

Implementing the Wild Bootstrap using a Two-Point Distribution

James Davidson*
University of Exeter

Andrea Monticini
University of Exeter

David Peel
University of Lancaster

June 2006

Keywords: Wild Bootstrap, two-point distribution, heteroscedsticity, skewness
JEL Classification: C15

Abstract

We consider the problem of selecting the auxiliary distribution to implement the wild bootstrap for regressions featuring heteroscedasticity of unknown form. Asymptotic refinements are nominally obtained by choosing a distribution with second and third moments equal to 1. We show that this stipulation may fail in practice, due to the distortion imposed on higher moments. We propose a new class of two-point distributions and suggest using the Kolmogorov-Smirnov statistic as a selection criterion. The results are illustrated by a Monte Carlo experiment.

*Corresponding author, james.davidson@exeter.ac.uk. We thank Bernard Pearson for helpful discussion.

1 Introduction

The wild bootstrap is a variant of the bootstrap method for application to data which are not i.i.d. and which, in particular, are heteroscedastic; see Wu (1986), Beran (1986), Liu (1988), Mammen (1993), Davidson and Flachaire (2001), and for a recent empirical application, Paya and Peel (2006).

Consider a regression model

$$y_t = \boldsymbol{\beta}' \mathbf{x}_t + u_t$$

where \mathbf{x}_t are fixed in repeated samples and $u_t \sim \text{iid}(0, \sigma_t^2)$, with $\sigma_t^2 \neq \sigma_s^2$ for $t \neq s$, in general. Letting $\hat{\boldsymbol{\beta}}$ denote the OLS estimator, our object is to construct a bootstrap analogue for the distribution of

$$\sqrt{n} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sqrt{n} \left(\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \sum_{t=1}^n \mathbf{x}_t u_t$$

whose variance matrix, in particular, is

$$nE(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' = n \left(\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \sum_{t=1}^n \sigma_t^2 \mathbf{x}_t \mathbf{x}_t' \left(\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \right)^{-1}. \quad (1.1)$$

Let \hat{u}_t^* denote a random resampling of the rescaled¹ least squares residuals, such that

$$P \left(\hat{u}_t^* = \sqrt{\frac{n}{n-k}} \hat{u}_t \right) = n^{-1}$$

and let

$$\sqrt{n} (\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}}) = \sqrt{n} \left(\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \sum_{t=1}^n \mathbf{x}_t \hat{u}_t^*$$

denote the bootstrap analogue of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Letting E_* denotes the expected value under the bootstrap distribution, it is easily established that $E_*(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}}) = \mathbf{0}$, but if s^2 denotes the usual unbiased residual variance estimator then

$$\begin{aligned} nE_*(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}})' &= ns^2 \left(\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \\ &= \left(\sum_{t=1}^n \hat{u}_t^2 \right) \left(\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \\ &= \left(\sum_{t=1}^n \sigma_t^2 \right) \left(\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \right)^{-1} + O_p(n^{-1/2}) \end{aligned} \quad (1.2)$$

Note that the implicit variance estimator is inconsistent for (1.1), so that the bootstrap fails in this case. However, letting (η_1, \dots, η_n) be random drawings from a distribution having $E\eta_t = 0$ and $E\eta_t^2 = 1$, and independent of $(\hat{u}_1, \dots, \hat{u}_n)$, so-called ‘wild bootstrap’ replicates have the form

$$\hat{u}_t^{**} = \sqrt{\frac{n}{n-k}} \hat{u}_t \eta_t.$$

Letting E_{**} denote expected values under the wild bootstrap distribution, note that

$$\begin{aligned} nE_{**}(\hat{\boldsymbol{\beta}}^{**} - \hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}^{**} - \hat{\boldsymbol{\beta}})' &= \frac{n^2}{n-k} \left(\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \sum_{t=1}^n \hat{u}_t^2 \mathbf{x}_t \mathbf{x}_t' \left(\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \\ &= nE(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' + O_p(n^{-1/2}). \end{aligned}$$

¹This is one of several possible transformations of the residuals - see Davidson and Flachaire (2001).

Thus, under the usual regularity conditions this distribution is consistent for the sampling distribution of $\sqrt{n}(\hat{\beta} - \beta)$.

The regular bootstrap still provides asymptotically valid t tests, because although the variance in (1.2) does not match the sampling distribution the bootstrap t -ratios are nonetheless asymptotically $N(0, 1)$. However, the coverage probabilities of confidence intervals are asymptotically biased, and asymptotic refinements of the error in rejection probability (ERP) are not attained.

As well as asymptotically valid confidence intervals, the wild bootstrap has been shown (Liu 1988, Davidson and Flachaire 2001) to yield asymptotic refinements in the distributions of pivotal statistics. The key fact is that if $E\eta_t^3 = 1$, then $E\hat{u}_t^{**3} = E\hat{u}_t^3$ in view of the independence of the components. Agreement of the third moments of the bootstrap shocks with that of the parent distribution of the rescaled residuals means that the first-order terms in the Edgeworth expansions of an asymptotically pivotal statistic in the two cases agree likewise, with a corresponding reduction of ERP.

By the same token, if we could arrange for additional higher moments of η to equal 1, then we should correspondingly match the higher moments of \hat{u}_t^{**} and \hat{u}_t , leading to additional refinements. Unfortunately, no distributions with this desirable property exist, in view of the inequality

$$E\eta^4 \geq 1 + (E\eta^3)^2 \quad (1.3)$$

(Pearson, 1916). However, note that if $E\hat{u}_t^3 = 0$, then $E(\hat{u}_t^{**})^3 = 0$ regardless of the value of $E\eta_t^3$. In this case, if we could arrange to have $E\eta_t^4 = 1$, which is possible according to (1.3), then $E\hat{u}_t^{**4} = E\hat{u}_t^4$ in addition to $E\hat{u}_t^{**3} = E\hat{u}_t^3$, implying agreement of the second-order terms of the expansions. For example, it is shown by Davidson and Flachaire that in the first of these cases, the leading term in the asymptotic expansion of the ERP of a 1-tailed t test is of $O(n^{-1})$. In the second case, however, the leading term of $O(n^{-1})$ in the expansion also vanishes.

2 Choice of the Auxiliary Distribution

A number of distributions can be considered to play the role of generation process for the η_t , fulfilling one or more of the requirements detailed above. Liu (1988) and Mammen (1993) suggest alternative schemes to meet the requirement $E\eta_t^3 = 1$, of which the most widely adopted appears to be the two-point distribution,

$$\eta_A = \begin{cases} \frac{1 + \sqrt{5}}{2} & \text{with probability } p = \frac{\sqrt{5} - 1}{2\sqrt{5}} \\ \frac{1 - \sqrt{5}}{2} & \text{with probability } 1 - p. \end{cases} \quad (2.1)$$

This has the properties $E\eta_A = 0$, $E\eta_A^2 = E\eta_A^3 = 1$ and $E\eta_A^4 = 2$. Two-point distributions are the only class for which (1.3) holds as an equality, an important property in their favour for this role. An alternative case is the so-called Rademacher distribution, taking the form

$$\eta_1 = \begin{cases} 1 & \text{with probability } p = \frac{1}{2} \\ -1 & \text{with probability } 1 - p. \end{cases} \quad (2.2)$$

and has the properties $E\eta_1 = 0$, $E\eta_1^2 = 1$, $E\eta_1^3 = 0$ and $E\eta_1^4 = 1$. This latter distribution offers the possibility of the higher-order improvements noted in the last section when the parent distribution is symmetric.

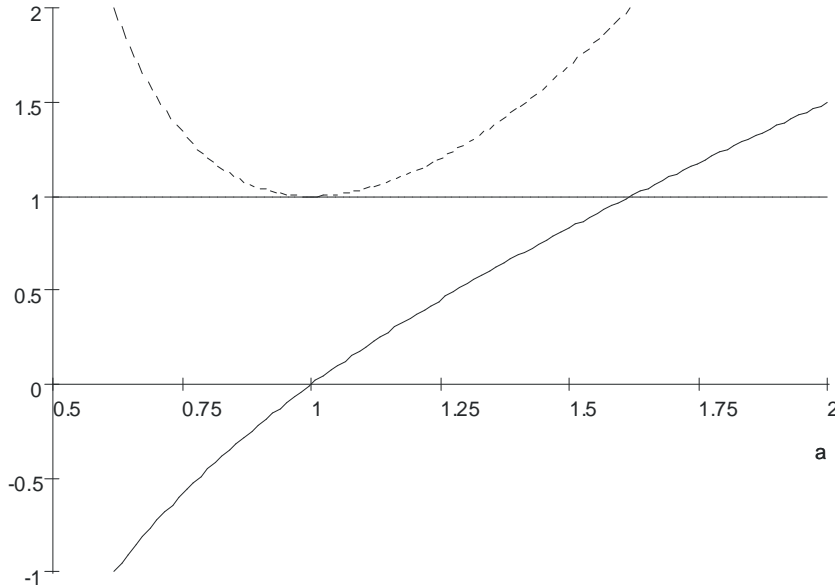


Figure 1: Skewness (solid line) and kurtosis (dashed line) of 2-point distributions η^a

Focussing on the choice between these two-point alternatives, there is evidently a conflict between achieving the best improvement of the ERP in the ‘worst case’ in which the parent distribution is skewed, hence favouring η_A , and taking advantage of possible symmetry of the parent distribution to achieve a better refinement by using η_1 . Instead, we propose extending the range of possibilities to achieve a potential balance of advantages.

A two-point distribution being completely specified by two point values and the associated probability, the conditions of zero mean and unit variance restrict the remaining free parameters to one. Define a class of two-point distributions η_a indexed on a parameter $a > 0$, such that $E\eta_a = 0$ and $E\eta_a^2 = 1$, by

$$\eta_a = \begin{cases} a & \text{with probability } \frac{1}{1+a^2} \\ -\frac{1}{a} & \text{with probability } \frac{a^2}{1+a^2}. \end{cases} \quad (2.3)$$

Setting $a = A = \frac{1}{2}(1 + \sqrt{5}) \approx 1.618$ yields (2.1) while setting $a = 1$ yields (2.2). Moreover, since these are two-point distributions, the relation

$$E\eta_a^4 = (E\eta_a^3)^2 + 1$$

holds for each a implying $E\eta_a^4 < E\eta_a^3 + 1$ in the range $1 < a < A$. Figure 1 shows the relationship between the two moments, plotted as functions of a . The difference between the curves attains its minimum of 0.8 at $a \approx 1.147$. Rather than choosing between the two extremes, these facts suggest choosing a to optimize the agreement between bootstrap and parent distributions. We denote the class of wild bootstrap distributions so defined by

$$\hat{u}_{at}^* = \hat{u}_t \eta_{at}$$

where $(\eta_{a1}, \dots, \eta_{an})$ are independent drawings from (2.3).

An easily implemented method for checking the choice of a empirically is to compute the Kolmogorov-Smirnov (KS) statistic for the residuals \hat{u}_t , relative to the quantiles of the bootstrap

distribution of \hat{u}_{at}^* . Let $\hat{u}_{(t)}$ denote the t th order statistic associated with the observed residuals, such that $\hat{u}_{(1)} \leq \dots \leq \hat{u}_{(n)}$. Given B independent replications drawn from some chosen wild bootstrap distribution, say $\{\hat{u}_1^{**}(j), \dots, \hat{u}_n^{**}(j)\}$ for $j = 1, \dots, B$, define the function

$$\hat{G}(t) = \frac{1}{Bn} \sum_{j=1}^B \sum_{s=1}^n 1(\hat{u}_s^{**}(j) \leq \hat{u}_{(t)})$$

where $1(\cdot)$ is the indicator function taking the value 1 when its argument is true, and 0 otherwise. $\hat{G}(t)$ can be viewed as an estimate of the quantity $F_{**}(\hat{u}_{(t)})$, where F_{**} is the CDF of the wild bootstrap distribution, and we expect to observe $\hat{G}(t) \approx t/n$ when $\{\hat{u}_1, \dots, \hat{u}_n\}$ is a drawing from this distribution. Accordingly, define

$$\text{KS} = n^{-1/2} \max_{1 \leq t \leq n} \left(|\hat{G}(t) - t/n| \right) \quad (2.4)$$

as our indicator of the agreement between the wild bootstrap and parent distributions. Note, in this application the statistic is used solely as a basis for ranking alternative choices of the parameter a . In the presence of skewness, the hypothesis of actual agreement between the distributions cannot truly hold.

3 Experimental Evidence

We performed Monte Carlo experiments to compare different choices of a in a regression model exhibiting both heteroscedasticity and skewness of the disturbances. The model is

$$y_t = \alpha + \beta x_t + u_t, \quad t = 1, \dots, n$$

where $\alpha = \beta = 0$, $x_t = \sin(\pi t/n)^2$ and

$$u_t = \sqrt{\exp(1 + \gamma t/n)} \varepsilon_t$$

where ε_t has a ‘skew-Student t ’ distribution. The latter is a mixture distribution of the form

$$\varepsilon = |z_\nu| \left(e^\xi x - e^{-\xi} (1 - x) \right)$$

where z_ν is a Student t variate with ν degrees of freedom ($\nu > 2$) and x is independently Bernoulli distributed with probability of success

$$P(x = 1) = \frac{1}{1 + e^{2\xi}}.$$

ξ is a parameter to capture the skewness, with $\xi = 0$ representing symmetry (see Fernandez and Steel, 1998). By choice of ξ and ν , we may arrange for ε_t to have any desired configuration of skewness and kurtosis. Note the array formulation of the model, ensuring a comparable pattern of heteroscedasticity at each sample size.

In our experiments we chose $\gamma = 5.5$ and $\xi = 2$, and $\nu = 25$. Table 1 shows the rejection relative frequencies in 100,000 replications, in the two-sided t -test of the true hypothesis $\beta = 0$. The t test was conducted using the robust statistic $|t_\beta|$ where

$$t_\beta = \sqrt{\frac{n-2}{n}} \frac{\sum_{t=1}^n (x_t - \bar{x}) \hat{u}_t}{\sqrt{\sum_{t=1}^n (x_t - \bar{x})^2 \hat{u}_t^2}}. \quad (3.1)$$

²The choice of trigonometric trend ensures the estimator of β is consistent, in this setup.

n	Asymptotic Criterion	Regular Bootstrap	Wild Bootstrap			
			$a = 1$	$a = 1.206$	$a = 1.412$	$a = 1.618$
15	0.114	0.076	0.068 [0.398]	0.073 [0.561]	0.087 [0.667]	0.111 [0.770]
30	0.084	0.072	0.053 [0.548]	0.056 [0.730]	0.063 [0.849]	0.072 [0.954]
60	0.069	0.066	0.056 [0.765]	0.058 [0.990]	0.061 [1.166]	0.062 [1.314]
120	0.060	0.059	0.055 [1.089]	0.053 [1.389]	0.056 [1.587]	0.058 [1.773]
240	0.055	0.053	0.051 [1.587]	0.052 [1.928]	0.052 [2.169]	0.052 [2.453]

Table 1: Rejection Frequencies of the Null Hypothesis in 100,000 Replications. (Mean KS statistics in square brackets.)

In the table, ‘asymptotic criterion’ means that the rejection criterion took the form $|t_\beta| > 1.96$, and note that this test is asymptotically correctly sized. The rejection criterion for the bootstrap tests is $\tilde{P}_*(|t_\beta^*| \leq |t_\beta|) < 0.05$, where t_β^* is defined by (3.1) with \hat{u}_t replaced by \hat{u}_s^* or \hat{u}_{as}^* , and \tilde{P}_* denotes the probability under the bootstrap EDP, estimated by Monte Carlo with 99 replications. The figures in square brackets in the last four columns are the mean values of the KS statistic from (2.4), which is computed in each Monte Carlo replication. Repeating the experiments with $\nu = 100$, the case of virtually normal kurtosis, revealed little or no difference in the results.

4 Discussion

The example chosen is deliberately a “worst case”, with heavy skewness as well as pronounced heteroscedasticity. The skew-corrected wild bootstrap might be expected to perform best relative to its rivals, and it is therefore notable that the wild bootstrap ERP nonetheless increases monotonically with a , the case $a = A$ barely improving on the asymptotic criterion. The wild bootstrap with the Rademacher distribution ($a = 1$) is unambiguously the winner amongst the six alternatives, at all sample sizes. Moreover, this result is reflected accurately in the reported KS statistics. We cite this as strong evidence that the KS statistic is a reliable guide to the relative ERP of the corresponding wild bootstrap test.

We note in conclusion that Liu (1988) and Mammen (1993) suggest an alternative auxiliary distribution based on normal variates, $\eta = W_1W_2 - EW_1EW_2$ where $W_i \sim N(\delta_i, \frac{1}{2})$ and

$$\begin{aligned}\delta_1 &= \frac{1}{2} \left(\sqrt{17/6} + \sqrt{1/6} \right) \\ \delta_2 &= \frac{1}{2} \left(\sqrt{17/6} - \sqrt{1/6} \right).\end{aligned}$$

It can be verified that this distribution has $E\eta = 0$, $E\eta^2 = E\eta^3 = 1$. However, it is also easily verified that $E\eta^4 = 5.625$. As we have noted previously, two-point distributions offer the most favourable trade-off between third and fourth moments. It appears unlikely that alternative distributions of this type could provide a better remedy for our problem.

References

- Beran, R. (1986) Discussion: Jackknife, bootstrap and other resampling methods in regression analysis, *Annals of Statistics* 14, 1295-1298.
- Davidson, J. (2006) *Time Series Modelling 4.18*, at <http://www.timeseriesmodelling.com>
- Davidson, R., Flachaire, E. (2001) The wild bootstrap, tamed at last. Queen's Institute for Economic Research Working Paper No. 1000.
- Doornik, J. A. (2004) *Ox: an Object-Oriented Matrix Programming Language*. Timberlake Consultants Ltd.
- Fernandez, C., and M. Steel (1998) On Bayesian modelling of fat tails and skewness, *Journal of the American Statistical Association* 93, 359-371.
- Liu, R. Y. (1988) Bootstrap procedure under some non-I.I.D. models. *Annals of Statistics* 16, 1696-1708.
- Mammen, E. (1993) Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics* 21, 255-285.
- Paya, I., Peel, D. A. (2009) On the speed of adjustment in ESTAR models when allowance is made for bias in estimation. *Economics Letters* 90, 272-277.
- Pearson, K. (1916) Mathematical contributions to the theory of evolution XIX; second supplement to a memoir on skew variation, *Philos. Trans. Roy. Soc. London*, Ser A 216, 432.
- Wu, C. F. J.,(1986). Jackknife bootstrap and other resampling methods in regression analysis. *Annals of Statistics* 14, 1261-1295.