstrap DGP in  $\mathbb{M}$ , determined in some suitable way from the same data as those used to compute  $\hat{\tau}$ . We denote by  $\mu^*$  the random DGP of which  $\hat{\mu}$  is a realisation.

Let the asymptotic CDF of the asymptotic pivot  $\tau$  be denoted by  $F$ . At nominal level  $\alpha$ , an asymptotic test rejects if the asymptotic  $P$  value  $1 - F(\hat{\tau}) < \alpha$ . In order to avoid having to deal with different asymptotic distributions, it is convenient to replace the raw statistic  $\tau$  by the asymptotic  $P$  value  $1 - F(\tau)$ , of which the asymptotic distribution is always  $U(0, 1)$ . For the remainder of this section,  $\tau$  denotes such an asymptotic  $P$  value.

For the sample size of the observed data, the “rejection probability function,” or RPF, provides a measure of the true rejection probability of the asymptotic test. This function, which gives the rejection probability under  $\mu$  of a test at nominal level  $\alpha$ , is defined as follows:

$$R(\alpha, \mu) \equiv \Pr_{\mu}(\tau < \alpha). \quad (3)$$

It is clear that  $R(\cdot, \mu)$  is the CDF of  $\tau$  under  $\mu$ . The information contained in the function  $R$  is also provided by the “critical value function,” or CVF,  $Q$ , defined implicitly by the equation

$$\Pr_{\mu}(\tau < Q(\alpha, \mu)) = \alpha. \quad (4)$$

$Q(\alpha, \mu)$  is just the  $\alpha$  quantile of  $\tau$  under  $\mu$ . It follows from (3) and (4) that

$$R(Q(\alpha, \mu), \mu) = \alpha, \quad \text{and, conversely,} \quad Q(R(\alpha, \mu), \mu) = \alpha, \quad (5)$$

from which it is clear that, for given  $\mu$ ,  $R$  and  $Q$  are inverse functions.

The bootstrap test rejects at nominal level  $\alpha$  if  $\tau < Q(\alpha, \mu^*)$ , that is, if  $\tau$  is smaller than the the  $\alpha$ -quantile of the bootstrap DGP. By acting on both sides with  $R(\cdot, \mu^*)$ , this condition can also be expressed as

$$R(\tau, \mu^*) < R(Q(\alpha, \mu^*), \mu^*) = \alpha.$$

This makes it clear that the bootstrap  $P$  value is just  $R(\tau, \mu^*)$ . It follows that, if  $R$  actually depends on  $\mu^*$ , that is, if  $\tau$  is not an exact pivot, the bootstrap test is not equivalent to the asymptotic test, because the former depends not only on  $\tau$ , but also on the random  $\mu^*$ .

### 3. The Wild Bootstrap

Consider the linear regression model

$$y_t = x_{t1}\beta_1 + \mathbf{X}_{t2}\beta_2 + u_t, \quad t = 1, \dots, n, \quad (6)$$

in which the explanatory variables are assumed to be strictly exogenous, in the sense that, for all  $t$ ,  $x_{t1}$  and  $\mathbf{X}_{t2}$  are independent of all of the error terms  $u_s$ ,  $s = 1, \dots, n$ . The row vector  $\mathbf{X}_{t2}$  contains observations on  $k - 1$  variables, of which, if  $k > 1$ , one is a constant. We wish to test the null hypothesis that the coefficient  $\beta_1$  of the first regressor  $x_{t1}$  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  $E(u_t^2) = \sigma_t^2$ . We write  $u_t = \sigma_t v_t$ , where  $E(v_t^2) = 1$ . We consider only *unconditional* heteroskedasticity, which means that the  $\sigma_t^2$  may depend on the exogenous regressors, but not, for instance, on lagged dependent variables. The model represented by (6), in the sense of the previous section, is thus generated by the variation of the parameters  $\beta_1$  and  $\beta_2$ , the variances  $\sigma_t^2$ , 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 pseudo- $t$  statistics for testing whether  $\beta_1 = 0$  are then asymptotically pivotal for the restricted model in which we set  $\beta_1 = 0$  if we also impose the weak condition that the  $\sigma_t^2$  are bounded away from zero and infinity.

We write  $\mathbf{x}_1$  for the  $n$ -vector with typical element  $x_{t1}$ , and  $\mathbf{X}_2$  for the  $n \times (k - 1)$  matrix with typical row  $\mathbf{X}_{t2}$ . By  $\mathbf{X}$  we mean the full  $n \times k$  matrix  $[\mathbf{x}_1 \ \mathbf{X}_2]$ . Then the basic HCCME for the OLS parameter estimates of (6) is

$$(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \hat{\boldsymbol{\Omega}} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}, \quad (7)$$

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 (6) or the constrained model in which  $\beta_1 = 0$  is imposed. We refer to the version (7) 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  $h_t \equiv \mathbf{X}_t(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}_t^\top$ , the  $t^{\text{th}}$  diagonal element of the orthogonal projection matrix on to the span of the columns of  $\mathbf{X}$ . 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

$$y_t^* = \mathbf{X}_t \hat{\boldsymbol{\beta}} + u_t^*, \quad (8)$$

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

$$u_t^* = f_t(\hat{u}_t) \varepsilon_t, \quad (9)$$

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

$$E(\varepsilon_t) = 0 \quad \text{and} \quad E(\varepsilon_t^2) = 1. \quad (10)$$

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  $E(\varepsilon_t^3) = 0$  is often added. Liu (1988) considers model (6) 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  $\varepsilon_t$ , namely the following two-point distribution:

$$F_1 : \quad \varepsilon_t = \begin{cases} -(\sqrt{5} - 1)/2 & \text{with probability } p = (\sqrt{5} + 1)/(2\sqrt{5}) \\ (\sqrt{5} + 1)/2 & \text{with probability } 1 - p. \end{cases} \quad (11)$$

Liu also mentions the possibility of Rademacher variables, defined as

$$F_2 : \quad \varepsilon_t = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2, \end{cases} \quad (12)$$

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 (12), being a lattice distribution, does not lend itself to rigorous techniques

based on Edgeworth expansion. In this paper, we show by other methods that (12) is, for all the cases we consider, the best choice of distribution for the  $\varepsilon_t$ , and obtain non-rigorous confirmation of this by use of formal Edgeworth expansions. Another variant of the wild bootstrap that we consider later is obtained by replacing (9) by

$$u_t^* = f_t(|\hat{u}_t|)\varepsilon_t, \quad (13)$$

in which the absolute values of the residuals are used instead of the signed residuals.

Conditional on the random elements  $\hat{\beta}$  and  $\hat{u}_t$ , the wild bootstrap DGP (8) clearly belongs to the null hypothesis if the first component of  $\hat{\beta}$ , corresponding to the regressor  $\mathbf{x}_1$ , is zero, since the bootstrap error terms  $u_t^*$  have mean zero and are heteroskedastic, for both formulations, (9) or (13), for any distribution for the  $\varepsilon_t$  satisfying (10). Since (6) is linear, we may also set the remaining components of  $\beta$  to zero, since the distribution of any HCCME-based pseudo- $t$  statistic does not depend on the value of  $\beta_2$ . Since the HCCME-based statistics we have discussed are asymptotically pivotal, inference based on the wild bootstrap using such a statistic applied to model (6) is asymptotically valid. In the case of a nonlinear regression, the distribution of the test statistic does depend on the specific value of  $\beta_2$ , 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 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 (8), since the squared residuals are not consistent estimators of the  $\sigma_t^2$ . In a somewhat different context from the present one, Davidson and MacKinnon (1999) show that bootstrap inference can be refined, sometimes beyond Beran's refinement, if the statistic that is bootstrapped is asymptotically independent of the bootstrap DGP. It is tempting to see if a similar refinement is available for the wild bootstrap. In the theorems that follow, we show that such asymptotic independence can certainly be attained if the wild bootstrap makes use of the  $F_2$  distribution, and that, in a very specific case, it leads to exact inference.

As discussed by Davidson and MacKinnon, it is often useful for achieving this asymptotic independence to base the bootstrap DGP  $\mu^*$  exclusively on estimates *under the null hypothesis*. If one does that, (8) becomes just

$$y_t^* = u_t^*, \quad u_t^* = f_t(\tilde{u}_t)\varepsilon_t, \quad (14)$$

where the OLS residuals  $\tilde{u}_t$  are obtained from the regression  $y_t = \mathbf{X}_{t2}\beta_2 + u_t$  that incorporates the constraint of the null hypothesis. The transformation  $f$  may involve taking the absolute value of the argument. It turns out that, when asymptotic independence can be achieved for the wild bootstrap, it does not matter whether constrained or unconstrained residuals are used. However, for the case in which exact inference is possible, constrained residuals are necessary, not only for the bootstrap DGP, but also for the construction of the HCCME (7). Simulation evidence shows that, when exact inference is no longer possible, size distortion remains less when constrained residuals are used.

We now show that, if the model (6) is restricted so that the error terms have distributions that are symmetric about the mean of zero, all HCCME-based  $t$  statistics for  $\beta_1 = 0$  are asymptotically independent of some versions of the wild bootstrap DGP (14). First, an easy Lemma.

**Lemma 1:** A mean-zero random variable  $u$  which has zero probability mass on the origin and the density of which is symmetric about the origin is the product of two independent random variables: the absolute value  $|u|$  and the sign  $\text{sgn}(u)$ .

**Proof:** Denote the density of  $u$  by  $f(u)$ . Since  $f(-u) = f(u)$ , the density of  $|u|$  is  $g(|u|) \equiv 2f(|u|)$ . The density of  $\text{sgn}(u)$  can be written in terms of indicator functions as  $0.5I(u < 0) + 0.5I(u > 0) = 0.5$ . Here we use the fact that there is no positive probability mass on the origin itself. The product of the two densities is  $2f(|u|) \cdot 0.5 = f(|u|) = f(u)$ , the density of  $u$  itself. The factorisation of this density shows that  $|u|$  and  $\text{sgn}(u)$  are independent. ■

**Theorem 1:** Consider the linear regression model

$$y_t = x_{t1}\beta_1 + \mathbf{X}_{t2}\beta_2 + u_t, \quad (15)$$

where the regressors are strictly exogenous, and the error terms are mutually independent with mean zero and distributions symmetric about the origin with no positive probability mass on the origin. All pseudo- $t$  statistics based on the HCCME for the hypothesis that  $\beta_1 = 0$  can be written as

$$\tau \equiv \mathbf{x}_1^\top \mathbf{M}_2 \mathbf{y} / (\mathbf{x}_1^\top \mathbf{M}_2 \hat{\mathbf{\Omega}} \mathbf{M}_2 \mathbf{x}_1)^{1/2}. \quad (16)$$

Here  $\mathbf{y}$  is the  $n$ -vector with typical element  $y_t$ ,  $\hat{\mathbf{\Omega}}$  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  $\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top$  is the orthogonal projection matrix on to the orthogonal complement of the span of the columns of  $\mathbf{X}_2$ .

If the regressors obey the usual regularity condition that  $n^{-1} \mathbf{X}^\top \mathbf{X}$  tends as  $n \rightarrow \infty$  to a deterministic positive definite finite matrix, and if, in addition, there exist positive bounds  $\underline{\sigma}^2$  and  $\bar{\sigma}^2$  such that  $\underline{\sigma}^2 \leq \sigma_t^2 \leq \bar{\sigma}^2$  for all  $t$ , then the statistic  $\tau$  of (16) is asymptotically independent, under the null hypothesis, of the absolute values  $|\tilde{u}_t|$  of the residuals, and consequently also of the wild bootstrap DGP  $\mu^*$  defined by (14) if the transformation  $f$  depends only on the absolute value of its argument.

**Proof:** By the Frisch-Waugh-Lovell theorem (see, for instance, Davidson and MacKinnon (1993), Chapter 1), the OLS estimate of  $\beta_1$  from (15) is the same as the OLS estimate from the regression  $\mathbf{M}_2 \mathbf{y} = \mathbf{M}_2 \mathbf{x}_1 \beta_1 + \text{residuals}$ , that is,

$$\hat{\beta}_1 = (\mathbf{x}_1^\top \mathbf{M}_2 \mathbf{x}_1)^{-1} \mathbf{x}_1^\top \mathbf{M}_2 \mathbf{y}. \quad (17)$$

The HCCME of the variance of  $\hat{\beta}_1$  is obtained by applying (7) to the regression (17). The estimated variance is thus

$$\mathbf{x}_1^\top \mathbf{M}_2 \hat{\mathbf{\Omega}} \mathbf{M}_2 \mathbf{x}_1 (\mathbf{x}_1^\top \mathbf{M}_2 \mathbf{x}_1)^{-2}, \quad (18)$$

Equation (16) follows from (17) and (18).

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

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

where  $\ddot{u}_t$  is the  $t^{\text{th}}$  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  $HC_3$ . But the squared residuals  $\ddot{u}_t^2$  are asymptotically equal to the squared error terms  $u_t^2$ , and under the conditions of the theorem,  $a_t = 1 + O(n^{-1})$  for all  $t$ . Under the null,  $\mathbf{M}_2 \mathbf{y} = \mathbf{M}_2 \mathbf{u}$ , and so the statistic (16) is asymptotically equal to

$$\frac{\sum_{t=1}^n (\mathbf{M}_2 \mathbf{x}_1)_t u_t}{\left(\sum_{t=1}^n (\mathbf{M}_2 \mathbf{x}_1)_t^2 u_t^2\right)^{1/2}} \quad (19)$$

Let us write  $u_t = |u_t|s_t$ , where  $s_t$ , the sign of  $u_t$ , is equal to either  $+1$  or  $-1$ . In addition, let us write  $z_t \equiv (\mathbf{M}_2 \mathbf{x}_1)_t |u_t|$ . Then (19) is equal to

$$\frac{\sum_{t=1}^n z_t s_t}{\left(\sum_{t=1}^n z_t^2\right)^{1/2}}. \quad (20)$$

By Lemma 1, the  $z_t$  and the  $s_t$  are independent, and so, conditional on the  $z_t$ , the  $s_t$  are mutually independent and distributed according to the law  $F_2$  of (12). Under the regularity conditions of the second part of the theorem, the central limit theorem can be applied to show that the asymptotic distribution of (20) conditional on the  $z_t$  is standard normal. Since this asymptotic distribution is independent of the  $z_t$  and so of the  $|u_t|$ , it follows that  $\tau$  is asymptotically independent of the  $|u_t|$ , and so also of any  $\mu^*$  defined exclusively in terms of the exogenous regressors and the absolute values of the residuals, constrained or unconstrained. ■

**Remarks and Corollaries:** Note that the theorem applies to any wild bootstrap defined by (13) and based on the absolute values of the residuals, provided only that the distribution of the  $\varepsilon_t$  satisfies (10). If this distribution is itself symmetric about the origin, like  $F_2$ , then the theorem applies with any transformation  $f$  that is either even or odd, because then the vector with typical element  $f_t(\ddot{u}_t)\varepsilon_t$  has the same distribution as that with typical element  $f_t(|\ddot{u}_t|)\varepsilon_t$ . We refer to any wild bootstrap DGP that can be expressed, implicitly or explicitly, in terms of the  $|\ddot{u}_t|$  only as a *tamed* wild bootstrap DGP.

It is easy to adapt the above proof so that it applies to the case in which a joint hypothesis is tested with more than one degree of freedom. The statistic takes on a chi-squared form. If we replace the single column  $\mathbf{x}_1$  by a matrix  $\mathbf{X}_1$ , we may define a matrix  $\mathbf{Z}$  with typical row  $(\mathbf{M}_2 \mathbf{X}_1)_t |u_t|$ , and the  $n$ -vector  $\mathbf{s}$  with typical element  $s_t$ . The statistic is asymptotically equivalent to

$$\mathbf{s}^\top \mathbf{Z} (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{s} = \mathbf{s}^\top \mathbf{P}_Z \mathbf{s}, \quad (21)$$

where  $\mathbf{P}_Z$  is the orthogonal projection on to the columns of  $\mathbf{Z}$ . Since  $\mathbf{s}$  and  $\mathbf{Z}$  are independent, the asymptotic distribution of (21) conditional on the  $|u_t|$  is chi-squared with as many degrees of freedom as  $\mathbf{X}_1$  has columns, and, since this does not depend on the  $|u_t|$ , the statistic is asymptotically independent of the  $|u_t|$  and so of any tamed bootstrap DGP  $\mu^*$ .

In the case of nonlinear regression, the bootstrap DGP must be constructed using a consistent estimate of  $\beta_2$ . The NLS estimate  $\tilde{\beta}_2$  obtained by estimating the restricted model with  $\beta_1 = 0$  is asymptotically independent of the NLS residuals  $\tilde{u}_t$  from the same regression. A tamed bootstrap DGP can thus be defined exclusively in terms of  $\tilde{\beta}_2$  and the  $|\tilde{u}_t|$ . The statistic  $\tau$  is then asymptotically independent of this tamed bootstrap DGP.

Although for the proof of the theorem it is easier to speak loosely of the asymptotic independence of  $\tau$  and the absolute values of the residuals, what is meant more correctly is that  $\tau$  is asymptotically independent of the random CDF of the bootstrap statistic generated by a wild bootstrap DGP based on these absolute values. This point will be clearer in Section 4, when we make formal Edgeworth expansions of this random distribution.

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

**Theorem 2:** Consider the linear regression model

$$y_t = \mathbf{X}_t \beta + u_t \quad (22)$$

where the  $n \times k$  matrix  $\mathbf{X}$  with typical row  $\mathbf{X}_t$  is independent of all the symmetrically distributed error terms  $u_t$ , which satisfy the same regularity conditions as for Theorem 1. Under the null hypothesis that  $\beta = \mathbf{0}$ , the  $\chi^2$  statistic for a test of that null against the alternative represented by (22), based on any of the four HCCMEs considered here constructed with constrained residuals, has exactly the same distribution as the same statistic bootstrapped, if the bootstrap DGP is the tamed wild bootstrap (14), with  $f(u) = u$  or equivalently  $f(u) = |u|$ , for which the  $\varepsilon_t$  are generated by the symmetric two-point distribution  $F_2$  of (12).

For sample size  $n$ , the bootstrap  $P$  value  $p^*$  follows a discrete distribution supported by the set of points  $p_i = i/2^n$ ,  $i = 0, \dots, 2^n - 1$ , with equal probability mass  $2^{-n}$  on each point.

**Proof:** The OLS estimates from (22) are given by  $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ , and any of the HCCMEs we consider for  $\hat{\beta}$  can be written in the form (7), with an appropriate choice of  $\hat{\Omega}$ . The  $\chi^2$  statistic thus takes the form

$$\tau \equiv \mathbf{y}^\top \mathbf{X} (\mathbf{X}^\top \hat{\Omega} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (23)$$

Under the null,  $\mathbf{y} = \mathbf{u}$ , and each component  $u_t$  of  $\mathbf{u}$  can be written as  $|u_t|s_t$ , where  $|u_t|$  and  $s_t$  are independent. Define the  $1 \times k$  row vector  $\mathbf{Z}_t$  as  $|u_t| \mathbf{X}_t$ , and the

$n \times 1$  column vector  $\mathbf{s}$  with typical element  $s_t$ . It follows from Lemma 1 that the entire  $n \times k$  matrix  $\mathbf{Z}$  with typical row  $\mathbf{Z}_t$  is independent of the vector  $\mathbf{s}$ . If the constrained residuals, which are just the elements of  $\mathbf{y}$ , are used to form  $\hat{\Omega}$ , the statistic (23) becomes

$$\mathbf{s}^\top \mathbf{Z} \left( \sum_{t=1}^n a_t \mathbf{Z}_t^\top \mathbf{Z}_t \right)^{-1} \mathbf{Z}^\top \mathbf{s}, \quad (24)$$

where the  $a_t$  are defined as in the proof of Theorem 1.

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

$$\boldsymbol{\varepsilon}^\top \mathbf{Z} \left( \sum_{t=1}^n a_t \mathbf{Z}_t^\top \mathbf{Z}_t \right)^{-1} \mathbf{Z}^\top \boldsymbol{\varepsilon}, \quad (25)$$

where  $\boldsymbol{\varepsilon}$  denotes the vector containing the  $\varepsilon_t$ . The matrix  $\mathbf{Z}$  is exactly the same as in (24), because the exogenous matrix  $\mathbf{X}$  is reused unchanged by the wild bootstrap, 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  $\mathbf{Z}$  are independent under the wild bootstrap DGP. But we saw in the proof of Theorem 1 that, under the null hypothesis,  $\mathbf{s}$  follows exactly the same distribution as  $\boldsymbol{\varepsilon}$ , and so it follows that  $\tau$  under the null and  $\tau^*$  under the wild bootstrap DGP with  $F_2$  have the same distribution. This proves the first assertion of the theorem.

Conditional on the  $|u_t|$ , this common distribution of  $\tau$  and  $\tau^*$  is of course a discrete distribution, since  $\boldsymbol{\varepsilon}$  and  $\mathbf{s}$  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  $\tau$  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  $\tau_i$ ,  $i = 1, \dots, 2^n$ , 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  $\tau_i$ , is just  $1 - i/2^n$ . As  $i$  ranges from 1 to  $2^n$ , the  $P$  value varies over the set of points  $p_i$ ,  $i = 0, \dots, 2^n - 1$ , 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. ■

**Remarks:** For small enough  $n$ , it may be quite feasible to enumerate all the possible values of the bootstrap statistic  $\tau^*$ , and thus obtain the exact value of the realisation  $p^*(\hat{\tau}, \hat{\mu})$ .

Although the discrete nature of the bootstrap distribution means that it is not possible to perform exact inference for an arbitrary significance level  $\alpha$ , 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 sections,  $2^n = 1024$ , and so the bootstrap  $P$  value cannot be in error by more than 1 part in a thousand.

It is possible to imagine a case in which the discreteness problem is aggravated by the coincidence of some adjacent values of the  $\tau_i$  of the proof of the theorem. For instance, if the only regressor in  $\mathbf{X}$  is the constant, the value of (24) depends only on the number of positive components of  $\mathbf{s}$  and not on their ordering. For this case, of course, it is not necessary to base inference on an HCCME. Coincidence of

values of the  $\tau_i$  will otherwise occur if all the explanatory variables take on exactly the same values for more than one observation. However, since this phenomenon is observable, it need not be a cause for concern. A very small change in the values of the components of the  $\mathbf{X}_t$  would be enough to break the ties in the  $\tau_i$ .

The exact result of the theorem is specific to the wild bootstrap with  $F_2$ . The proof works because the signs in the vector  $\mathbf{s}$  also follow the distribution  $F_2$ .

In terms of the analysis in section 2, the result of Theorem 2 can be means that the bootstrap rejection probability is equal to  $\alpha$  whenever  $\alpha$  is one of the dyadic numbers  $i/2^n$ . Since  $\tau$  and  $\tau^*$  have the same distribution, the functions  $R(\cdot, \mu_0)$  and  $R(\cdot, \mu^*)$  are the same, as are  $Q(\cdot, \mu_0)$  and  $Q(\cdot, \mu^*)$ .

Given the exact result of the theorem, it is of great interest to see the extent of the size distortion of the  $F_2$  bootstrap with constrained residuals when the null hypothesis involves only a subset of the regression parameters. This question will be investigated by Edgeworth expansion in the next section, where we will see that the distortion is of order lower than  $n^{-1}$ , and by simulation in the following section. At this stage, it is possible to see why the theorem does not apply more generally. The expressions (24) and (25) for  $\tau$  and  $\tau^*$  continue to hold if  $\mathbf{Z}_t$  is redefined as  $|\ddot{u}_t|(\mathbf{M}_2 \mathbf{X}_1)_t$ , where  $\mathbf{X}_1$  is the matrix of regressors admitted under the null. However, although  $\boldsymbol{\varepsilon}$  in  $\tau^*$  is by construction independent of  $\mathbf{Z}$ ,  $\mathbf{s}$  in  $\tau$  is not. This is because the covariance matrix of the residual vector  $\ddot{\mathbf{u}}$  is not diagonal in general, unlike that of the error terms  $\mathbf{u}$ . In Figure 1, this point is illustrated for the bivariate case. In panel a), two level curves are shown of the joint density of two symmetrically distributed and independent variables  $u_1$  and  $u_2$ . In panel b), the two variables are no longer independent. For the set of four points for which the absolute values of  $u_1$  and  $u_2$  are the same, it can be seen that, with independence, all four points lie on the same level curve of the joint density, but that this is no longer true without independence. The *vector* of absolute values is no longer independent of the *vector* of signs, even though independence still holds for the marginal distribution of each variable. Of course, by Theorem 1, the asymptotic distributions of  $\tau$  and  $\tau^*$  still coincide.

#### 4. Formal Edgeworth Expansions

In order to relate the results of the previous section to the conventional asymptotic theory of the bootstrap as found, for instance, in Hall (1992), and, in particular, to the theory presented in Liu (1988) and Mammen (1993), it may be helpful to develop formal Edgeworth expansions of the distribution of an HCCME-based pseudo- $t$  statistic and its wild bootstrap counterpart. Since this is an algebraically complicated undertaking, we present here the result from which we can draw conclusions, and relegate the complicated proof to an Appendix. In order to simplify some of the expressions, we assume for the purposes of this theorem that the  $v_t$  share the same distribution for all  $t$ .

**Theorem 3:** For the HCCME-based pseudo- $t$  statistics  $\tau$  considered in Theorem 1, and a wild bootstrap DGP with bootstrap error terms  $u_t^* = f(\ddot{u}_t)\varepsilon_t$ , where the  $\ddot{u}_t$  are the constrained or unconstrained residuals,

the transformation  $f$  corresponds to one of the  $HC_i$ ,  $i = 0, 1, 2, 3$ , and the  $\varepsilon_t$  are independent drawings from a distribution satisfying (10), the error in the rejection probability of a one-tailed bootstrap test at nominal level  $\alpha$  with rejection in the left-hand tail of the distribution, under a DGP of the form (6) with  $\beta_1 = 0$ , has the following formal expansion through order  $n^{-1}$ :

$$\begin{aligned} \phi(z_\alpha) & \left( \frac{1}{6}n^{-1/2}(1 - e_3^*)e_3C(1 + 2z_\alpha^2) + n^{-1} \left( e_3^2C^2\left(\frac{1}{6}z_\alpha - \frac{1}{9}z_\alpha^3 - \frac{1}{18}z_\alpha^5\right) \right. \right. \\ & + e_3^*e_3^2C^2\left(\frac{1}{6}z_\alpha + \frac{7}{18}z_\alpha^3 + \frac{1}{9}z_\alpha^5\right) + (e_3^*)^2e_3^2C^2\left(-\frac{1}{12}z_\alpha + \frac{2}{9}z_\alpha^3 - \frac{1}{18}z_\alpha^5\right) \\ & + \frac{1}{12}e_4D\left(z_\alpha(3(e_4^* - 1) - 2e_3^*) - z_\alpha^3((e_4^* - 1) + 4e_3^*)\right) \\ & \left. \left. + e_3^* \sum_i E_i F_i \left(\frac{1}{2}z_\alpha + z_\alpha^3\right) \right) \right). \end{aligned} \quad (26)$$

Here,  $e_3 = E(v_t^3)$ ,  $e_4 = E(v_t^4)$ ,  $e_3^* = E(\varepsilon_t^3)$ , and  $e_4^* = E(\varepsilon_t^4)$ . In addition,  $\phi(\cdot)$  is the standard normal density and  $z_\alpha$  is the  $\alpha$ -quantile of that distribution.

The other parameters in (26) depend on the  $\mathbf{X}$  matrix. Without loss of generality, we may choose the regressors in the matrix  $\mathbf{X}_2$  to be such that  $\mathbf{X}_2^\top \mathbf{X}_2 = n\mathbf{I}$ . Define the vector  $\mathbf{x}$ , with typical element  $x_t$ , to be  $\mathbf{M}_2 \mathbf{x}_1$  normalised so that  $\mathbf{x}^\top \mathbf{x} = n$ . Then the quantities  $C$ ,  $D$ ,  $E_i$ ,  $F_i$  in (26) are defined by

$$\begin{aligned} S^2 &= n^{-1} \sum_t x_t^2 \sigma_t^2, \quad C = S^{-3} n^{-1} \sum_t x_t^3 \sigma_t^3, \quad D = S^{-4} n^{-1} \sum_t x_t^4 \sigma_t^4, \\ E_i &= S^{-2} n^{-1} \sum_t x_t X_{ti} \sigma_t^2, \quad \text{and} \quad F_i = S^{-2} n^{-1} \sum_t x_t^3 X_{ti} \sigma_t^2. \end{aligned} \quad (27)$$

Here, if  $\mathbf{X} \equiv [\mathbf{x} \ \mathbf{X}_2]$ ,  $X_{ti}$  is the  $(t, i)$ <sup>th</sup> element of  $\mathbf{X}$ . The sum over  $i$  in (26) 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:** Note that, for the distribution  $F_1$  of (11), we have  $e_3^* = 1$ ,  $e_4^* = 2$ , and for  $F_2$  in (12),  $e_3^* = 0$  and  $e_4^* = 1$ .

The approximate error in rejection probability (ERP) in (26) provides an approximation to the distribution of the bootstrap  $P$  value  $p^*$ . Thus the same formula can be used to obtain approximations of ERPs for one-tailed tests in the other direction and two-tailed tests.

The leading-order term in (26) is clearly of order  $n^{-1/2}$  in general, the same as the leading-order term in the ERP of the asymptotic test. The calculations carried out in the proof of the theorem can also provide an expression for the ERP of the asymptotic test based on  $\tau$ . This expression is given after the proof of the theorem in the appendix. Interestingly enough, the contribution of order  $n^{-1/2}$  to this ERP is the *same* as that of the wild bootstrap test based on  $F_2$ . However,



simulation results in the next section will show that, even when the leading-order term in (26) does not vanish, the ERP of the bootstrap test is in many cases very much smaller than that of the asymptotic test.

The regressor design and the pattern of heteroskedasticity influence (26) through just three quantities,  $C$ ,  $D$ , and  $\sum_i E_i F_i$ .

With symmetric error terms,  $e_3 = 0$ , and (26) simplifies to  $n^{-1}\phi(z_\alpha)$  times

$$\frac{1}{12}e_4D(z_\alpha(3(e_4^* - 1) - 2e_3^*) - z_\alpha^3((e_4^* - 1) + 4e_3^*)) + e_3^* \sum_i E_i F_i (\frac{1}{2}z_\alpha + z_\alpha^3). \quad (28)$$

If in addition the Rademacher distribution  $F_2$  is used, it can be seen that (28) vanishes completely, so that the ERP is at most of order  $n^{-3/2}$ . This is compatible with the result of Theorem 2, and shows that we may expect good performance with symmetric errors and the  $F_2$  bootstrap even if regressors are present under the null.

With asymmetric error terms, the ERP also simplifies to  $n^{-1}\phi(z_\alpha)$  times (28) if  $C = 0$ . This refinement is the analogue for heteroskedastic models of the result in Hall (1992) according to which bootstrap tests on the coefficients of regression models benefit from refinements unless *both* the regressors and the errors are skewed. Here, as the definition of  $C$  in (27) makes clear, skewness of the regressors must be interpreted in the metric of the error covariance matrix.

With asymmetric error terms and the  $F_1$  bootstrap, the term of order  $n^{-1/2}$  vanishes, and the terms of order  $n^{-1}$  become  $n^{-1}\phi(z_\alpha)$  times

$$e_3^2 C^2 (\frac{1}{4}z_\alpha + \frac{1}{2}z_\alpha^3) + \frac{1}{12}e_4D(z_\alpha - 5z_\alpha^3) + e_3^* \sum_i E_i F_i (\frac{1}{2}z_\alpha + z_\alpha^3). \quad (29)$$

With asymmetric errors and the  $F_2$  bootstrap, the order  $n^{-1/2}$  term does not vanish, but the order  $n^{-1}$  terms simplify to  $n^{-1}\phi(z_\alpha)$  times

$$e_3^2 C^2 (\frac{1}{6}z_\alpha - \frac{1}{9}z_\alpha^3 - \frac{1}{18}z_\alpha^5). \quad (30)$$

The coefficients in (30), except for that of  $z_\alpha^5$ , are all smaller in absolute magnitude than those of the corresponding term in (29), which also has two more terms. This suggests, at least, that even with asymmetric errors, the  $F_2$  bootstrap will have a smaller order  $n^{-1}$  contribution to the ERP than the  $F_1$  bootstrap.

It follows from the Cauchy-Schwarz inequality that

$$(n^{-1} \sum_t |x_t^3| \sigma_t^3)^2 \leq (n^{-1} \sum_t x_t^2 \sigma_t^2) (n^{-1} \sum_t x_t^4 \sigma_t^4) = S^6 D \quad (31)$$

Provided that there is a constant in the regression,  $\sum_t x_t = 0$ , so that some of the  $x_t$  must be positive and some negative. Thus  $S^6 C^2$  is strictly less than the left-hand side of the inequality (31), and so  $C^2 < D$ . Similarly,  $e_3^2 < e_4$ . This observation suggests that, for  $n$  not too large, the order  $n^{-1/2}$  term in (26), being proportional to  $e_3 C$ , may be less important numerically than the order  $n^{-1}$  term, if the latter contains a contribution proportional to  $e_4 D$ . This suggestion will be borne out by our simulation results for sample sizes in the range of 10 – 100.

The expansion (26) does not go far enough for there to be any effect associated with the choice of HCCME. Such effects will however be clear in our simulation results.

## 5. Experimental Design and Simulation Results

It was shown by Chesher and Jewitt (1987) that HCCMEs are most severely biased when the regression design has observations with high leverage, and that the extent of the bias depends on the amount of heteroskedasticity in the true DGP. Since in addition one expects bootstrap tests to behave better in large samples than in small, in order to stress-test the wild bootstrap, most of our experiments are performed with a sample of size 10 containing one regressor, denoted  $\mathbf{x}_1$ , all the elements but one of which are independent drawings from  $N(0, 1)$ , but the second of which is 10, so as to create an observation with exceedingly high leverage. All the tests we consider are of the null hypothesis that  $\beta_1 = 0$  in the model (6), with  $k$ , the total number of regressors, varying across experiments. In all regression designs,  $\mathbf{x}_1$  is always present; for the design we designate by  $k = 2$  a constant, denoted  $\mathbf{x}_2$ , is also present; and for  $k = 3, \dots, 6$ , additional regressors  $\mathbf{x}_i$ ,  $i = 3, \dots, 6$  are successively appended to  $\mathbf{x}_1$  and  $\mathbf{x}_2$ . In Table 1, the components of  $\mathbf{x}_1$  are given, along with those of the  $\mathbf{x}_i$ ,  $i = 3, \dots, 6$ . In Table 2 are given the diagonal elements  $h_t$  of the orthogonal projections on to spaces spanned by  $\mathbf{x}_1$ ,  $\mathbf{x}_1$  and  $\mathbf{x}_2$ ,  $\mathbf{x}_1$ ,  $\mathbf{x}_2$ , and  $\mathbf{x}_3$ , *etc.* The  $h_t$  measure the leverage of the 10 observations for the different regression designs.

The data in all the simulation experiments discussed here are generated under the null hypothesis. Since (6) is a linear model, we set  $\beta_2 = \mathbf{0}$  without loss of generality. Thus our data are generated by a DGP of the form

$$y_t = \sigma_t v_t, \quad t = 1, \dots, n, \quad (32)$$

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  $t^{\text{th}}$  component of  $\mathbf{x}_1$ . 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 standardised.

In Table 3, we give, for the regression designs considered and the above pattern of heteroskedasticity, the values of the quantities  $C$ ,  $D$ , and  $\sum_i E_i F_i$  on which the approximate ERP (26) depends. The quantity denoted  $EF_0$  is the sum over  $i = 2, \dots, k$ , appropriate if constrained residuals are used in the HCCME,  $EF_1$  adds in the term for  $k = 1$ .

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, we show that it always leads to less distortion than the former in sample sizes up to 100. We also conduct a few experiments comparing the wild bootstrap and the  $(y, X)$  bootstrap. In order to conduct a fair comparison, we use an improved version of the  $(y, X)$  bootstrap suggested by Mammen (1993), and subsequently modified by Flachaire (1999), in which we resample, not the  $(y, X)$  pairs as such, but rather regressors ( $X$ ) and the constrained residuals, transformed according to  $HC_3$ . For the wild bootstrap,

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 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  $\alpha$ . Since we are considering a one-degree-of-freedom test, it is possible to perform a one-tailed test for which the rejection region is the set of values of the statistic algebraically greater than the critical value. We choose to look at one-tailed tests because Edgeworth expansions predict – see Hall (1992) – that the ERPs of one-tailed bootstrap tests converge to zero with increasing sample size more slowly than those of two-tailed tests. In any event, it is easy to compute the ERP of a two-tailed test with the information in the  $P$  value discrepancy plot. 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 high leverage, is the wild bootstrap capable of reducing the ERP relative to asymptotic tests?

Figure 2 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, 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  $\varepsilon_t$ , and the statistic that is bootstrapped is the  $HC_3$  form. To avoid redundancy, the plots are drawn only for the range  $0 \leq \alpha \leq 0.5$ , since, as is clear from (19), 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  $\alpha$  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  $k = 1$  satisfies the conditions of Theorem 2 when the errors are symmetric and the HCCME and the bootstrap DGP are based on constrained residuals. If we maintain all these conditions but consider the cases with  $k > 1$ , bootstrap inference is no longer perfect, but, according to the approximate theory of Section 4, should still be good. To what extent is this so? Do the design-dependent quantities given in Table 3 have any predictive power for the ERP?

$P$  value discrepancy plots are shown in Figure 3 for the designs  $k = 1, \dots, 6$  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. By what is presumably a coincidence induced by the specific form of the data, they are not at all large for  $k = 5$  or  $k = 6$ . Such a result might perhaps be predicted on the basis of Table 3, given

that the figures in the columns for  $k = 5$  and  $k = 6$  are smaller than in the other columns. In any case, we can conclude that the ERP does indeed depend on the regression design, but is not quantitatively very severe, given the small sample size.

- 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  $F_1$ ? What about the  $(y, X)$  bootstrap?

In Figure 4 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 for  $k > 2$ . Plots are also shown for the same designs and the preferred form of the  $(y, X)$  bootstrap. It is clear that the ERPs are quite different from those of the wild bootstrap, in either of its forms, and substantially greater.

- 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  $F_1$ ,  $F_2$ , and the  $(y, X)$  bootstrap?

Again we use the  $k = 3$  design. We see from Figure 5, which is like Figure 4 except that the errors are homoskedastic, that, with normal errors, the ERP is very slight with  $F_2$ , but remains significant for  $F_1$  and  $(y, X)$ . 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, which in turn is somewhat better than the  $(y, X)$  bootstrap.

- 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  $n = 20, 30, \dots$ . In this way, the design-dependent quantities like  $C$  and  $D$  do not depend on  $n$ . The plots shown in Figures 4 and 5 are repeated in Figure 6 for  $n = 100$ . The rankings found for  $n = 10$  remain unchanged, but, as suggested by the results of Section 4, 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 7, we plot the ERP for  $\alpha = 0.05$  as a function of  $n$ ,  $n = 10, 20, \dots$ , 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 (26) is antisymmetric with respect to  $z_\alpha$ , implying an ERP that is antisymmetric about  $\alpha = 0.5$ . We see from Figure 4 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  $\alpha$ . 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 8a, we plot the order  $n^{-1/2}$  and order  $n^{-1}$  contributions in (26) as functions of  $\alpha$ . For  $n = 10$ , the order  $n^{-1}$  term is clearly quantitatively greater than the order  $n^{-1/2}$  term. Then, in Figure 8b, we plot the differences between the approximate ERP (26) 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  $\mathbf{x}_1$  is the constant, and so  $h_t$  does not depend on  $t$ . For  $k > 2$ , significant differences appear, as seen in Figure 10 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?

Because the ERPs of the asymptotic tests are greater with a high leverage observation, we might expect the same to be true of bootstrap tests. In fact, although this is true if the  $HC_0$  statistic is used, the use of widely varying  $h_t$  with  $HC_3$  provides a good enough correction that, with it, the presence or absence of leverage has little impact. In Figure 10, 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  $\beta_1$ . In that case, the  $h_t$  differ more among themselves, since  $\mathbf{x}_1$  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 2 to hold, it is essential, Theorem 1 is agnostic on the point, and Theorem 3 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 3, where the difference between  $EF_1$  and  $EF_0$  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 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 never any worse behaved than the  $F_1$  version, or either version with unconstrained residuals, 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. This recommendation is supported by the results of simulation experiments designed to expose potential weaknesses of both versions, and also to some extent 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. This is not a problem for many econometric applications, for which hypothesis tests may be sufficient. In those cases in which 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 7 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 (1993). *Estimation and Inference in Econometrics*, Oxford University Press, New York.
- Davidson, R. and J. G. MacKinnon (1999). “The size distortion of bootstrap tests,” *Econometric Theory*, 15, 361–376.
- 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.
- Flachaire, E. (1999). “A better way to bootstrap pairs,” *Economics Letters*, 64, 257–262.

- Freedman, D. A. (1981). “Bootstrapping regression models,” *Annals of Statistics*, 9, 1218–1228.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, New York, Springer-Verlag.
- Horowitz, J. L. (1997). “Bootstrap Methods in Econometrics: Theory and Numerical Performance,” in David M. Kreps and Kenneth F. Wallis, eds., *Advances in Economics and Econometrics: Theory and Applications*, Vol. 3, pp. 188–222, Cambridge, Cambridge University Press.
- Kendall, M., and A. Stuart (1977). *The Advanced Theory of Statistics*, 4th ed., Vol. 1, “Distribution Theory,” London, Griffin.
- Liu, R. Y. (1988). “Bootstrap procedure 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 Theorem 3:

The proof is divided into several steps. First, we develop a formal stochastic expansion through order  $n^{-1}$  for any of the HCCME-based pseudo- $t$  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  $\alpha$ -quantile. This quantile is needed in order to express the condition  $\tau < Q(\alpha, \mu^*)$ , the probability of which is the rejection probability of the bootstrap test at nominal level  $\alpha$ . Then, because  $Q(\alpha, \mu^*)$  is random, the inequality 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  $\tau$ . Finally, the approximate bootstrap rejection probability, and hence also the ERP, is found by evaluating the Edgeworth expansion at the desired nominal level  $\alpha$ .

*Step 1: stochastic expansion of the statistic.*

In the proof of Theorem 1, it was seen that, under the null hypothesis, any of the statistics we consider can be written as

$$\tau = \sum_{t=1}^n (\mathbf{M}_2 \mathbf{x}_1)_t u_t / \left( \sum_{t=1}^n a_t (\mathbf{M}_2 \mathbf{x}_1)_t^2 \ddot{u}_t^2 \right)^{1/2}, \quad (33)$$

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,  $(\mathbf{M}_2 \mathbf{x}_1)_t = x_t$ . If constrained residuals are used, the vector  $\tilde{\mathbf{u}}$  with typical element  $\tilde{u}_t$  is

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

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

$$\tilde{u}_t = \sigma_t v_t - n^{-1} \sum_{i=2}^k X_{ti} \sum_{s=1}^n X_{si} \sigma_s v_s. \quad (34)$$

(Recall that the  $X_{ti}$  are defined so that, for  $i > 1$ ,  $X_{ti}$  is the element of  $\mathbf{X}_2$  in the  $t^{\text{th}}$  row and the  $(i-1)^{\text{th}}$  column.) If unconstrained residuals  $\hat{u}_t$  are used, (34) 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  $\tilde{u}_t$  for the residuals.

Let us make the definition

$$w_i \equiv n^{-1/2} S^{-1} \sum_{t=1}^n X_{ti} \sigma_t v_t, \quad i = 1, \dots, k; \quad (35)$$

recall the definition of  $S$  in (27). By the central limit theorem, the  $w_i$  are asymptotically normal, mean zero, and  $O(1)$ . The numerator of the statistic (33), divided by  $n^{1/2}$ , is just  $Sw_0$ . For the denominator, we need the stochastic expansion of the residuals, which from (34) and (35) is

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

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

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

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

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 (33), also divided by  $n^{1/2}$ , is  $S$  times

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

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

$$\tau_1 \stackrel{a}{=} w_0 \left(1 - \frac{1}{2} n^{-1/2} q + n^{-1} \sum_i w_i w_{ai} - \frac{1}{2} n^{-1} H - \frac{1}{2} n^{-1} \sum_i \sum_j A_{ij} w_i w_j + \frac{3}{8} n^{-1} q^2\right). \quad (36)$$

A wild bootstrap statistic is defined by the same formula (33) as  $\tau$  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

$$s_t = a_t^{1/2} (\sigma_t v_t - n^{-1/2} S \sum_i X_{ti} w_i), \quad (37)$$

or possibly the absolute value of that expression. Since a realisation  $\hat{\mu}$  of the bootstrap DGP generates data conditional on the realised  $v_t$ , the only random elements in a drawing from  $\hat{\mu}$  are the  $\varepsilon_t$ , just as the random elements in a drawing from the true DGP are the realised  $v_t$ . 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 DGP (conditional on the  $v_t$ ) is given by (36), with all the variables redefined with  $s_t$  and  $\varepsilon_t$  in place of  $\sigma_t$  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  $\tau$  can be written as

$$F(x) = \Phi(x) - n^{-1/2} \phi(x) \sum_{i=1}^{\infty} \lambda_i H e_{i-1}(x). \quad (38)$$

Here  $\Phi(\cdot)$  and  $\phi(\cdot)$  are respectively the CDF and the density of the  $N(0, 1)$  distribution,  $He_i(\cdot)$  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 (38) is more properly referred to as the Gram-Charlier series, but, unless truncated, the Edgeworth and Gram-Charlier series are equivalent. As we truncate everything in this proof of order lower than  $n^{-1}$ , we obtain true Edgeworth series. The  $\lambda_i$  in (38) are coefficients that are at most of order unity, defined by the relations

$$\lambda_j = \frac{n^{1/2}}{j!} E(He_j(\tau)), \quad (39)$$

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  $\mu_i$  is the uncentred moment of  $\tau$  of order  $i$ .

The leading-order term of the stochastic expansion  $\tau_1$  in (36) is  $w_0$ , which is a normalised sum of mean-zero variables that converges to the  $N(0, 1)$  distribution as  $n \rightarrow \infty$  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  $E(He_j(w_0))$  is equal to the “formal moment” of order  $j$  corresponding to a sequence of “formal cumulants,”  $\kappa_i$ , say, where  $\kappa_j$  is the  $j^{\text{th}}$  cumulant of  $w_0$ , except for  $j = 2$ , for which  $\kappa_2$  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

$$\xi \equiv -\frac{1}{2}w_0q + n^{-1/2}w_0\left(\sum_i w_i w_{ai} - \frac{1}{2}\sum_i \sum_j A_{ij}w_i w_j - \frac{1}{2}H + \frac{3}{8}q^2\right). \quad (40)$$

For  $j = 1$ , we find from (39) that  $\lambda_1 = n^{1/2}E(w_0 + n^{-1/2}\xi) = E(\xi)$ . To compute  $E(\xi)$ , note that

$$E(w_0q) = e_3(nS^3)^{-1} \sum_{t=1}^n a_t x_t^3 \sigma_t^3 = e_3C + O(n^{-1})$$

The expectations of all the terms in  $\xi$  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  $\lambda_1 = -\frac{1}{2}e_3C + O(n^{-1})$ .

In order to compute  $\lambda_2$ , we note that

$$\begin{aligned}\tau_1^2 &= w_0^2 + 2n^{-1/2}w_0\xi + n^{-1}\xi^2 \\ &= w_0^2 - n^{-1/2}w_0^2q + n^{-1}w_0^2\left(q^2 + 2\sum_i w_iw_{ai} - \sum_i \sum_j A_{ij}w_iw_j - H\right)\end{aligned}$$

Now we have

$$\begin{aligned}E(w_0^2q) &= n^{-1/2}D(e_4 - 1), & E(w_0^2w_iw_j) &= B_{ij} + 2E_iE_j, \\ E(w_0^2q^2) &= D(e_4 - 1) + 2e_3^2C^2, & E(w_0^2w_iw_{ai}) &= G_i + 2E_iF_i,\end{aligned}$$

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

$$\begin{aligned}\lambda_2 &= \frac{1}{2}n^{1/2}E(\tau_1^2 - 1) = n^{-1/2}\left(e_3^2C^2 + \sum_i (G_i + 2E_iF_i)\right. \\ &\quad \left. - \frac{1}{2}\sum_i \sum_j A_{ij}(B_{ij} + 2E_iE_j) - \frac{1}{2}H\right) + O(n^{-1}).\end{aligned}$$

Similar calculations, of which we skip the details, show that  $\lambda_3 = -\frac{1}{3}e_3C + O(n^{-1})$ , and

$$\lambda_4 = n^{-1/2}\left(-\frac{1}{12}e_4D + \frac{2}{3}e_3^2C^2 + \sum_i E_iF_i - \frac{1}{2}\sum_i \sum_j A_{ij}E_iE_j\right) + O(n^{-1}).$$

Finally,  $\lambda_6$  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  $6n^{-1/2}\lambda_3 = -2n^{-1/2}e_3C$ , and so  $\lambda_6 = \frac{1}{18}n^{-1/2}e_3^2C^2$ .

For the wild bootstrap statistic, we define coefficients  $\lambda_j^*$  by the same formulas as those for the  $\lambda_j$ , but with  $s_t$  instead of  $\sigma_t$  and  $\varepsilon_t$  instead of  $v_t$  in the definitions. As with the  $\lambda_j$ , 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 (37) 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

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

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

$$\lambda_1^* = -\frac{1}{2} e_3^* C^* = -\frac{1}{2} e_3^* e_3 C + n^{-1/2} e_3^* (\frac{3}{4} q e_3 C - \frac{1}{2} c + \frac{3}{2} \sum_i w_i F_i), \quad (41)$$

$$\lambda_3^* = \frac{2}{3} \lambda_1^*, \quad \text{and} \quad \lambda_6^* = \frac{1}{18} n^{-1/2} (e_3^*)^2 e_3^2 C^2.$$

Since  $\lambda_2^*$  and  $\lambda_4^*$  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  $i, j = 1, \dots, k$ ,  $B_{ij}^* = B_{ij}$ ,  $E_i^* = E_i$ ,  $F_i^* = F_i$ , and  $G_i^* = G_i$ . Since  $A_{ij}$  and  $H$ , depend neither on the  $\sigma_t$  nor on the  $v_t$ , they are the same for the true and the bootstrap DGP. We thus see that, to order  $n^{-1/2}$ ,

$$\lambda_2^* = n^{-1/2} \left( (e_3^*)^2 e_3^2 C^2 + \sum_i (G_i + 2E_i F_i) - \frac{1}{2} \sum_i \sum_j A_{ij} (B_{ij} + 2E_i E_j) - \frac{1}{2} H \right),$$

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

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

For the CDF (38), the  $\alpha$ -quantile is defined implicitly by the equation  $F(x_\alpha) = \alpha$ . An expansion of  $x_\alpha$  in powers of  $n^{-1/2}$ , usually called the Cornish-Fisher expansion, can be obtained by inverting the definition (38) of  $F$ . The result is

$$x_\alpha = z_\alpha + n^{-1/2} \sum_i \lambda_i H e_{i-1}(z_\alpha) + \frac{1}{2} n^{-1} \sum_i \sum_j \lambda_i \lambda_j h_{ij}(z_\alpha) + O(n^{-3/2}), \quad (42)$$

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 (42) for the sequence  $\{\lambda_i\} \equiv \{\lambda_1, \lambda_2, \dots\}$ . Let  $l_i \equiv \lambda_i^* - \lambda_i$ . For a one-tailed bootstrap test at nominal level  $\alpha$ , 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

$$q_\alpha^* = \sum_i l_i H e_{i-1}(z_\alpha) + \frac{1}{2} n^{-1/2} \sum_i \sum_j (l_i \lambda_j + \lambda_i l_j) h_{ij}(z_\alpha) + O(n^{-1}), \quad (43)$$

Write  $Q_\alpha = Q_\alpha(\{\lambda_i\})$ ,  $\nu_i = E_{\mu_0}(l_i)$ , and let  $q_\alpha = E_{\mu_0}(q_\alpha^*)$ . We have

$$q_\alpha = \sum_i \nu_i H e_{i-1}(z_\alpha) + \frac{1}{2} n^{-1/2} \sum_i \sum_j (\nu_i \lambda_j + \lambda_i \nu_j) h_{ij}(z_\alpha) + O(n^{-1}). \quad (44)$$

Comparison with (43) shows that, through order  $n^{-1}$ ,  $Q_\alpha + n^{-1/2}q_\alpha$  is the  $\alpha$ -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 (38) 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

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

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

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

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

$$n^{-1/2}\phi(Q_\alpha + n^{-1/2}q_\alpha) \sum_i (\nu_i - \eta_i) H e_{i-1}(Q_\alpha + n^{-1/2}q_\alpha). \quad (45)$$

*Step 4: Computation of the bootstrap RPE.*

We begin with the computation of the  $\nu_i$ . From (41), we see that, through order  $n^{-1/2}$ ,

$$\nu_1 = E(\lambda_1^* - \lambda_1) = -\frac{1}{2}(e_3^* - 1)e_3C,$$

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

$$\begin{aligned} \nu_3 &= -\frac{1}{3}(e_3^* - 1)e_3C, & \nu_6 &= \frac{1}{18}n^{-1/2}((e_3^*)^2 - 1)e_3^2C^2, \\ \nu_2 &= n^{-1/2}((e_3^*)^2 - 1)e_3^2C^2, & \nu_4 &= n^{-1/2}\left(-\frac{1}{12}(e_4^* - 1)e_4D + \frac{2}{3}((e_3^*)^2 - 1)e_3^2C^2\right). \end{aligned}$$

Since  $\lambda_2^*$ ,  $\lambda_4^*$ , and  $\lambda_6^*$  are nonrandom through order  $n^{-1/2}$ , we see that  $l_2 = \nu_2$ ,  $l_4 = \nu_4$ , and  $l_6 = \nu_6$ . Thus, from (43) and (44), we have that

$$\begin{aligned}\gamma_\alpha = q_\alpha^* - q_\alpha &= l_1 - \nu_1 + (l_3 - \nu_3)(z_\alpha^2 - 1) \\ &+ n^{-1/2}(l_1 - \nu_1) \sum_j \lambda_j h_{1j}(z_\alpha) + n^{-1/2}(l_3 - \nu_3) \sum_j \lambda_j h_{3j}(z_\alpha) + O(n^{-1}),\end{aligned}\quad (46)$$

Define the zero mean random variable

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

We see from this that we need only the first two terms in the expression (46) for  $\gamma_\alpha$ , since the last two are  $O(n^{-1})$ . To the desired order,

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

For the  $\eta_i$ , we note first that  $\lambda_1 + \eta_1 = n^{1/2} E(\tau - n^{-1/2} \gamma_\alpha) = n^{1/2} E(\tau) = \lambda_1$ , so that  $\eta_1 = 0$ . For  $\eta_2$ , since  $\gamma_\alpha = O(n^{-1/2})$ , we have

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

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

$$\begin{aligned}\eta_2 &= -E(\tau \gamma_\alpha) = -E(w_0 \gamma_\alpha) + O(n^{-1}) = -n^{-1/2} (1 + 2z_\alpha^2) E(w_0 \zeta) + O(n^{-1}) \\ &= -n^{-1/2} (1 + 2z_\alpha^2) e_3^* \left( \frac{1}{4} e_3^2 C^2 - \frac{1}{6} e_4 D + \frac{1}{2} \sum_i E_i F_i \right) + O(n^{-1}),\end{aligned}$$

since, as we have already seen,  $E(w_0 q) = e_3 C$ , and, as can easily be checked,  $E(w_0 c) = e_4 D$  and  $E(w_0 w_i) = E_i$ . For  $\eta_3$ , we compute

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

But  $E(w_0^2 q)$ ,  $E(w_0^2 c)$ , and  $E(w_0^2 w_i)$  are all  $O(n^{-1/2})$ , and so  $\eta_3 = O(n^{-1})$ . For  $\eta_4$ , we find that

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

Now it can be checked that  $E(w_0^3 \zeta) = 3E(w_0 \zeta)$ , since  $E(w_0^3 q) = 3e_3 C$ , with similar results for  $E(w_0^3 c)$  and  $E(w_0^3 w_i)$ . Thus  $\eta_4 = O(n^{-1})$ . Since in general  $\lambda_6$  is through order  $n^{-1}$  a function of  $\lambda_3$ , and since  $\eta_3 = O(n^{-1})$ , it follows that  $\eta_6 = O(n^{-1})$ .

We now return to the evaluation of (45). Since to relevant order  $Q_\alpha + n^{-1/2} q_\alpha$  is the  $\alpha$ -quantile of the distribution characterised by  $\{\lambda_i + \nu_i\}$ , we obtain from (42) that

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

Performing a Taylor expansion of (45) about  $z_\alpha$  gives for the ERP of the bootstrap test

$$\begin{aligned} & n^{-1/2}\phi(z_\alpha) \sum_i (\nu_i - \eta_i) H e_{i-1}(z_\alpha) \\ & - n^{-1}\phi(z_\alpha) \sum_j (\lambda_j + \nu_j) H e_{j-1}(z_\alpha) \sum_i (\nu_i - \eta_i) H e_i(z_\alpha) + o(n^{-1}), \end{aligned} \quad (47)$$

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

$$\sum_j (\lambda_j + \nu_j) H e_{j-1}(z_\alpha) = -e_3^* e_3 C \left( \frac{1}{2} + \frac{1}{3}(z_\alpha^2 - 1) \right) = -\frac{1}{6} e_3^* e_3 C (1 + 2z_\alpha^2).$$

With this, the ERP (47) is

$$n^{-1/2}\phi(z_\alpha) \sum_i (\nu_i - \eta_i) \left( H e_{i-1}(z_\alpha) + \frac{1}{6} n^{-1/2} e_3^* e_3 C (1 + 2z_\alpha^2) H e_i(z_\alpha) \right). \quad (48)$$

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

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

in accord with (26). The contribution of order  $n^{-1}$  from these same two terms is

$$\frac{1}{36} n^{-1} \phi(z_\alpha) ((e_3^*)^2 - e_3^*) e_3^2 C^2 (3z_\alpha + 4z_\alpha^3 - 4z_\alpha^5). \quad (49)$$

For  $i = 2$ , the leading-order contribution to (48) is  $n^{-1}\phi(z_\alpha)$  times

$$z_\alpha \left( ((e_3^*)^2 - 1) e_3^2 C^2 + (1 + 2z_\alpha^2) e_3^* \left( \frac{1}{4} e_3^2 C^2 - \frac{1}{6} e_4 D + \frac{1}{2} \sum_i E_i F_i \right) \right). \quad (50)$$

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

$$(z_\alpha^3 - 3z_\alpha) \left( -\frac{1}{12} (e_4^* - 1) e_4 D + \frac{2}{3} ((e_3^*)^2 - 1) e_3^2 C^2 \right), \quad (51)$$

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

$$(z_\alpha^5 - 10z_\alpha^3 + 15z_\alpha) \frac{1}{18} ((e_3^*)^2 - 1) e_3^2 C^2. \quad (52)$$

Adding up the contributions (49), (50), (51), and (52) yields the term of order  $n^{-1}$  in (26). ■

The rejection probability of the asymptotic test based on  $\tau$  (one-tailed with rejection in the left-hand tail) is the probability mass to the left of  $z_\alpha$  in the distribution with expansion (38). Thus, with the values of the  $\lambda_i$  computed here, the ERP is  $n^{-1/2}\phi(z_\alpha)$  times

$$\begin{aligned} & \frac{1}{6} e_3 C (1 + 2z_\alpha^2) + n^{-1/2} \left( \frac{1}{2} H z_\alpha + e_3^2 C^2 \left( \frac{1}{6} z_\alpha - \frac{1}{9} z_\alpha^3 - \frac{1}{18} z_\alpha^5 \right) + \frac{1}{12} e_4 D (z_\alpha^3 - 3z_\alpha) \right. \\ & \left. + \left( \frac{1}{2} \sum_i \sum_j A_{ij} E_i E_j - \sum_i E_i F_i \right) (z_\alpha^3 - z_\alpha) + \left( \frac{1}{2} \sum_i \sum_j A_{ij} B_{ij} - \sum_i G_i \right) z_\alpha \right). \end{aligned}$$



**Table 1. Regressors**

Obs	$x_1$	$x_3$	$x_4$	$x_5$	$x_6$
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

**Table 2. Leverage measures**

Obs	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$
1	0.003537	0.101022	0.166729	0.171154	0.520204	0.560430
2	0.930524	0.932384	0.938546	0.938546	0.964345	0.975830
3	0.003357	0.123858	0.128490	0.137478	0.164178	0.167921
4	0.003497	0.124245	0.167158	0.287375	0.302328	0.642507
5	0.036190	0.185542	0.244940	0.338273	0.734293	0.741480
6	0.001562	0.102785	0.105276	0.494926	0.506885	0.880235
7	0.004260	0.126277	0.138399	0.143264	0.295007	0.386285
8	0.001490	0.102888	0.154378	0.162269	0.163588	0.218167
9	0.011385	0.100300	0.761333	0.879942	0.880331	0.930175
10	0.004197	0.100698	0.194752	0.446773	0.468841	0.496971

Notes: For  $k = 1$ , the only regressor is  $x_1$ , for  $k = 2$  there is also the constant, for  $k = 3$  there are the constant,  $x_1$ , and  $x_2$ , and so forth.

**Table 3. Influence of the design on Edgeworth expansion**

	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$
$C$	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
$EF_0$	0.00	0.98	6.67	6.98	1.71	1.02
$EF_1$	9.29	9.27	8.31	8.17	3.94	2.90

Note:  $EF_0$  is  $\sum_i E_i F_i$  for the case of constrained residuals;  $EF_1$  is for unconstrained residuals.

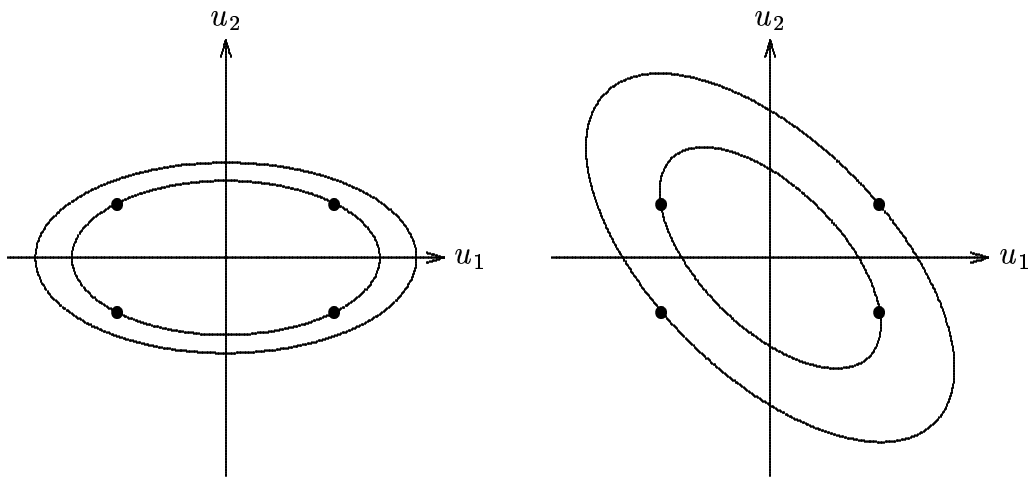


Figure 1. Absolute values and signs of two random variables

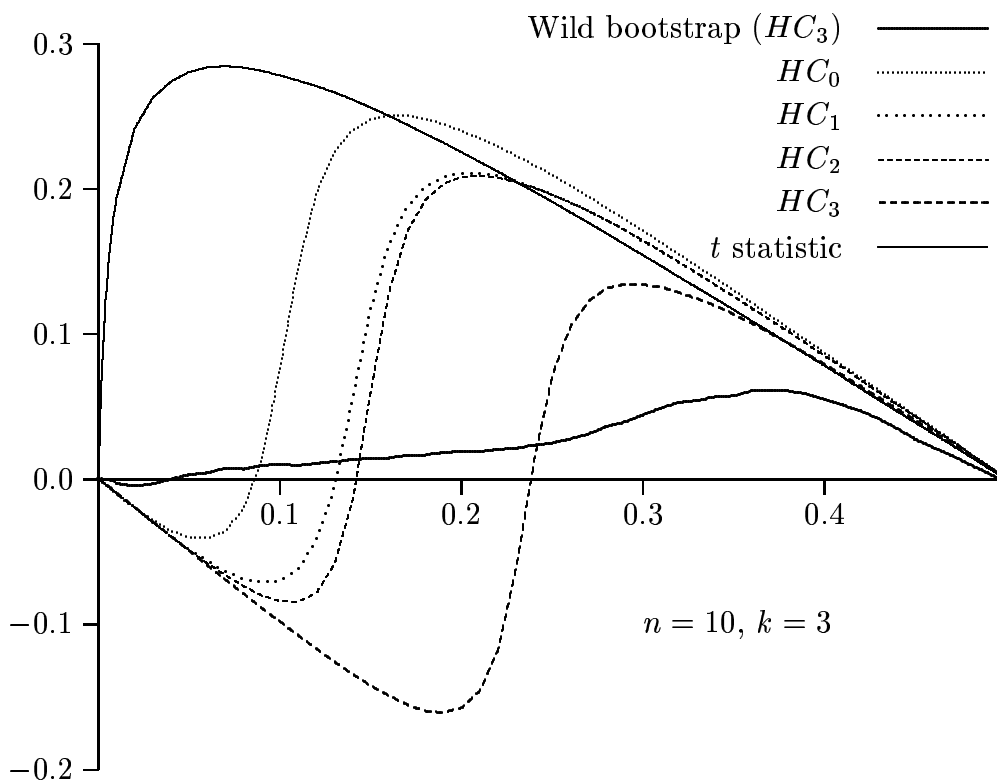


Figure 2. ERPs of asymptotic and bootstrap tests

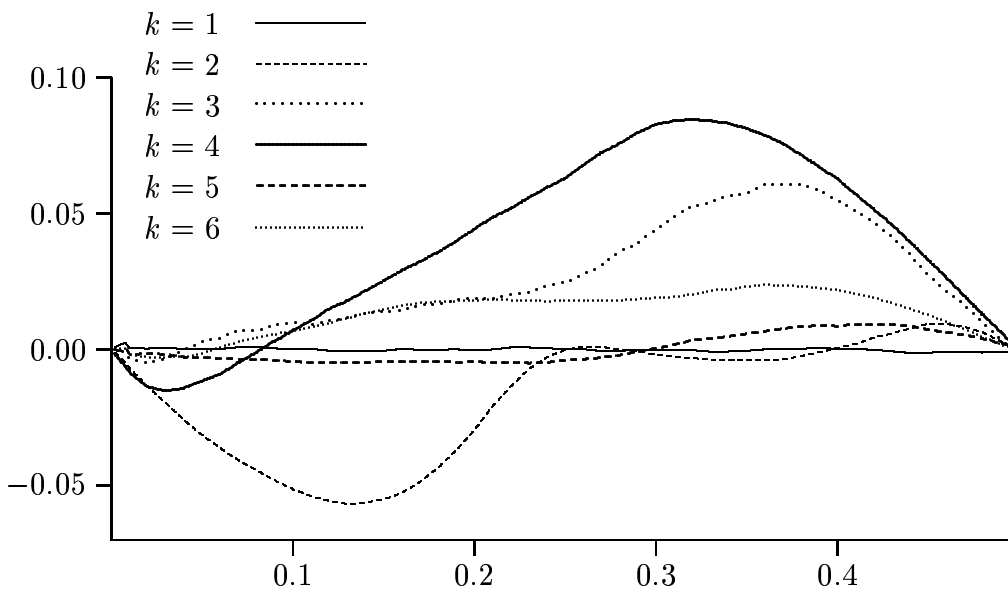


Figure 3. Base case with different designs

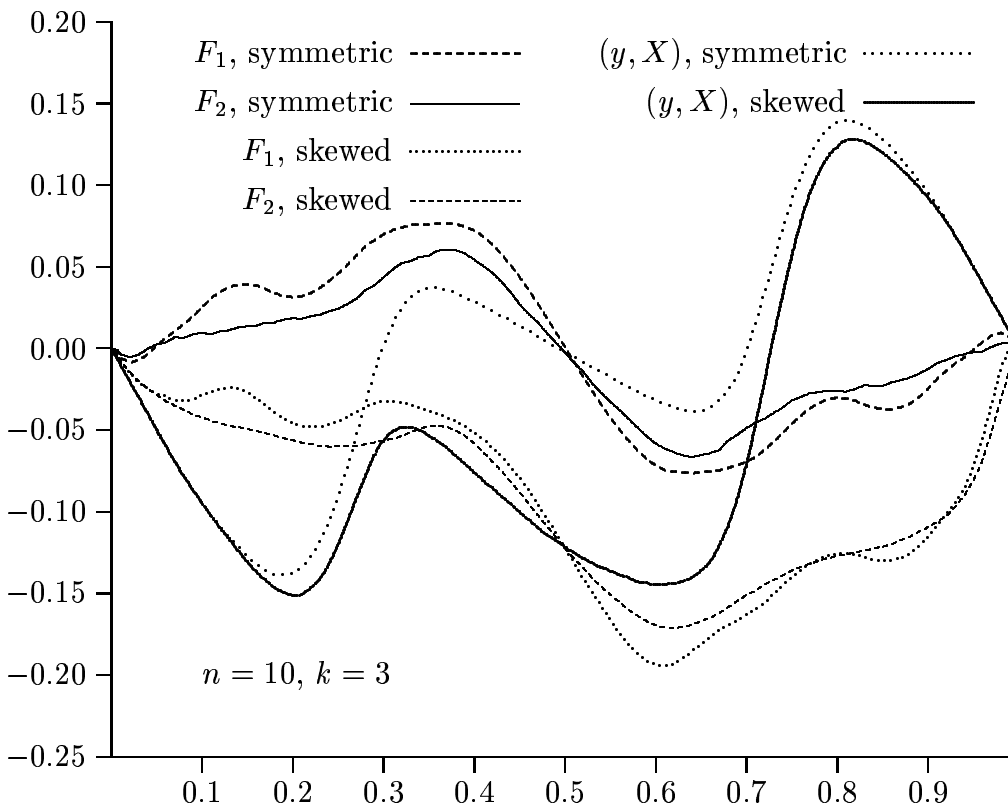
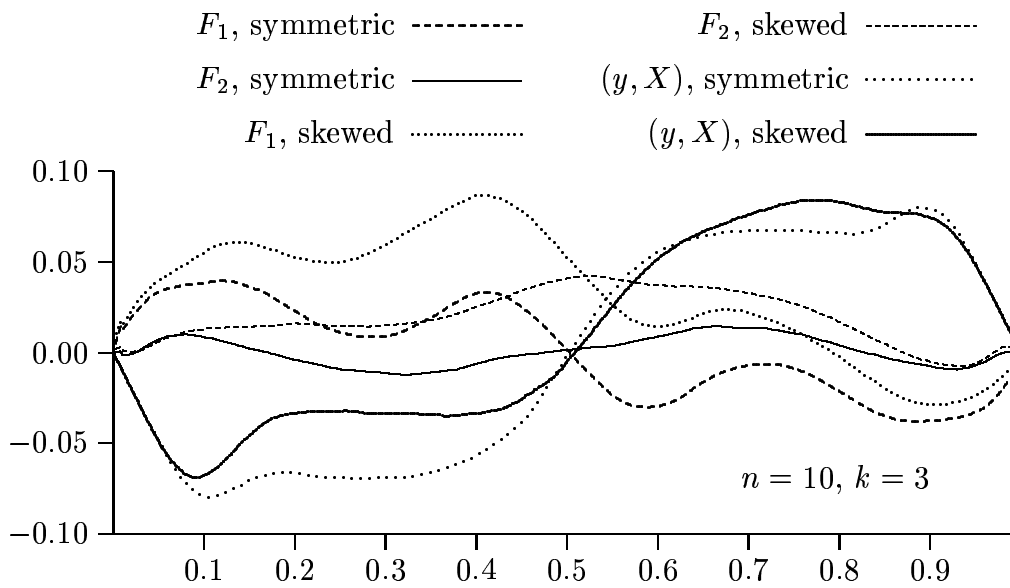
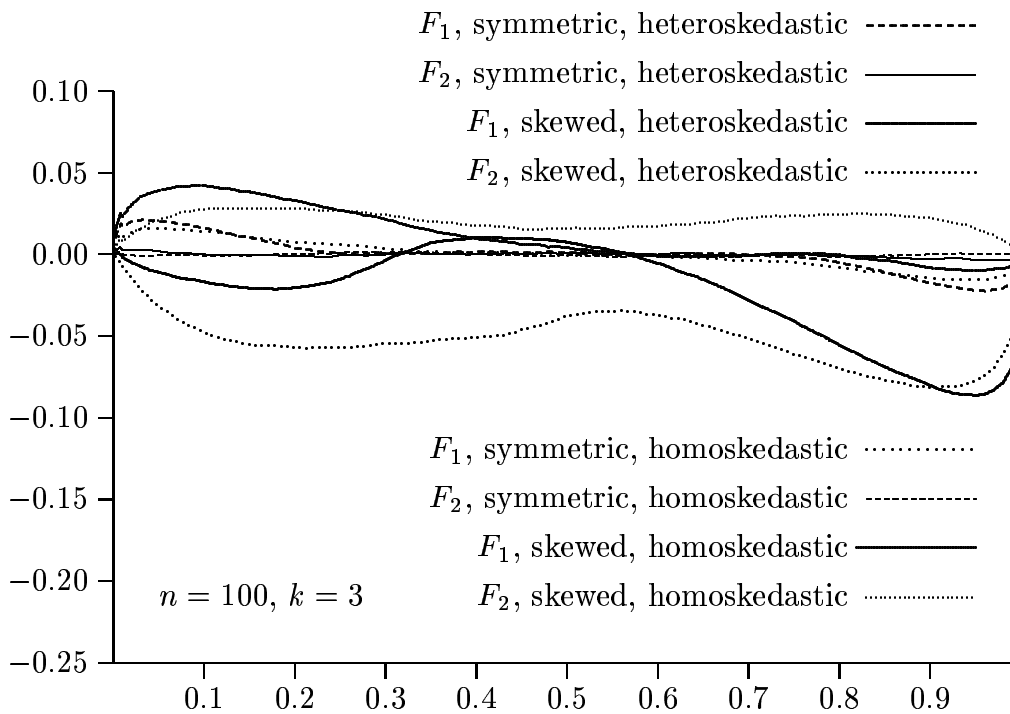


Figure 4. Symmetric and skewed errors,  $F_1$ ,  $F_2$ , and  $(y, X)$  bootstraps



**Figure 5. Homoskedastic errors**



**Figure 6. ERPs for  $n = 100$**

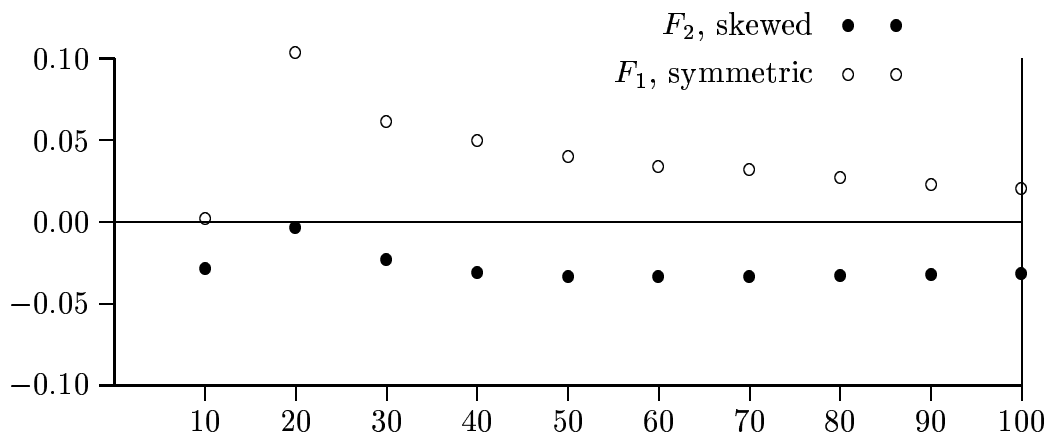


Figure 7. ERP for nominal level 0.05 as function of sample size

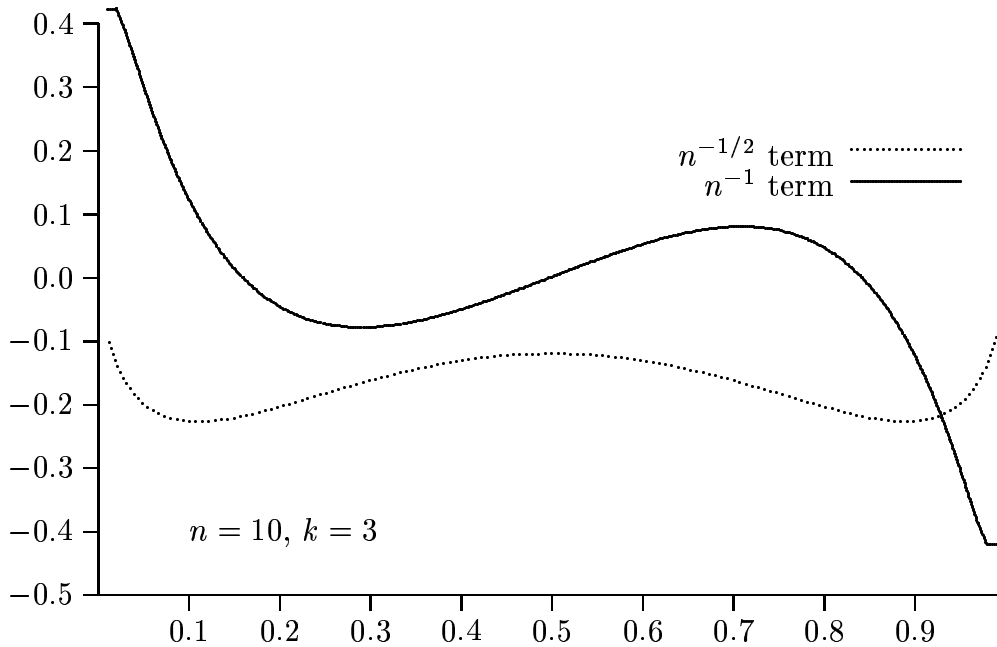


Figure 8a. The contributions of order  $n^{-1/2}$  and  $n^{-1}$  to the ERP

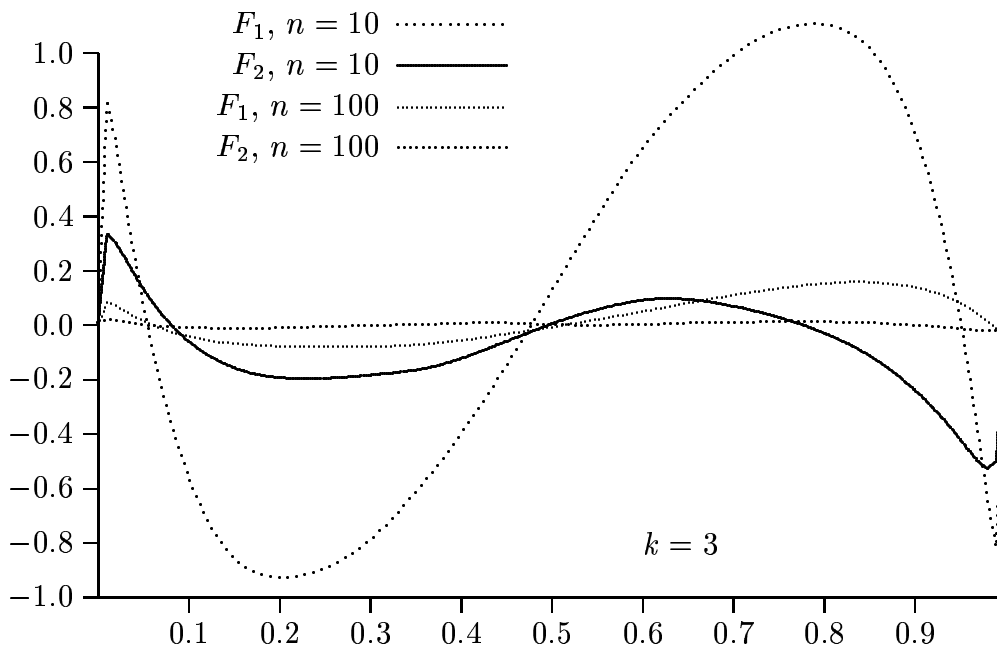


Figure 8b. Error in Edgeworth ERP,  $F_1$ ,  $F_2$ , skewed errors

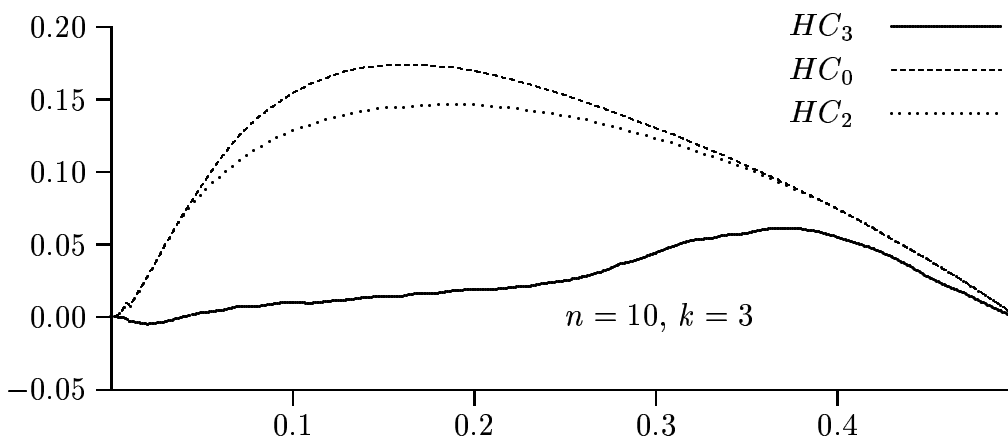
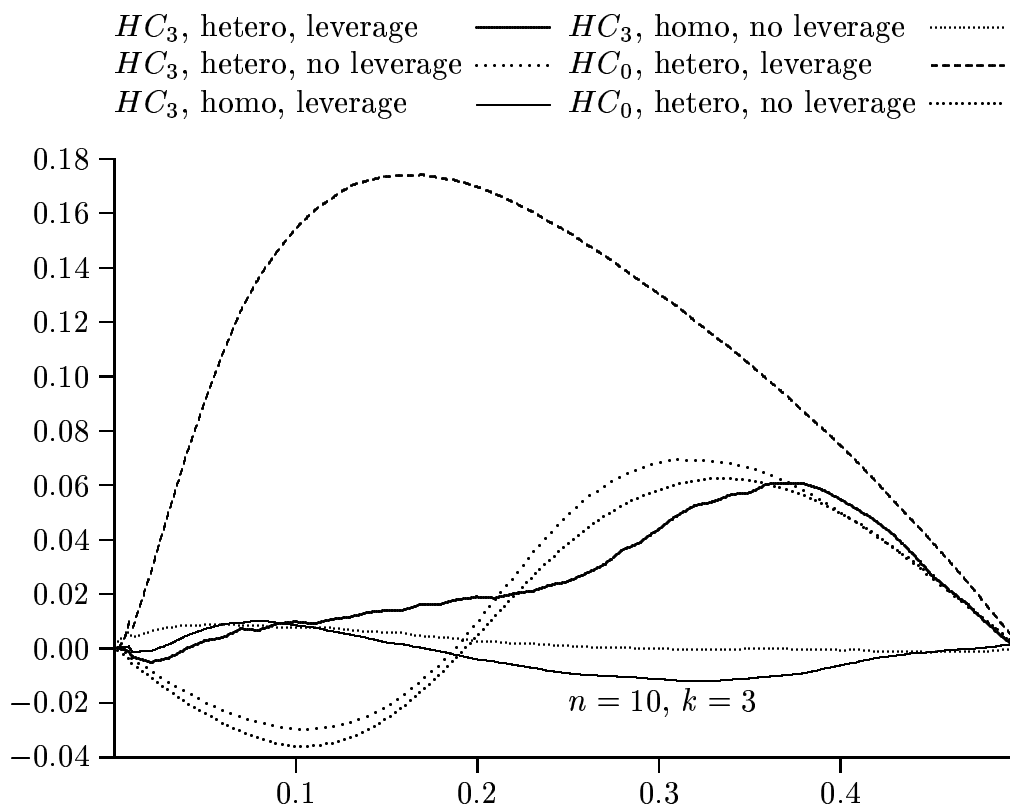


Figure 9.  $HC_3$  compared with  $HC_0$  and  $HC_2$



**Figure 10. Relative importance of leverage and heteroskedasticity**