Heteroskedasticity-and-Autocorrelation-Consistent Bootstrapping

by

Russell Davidson

Department of Economics and CIREQ
McGill University
Montréal, Québec, Canada
H3A 2T7

AMSE-GREQAM
Centre de la Vieille Charité
2 Rue de la Charité
13236 Marseille cedex 02, France

russell.davidson@mcgill.ca

and

Andrea Monticini

Catholic University
Via Necchi 5
20123 Milan, Italy

andrea.monticini@gmail.com

Abstract

In many, if not most, econometric applications, it is impossible to estimate consistently the elements of the white-noise process or processes that underlie the DGP. A common example is a regression model with heteroskedastic and/or autocorrelated disturbances, where the heteroskedasticity and autocorrelation are of unknown form.

A particular version of the wild bootstrap can be shown to work very well with many models, both univariate and multivariate, in the presence of heteroskedasticity. Nothing comparable appears to exist for handling serial correlation. Recently, there has been proposed something called the dependent wild bootstrap. Here, we extend this new method, and link it to the well-known HAC covariance estimator, in much the same way as one can link the wild bootstrap to the HCCME. It works very well even with sample sizes smaller than 50, and merits considerable further study.

Keywords: Bootstrap, time series, wild bootstrap, dependent wild bootstrap, HAC covariance matrix estimator

JEL codes: C12, C22, C32

This research was supported by the Canada Research Chair program (Chair in Economics, McGill University) and by grants from the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche du Québec - Société et culture.

February, 2014

1. Introduction

In this paper, we look at methods that have been proposed for bootstrap inference with time-series models, in which the main concerns are potential heteroskedasticity and autocorrelation. For a bootstrap based on conventional resampling to work well, it is necessary to be able to resample from a set of objects that are, at least approximately, IID realisations from some unknown univariate or multivariate distribution. A parametric bootstrap, if available, commonly delivers very reliable inference, but, in the presence of either heteroskedasticity or autocorrelation of unknown form, appropriate parametric models are not available.

We will see that the so-called wild bootstrap succeeds in overcoming all problems due to heteroskedasticity alone, provided only that it is used with care. Errors in inference with the wild bootstrap are very comparable in magnitude to those incurred by parametric or resampling bootstraps applied in suitable circumstances.

Matters are different with autocorrelation. So far, no bootstrap has been proposed that, in the presence of autocorrelation of unknown form, can deliver performance comparable to what can be obtained in its absence. Perhaps in consequence, a considerable number of bootstrap methods have been proposed, some a good deal better than others. By far the most popular of these are the various versions of the block bootstrap, although it has been seen that the block bootstrap often works poorly. In some circumstances, other schemes can sometimes work better.

The properties of the dependent wild bootstrap suggest that, in some cases, inference can benefit greatly from the Fast Double Bootstrap (FDB) of Davidson and MacKinnon (2007). In addition, in Section 9, a method is proposed for diagnosing bootstrap success or failure, and the extent to which inference can be improved by the FDB. A few concluding comments are made in Section 10.

2. The Wild Bootstrap

The so-called wild bootstrap was introduced as an alternative to the pairs bootstrap originally proposed by Freedman (1981). Early references to the wild bootstrap include Wu (1986), Liu (1988), and Mammen (1993).

Suppose, for simplicity, that we wish to undertake bootstrap inference for a hypothesis about the parameter vector β of the linear regression model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}, \tag{1}$$

If the assumption of no serial correlation is maintained but not that of homoskedasticity, then we have $\mathrm{E}(u_t^2) = \sigma_t^2$, $t = 1, \ldots, n$. The σ_t^2 must be considered to be unknown parameters. With possible heteroskedasticity, as with the pairs bootstrap, test statistics themselves must be modified so as to use an Eicker-White HCCME (heteroskedasticity-consistent covariance matrix estimator); see Eicker (1963) and White (1980). The form of this HCCME is

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

where $\hat{\boldsymbol{\Omega}}$ can take various forms, according to which version of the HCCME is desired. The easiest, but not necessarily the most desirable, is

$$\hat{\boldsymbol{\Omega}} = \operatorname{diag}\{\tilde{u}_t^2\}, \tag{3}$$

a diagonal matrix with diagonal elements the squared residuals obtained by estimating the model under the null hypothesis. Clearly, $\hat{\boldsymbol{\Omega}}$ cannot be consistent, since it is impossible to estimate the n variances with only n observations. But it is well-known that $n^{-1}\boldsymbol{X}^\top\hat{\boldsymbol{\Omega}}\boldsymbol{X}$ has a probability limit as $n \to \infty$ equal to that of $n^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X}$, where $\boldsymbol{\Omega} = \operatorname{diag}\{\sigma_t^2\}$ is the true disturbance covariance matrix.

When there is heteroskedasticity, meaning that the σ_t^2 are not all equal, resampling residuals of any sort ignores this fact. The wild bootstrap takes account of possible heteroskedasticity by using as bootstrap disturbances the residuals from estimation of the model under the null hypothesis, each multiplied by one of a set of mutually independent random variables of expectation zero and variance one. Thus the bootstrap disturbance terms can be written as

$$u_t^* = \tilde{u}_t\varepsilon_t^*, \tag{4}$$

where the ε_t^* are independent random variables, denoted with a star to indicate that they are generated by the investigator's random number generator.

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

$$\varepsilon_t^* = \begin{cases} -(\sqrt{5}-1)/2 & \text{with probability } p = (\sqrt{5}+1)/(2\sqrt{5}) \\ (\sqrt{5}+1)/2 & \text{with probability } 1-p, \end{cases}$$

which satisfies the condition on the third moment. Liu also mentions the possibility of Rademacher variables, defined as

$$\varepsilon_t^* = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2. \end{cases} \tag{5}$$

This amounts to giving each residual a random sign in the bootstrap DGP.

The wild bootstrap with GARCH disturbances

The wild bootstrap is probably most commonly used when it is suspected that there is unconditional heteroskedasticity. But it is equally useful with conditional heteroskedasticity. When it is used along with the Rademacher distribution, the covariance structure of the squared bootstrap disturbances is the same as that of the squared residuals from the original sample. This is because the squared bootstrap disturbances are always just the squared residuals, so that any relationship among the squared residuals, like that given by any GARCH model, is preserved unchanged by the Rademacher wild bootstrap.

In order to see whether this theoretical conclusion is borne out in simulation experiments, it is necessary to be able to generate GARCH disturbances for the DGP that generates the original data in a simulation experiment. If we limit attention to GARCH(1,1), then it is described by the two equations

$$\begin{aligned}\sigma_t^2 &= \alpha + \gamma u_{t-1}^2 + \delta \sigma_{t-1}^2, \\ u_t &= \sigma_t \varepsilon_t,\end{aligned} \tag{6}$$

where the series ε_t is white noise. For simulation, it is convenient to rewrite (6) as

$$\sigma_t^2 = \alpha + (\delta + \gamma \varepsilon_{t-1}^2)\sigma_{t-1}^2. \tag{7}$$

This recurrence relation has to be initialised before it can generate the variances of the GARCH process. If the process is stationary, then the expectation of σ_t^2 does not depend on t, and so it is

$$\mathrm{E}(\sigma_t^2) = \frac{\alpha}{1-\gamma-\delta}.$$

The right-hand side of this equation is a suitable value for what we may denote by σ_0^2.

It is easy to check that the recurrence (7) gives the explicit solution

$$\sigma_t^2 = \alpha\Big(1 + \sum_{s=1}^{t-1}\prod_{i=1}^{s}(\delta + \gamma\varepsilon_{t-i}^2)\Big) + \sigma_0^2 \prod_{j=1}^{t-1}(\delta + \gamma \varepsilon_j^2).$$

If we make the definition

$$V_t = \prod_{j=1}^{t-1}(\delta + \gamma\varepsilon_j^2),$$

then

$$\sum_{s=1}^{t-1}\prod_{i=1}^{s}(\delta + \gamma\varepsilon_{t-i}^2) = V_t \sum_{r=1}^{t-1}(1/V_r),$$

and

$$\sigma_t^2 = \alpha\big(1 + V_t\sum_{r=1}^{t-1}(1/V_r)\big) + \sigma_0^2 V_t. \tag{8}$$

The final step in generating disturbances which follow the GARCH(1,1) process is to generate

$$u_t = \sigma_t \varepsilon_t. \tag{9}$$

The above procedure is easy to implement, and is fast.

The model used in the simulation experiment is:

$$y_t = a + \rho y_{t-1} + u_t, \tag{10}$$

where u_t is generated using (8) and (9). For the moment, we limit our attention to the DGPs with $a = 1.5$, $y_0 = 0$, with sample sizes $n = 10, 30, 50$, and $\rho = 0.3, 0.5, 0.7$ and 0.9. For the GARCH disturbances, the parameter values are $\alpha = 1$, $\gamma = 0.4$, $\delta = 0.45$. In order to test the hypothesis that $\rho = \rho_0$, the test statistic used is

$$\tau = \frac{\hat\rho - \rho_0}{\hat\sigma_\rho},$$

where $\hat\rho$ is the OLS estimate from (10), run over observations 2 to n. The standard error $\hat\sigma_\rho$ is obtained by use of the HC_2 variant of the Eicker-White HCCME.

The bootstrap DGP is determined by first running the constrained regression

$$y_t - \rho_0 y_{t-1} = a + u_t, \quad t = 2, \ldots, n,$$

in order to obtain the estimate $\tilde a$, and the constrained residuals $\tilde u_t$, $t = 2, \ldots n$. A bootstrap sample is defined by

$$y_1^* = y_1 \quad \text{and} \quad y_t^* = \tilde a + \rho_0 y_{t-1}^* + \varepsilon_t \tilde u_t, \quad t = 2, \ldots, n,$$

where the ε_t are IID realisations from the Rademacher distribution. The bootstrap statistics are

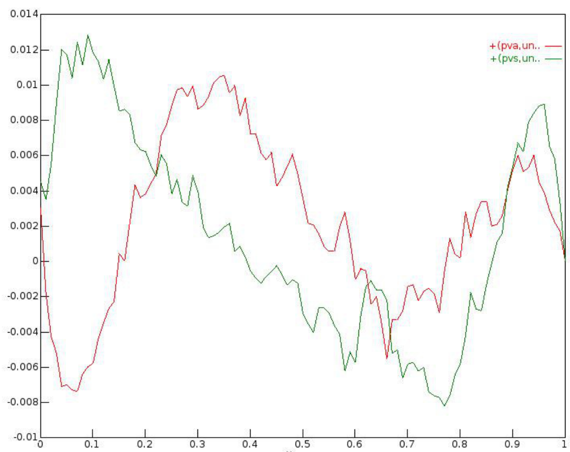
$$\tau_j^* = \frac{\hat\rho^* - \rho_0}{\hat\sigma_\rho^*}, \quad j = 1, \ldots, B$$

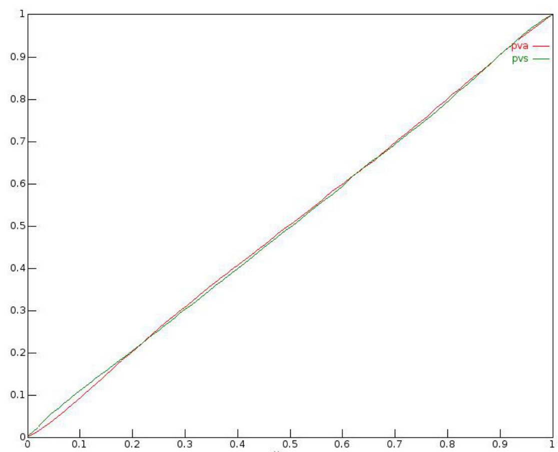
with $\hat\rho^*$ and $\hat\sigma_\rho^*$ defined as the bootstrap counterparts of $\hat\rho$ and $\hat\sigma_\rho$ respectively.

The bootstrap P value is the proportion of the τ_j^* that are more extreme than τ. We used a two-tailed test. Results for $N = 9999$ replications with $B = 199$ bootstrap samples each are as follows. The bootstrap discrepancy is the difference between the nominal level of 5% and the rejection rate in the simulations, that is, the proportion of the replications yielding a bootstrap P value less than 0.05.

n	ρ	bootstrap discrepancy
10	0.9	−0.006
30	0.9	+0.003
50	0.9	+0.002
10	0.7	−0.006
30	0.7	+0.002
50	0.7	+0.000
10	0.5	−0.008
30	0.5	+0.002
50	0.5	+0.001
10	0.3	−0.005
30	0.3	+0.003
50	0.3	+0.001

The acid test of a bootstrap procedure is to look at the P value discrepancy plot, which plots the (simulated) error in rejection probability (ERP) as a function of the nominal level α, for all values of α from 0 to 1. If the bootstrap works perfectly, the plot should be indistinguishable from the horizontal axis. Alternatively, the P value plot plots the rejection probability (RP) itself as a function of α, and should be close to the 45° line. The figures below, given for illustrative purposes, are for the case with $n = 10$ and $\rho = 0.3$. The curves in red are for a two-tailed test; those in green for a one-tailed test that rejects to the right. It can be seen that the small discrepancy for $\alpha = 0.05$ is not a coincidence, and that use of a two-tailed test confers no significant advantage.





We seem to be able to conclude that the wild bootstrap with the Rademacher distribution provides a very reliable procedure in the presence of heteroskedasticity, whether conditional or unconditional.

3. The Dependent Wild Bootstrap Based on the HAC Estimator

The procedure called the Dependent Wild Bootstrap (DWB) was introduced by Shao (2010) as an alternative to other methods, like the various versions of the block bootstrap, suggested to take account of autocorrelation. It is seen to provide rather good estimates of the distributions of quantities computed from autocorrelated time series if these satisfy the assumptions of the "smooth-function model" of Hall (1992) and Lahiri (2003); of course these quantities include the sample mean and variance.

Given a time series $\{x_t\}$, $t = 1, \ldots, n$, with mean $\bar{x}$, a DWB sample is defined by

$$x_t^* = \bar{x} + (x_t - \bar{x})\eta_t^*, \quad t = 1, \ldots, n,$$

where the η_t^* are independent of the sample, have expectation 0 and variance 1, but, unlike the ε_t^* of the plain wild bootstrap, are autocorrelated, with covariances $\text{cov}(\eta_t, \eta_s) = K(t - s)$, for some symmetric positive semi-definite kernel function K.

Appropriate choices for K, up to a choice of bandwidth, include the Bartlett, quadratic spectral, and Parzen kernels. Rather than discuss the DWB in detail following Shao's exposition, we move on directly to an extension of it more closely related to econometrics.

Here, we follow the same line of reasoning as that which leads from the HCCME to the wild bootstrap, but starting from a HAC (heteroskedasticity-and-autocorrelation consistent) covariance matrix estimator. For the linear regression (1), the covariance matrix of the OLS estimates of the parameter vector β is

$$(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}, \tag{11}$$

where $\boldsymbol{\Omega}$ is the covariance matrix of the disturbance vector $\boldsymbol{u}$. Estimators of this covariance matrix that are robust to heteroskedasticity or to both heteroskedasticity and autocorrelation all take the form (2), for various different choices of the inconsistent estimator $\hat{\boldsymbol{\Omega}}$. Any version of the HCCME has non-zero elements only on the principal diagonal, and these are functions of the squared residuals. With all forms of HAC estimator, the off-diagonal elements are estimated using cross-products of residuals. For instance, with the well-known Newey-West HAC covariance matrix, proposed in Newey and West (1987), the diagonal elements are estimated, as in the HCCME, by the squared residuals. But the (t, s) element, $t \neq s$ is

$$\hat{\omega}_{ts} = \Big(1 - \frac{|t-s|}{p+1}\Big)\tilde{u}_t\tilde{u}_s \quad \text{for } |t-s| < p+1,$$

and zero for $|t-s| \geq p+1$, where p is the lag truncation parameter.

More generally, if $k(\cdot)$ is a positive definite kernel, we have

$$\hat{\omega}_{ts} = k(|t-s|)\tilde{u}_t\tilde{u}_s.$$

The Newey-West estimator is based on the Bartlett kernel. The QS estimator of Andrews and Monahan (1992) uses the quadratic spectral kernel.

The variances of the bootstrap disturbances (4), conditional on the original data, are the $\tilde{u}_t^2$, the diagonal elements of the HCCME $\hat{\boldsymbol{\Omega}}$. With autocorrelation, it is natural to look for bootstrap disturbances of which the conditional covariance matrix is a HAC $\hat{\boldsymbol{\Omega}}$. If we denote by $\boldsymbol{U}$ the $n \times n$ diagonal matrix with typical diagonal element $\tilde{u}_t$, a HAC $\hat{\boldsymbol{\Omega}}$ can be written as $\boldsymbol{UKU}$, where $\boldsymbol{K}$ is a symmetric matrix wih typical element

$$k_{ts} = k(|t-s|),$$

where $k(\cdot)$ is the chosen kernel function. By a Cholesky decomposition, define the lower-triangular matrix $\boldsymbol{L}$ to be such that $\boldsymbol{L}\boldsymbol{L}^\top = \boldsymbol{K}$. If, as with the wild bootstrap, the vector $\boldsymbol{\varepsilon}^*$ has IID elements of expectation zero and unit variance, then we may form bootstrap disturbances as

$$\boldsymbol{u}^* \equiv \boldsymbol{UL}\boldsymbol{\varepsilon}^*,$$

since, conditional on $\boldsymbol{U}$, $\mathrm{E}(\boldsymbol{u}^*) = 0$, and $\mathrm{E}(\boldsymbol{u}^*(\boldsymbol{u}^*)^\top) = \boldsymbol{U}\boldsymbol{L}\boldsymbol{L}^\top\boldsymbol{U} = \boldsymbol{U}\boldsymbol{K}\boldsymbol{U}$, as desired.

We consider first the statistic

$$\tau \equiv \boldsymbol{y}^\top\boldsymbol{X}(\boldsymbol{X}^\top\hat{\boldsymbol{\Omega}}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}. \tag{12}$$

with a HAC $\hat{\boldsymbol{\Omega}}$, as appropriate to test the hypothesis that $\boldsymbol{\beta} = \boldsymbol{0}$ in the model (1). The bootstrap version of the statistic is then

$$\tau^* = (\boldsymbol{u}^*)^\top\boldsymbol{X}(\boldsymbol{X}^\top\hat{\boldsymbol{\Omega}}^*\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u}^*, \tag{13}$$

where $\hat{\boldsymbol{\Omega}}^* = \boldsymbol{U}^*\boldsymbol{K}\boldsymbol{U}^*$ and $\boldsymbol{U}^*$ is the diagonal matrix with typical element u_t^*. Let the vector $\boldsymbol{\eta}^* = \boldsymbol{L}\boldsymbol{\varepsilon}^*$, and let $\boldsymbol{H}^*$ be the diagonal matrix with elements those of $\boldsymbol{\eta}^*$. Then, since $\boldsymbol{u}^* = \boldsymbol{U}\boldsymbol{\eta}^*$, we also have $\boldsymbol{U}^* = \boldsymbol{U}\boldsymbol{H}^* = \boldsymbol{H}^*\boldsymbol{U}$, because the two diagonal matrices commute. Similarly, $\boldsymbol{X}^\top\boldsymbol{u}^* = \boldsymbol{X}^\top\boldsymbol{U}\boldsymbol{\eta}^* = \boldsymbol{X}^\top\boldsymbol{H}^*\boldsymbol{u}$. These relations show that

$$\hat{\boldsymbol{\Omega}}^* = \boldsymbol{H}^*\boldsymbol{U}\boldsymbol{K}\boldsymbol{U}\boldsymbol{H}^* = \boldsymbol{H}^*\hat{\boldsymbol{\Omega}}\boldsymbol{H}^* \quad \text{and} \quad \boldsymbol{X}^\top\hat{\boldsymbol{\Omega}}^*\boldsymbol{X} = \boldsymbol{X}^\top\boldsymbol{H}^*\hat{\boldsymbol{\Omega}}\boldsymbol{H}^*\boldsymbol{X},$$

and so

$$\tau^* = \boldsymbol{u}^\top\boldsymbol{H}^*\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{H}^*\hat{\boldsymbol{\Omega}}\boldsymbol{H}^*\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{H}^*\boldsymbol{u}.$$

There are a couple of other useful ways to express τ^*. Define the matrix $\boldsymbol{Z}$ to have row t equal to $\boldsymbol{X}_t u_t$. Then $\boldsymbol{X}^\top\boldsymbol{u}^* = \boldsymbol{X}^\top\boldsymbol{U}\boldsymbol{\eta}^* = \boldsymbol{Z}^\top\boldsymbol{\eta}^*$, and

$$\boldsymbol{X}^\top\hat{\boldsymbol{\Omega}}^*\boldsymbol{X} = \boldsymbol{X}^\top\boldsymbol{H}^*\boldsymbol{U}\boldsymbol{K}\boldsymbol{U}\boldsymbol{H}^*\boldsymbol{X} = \boldsymbol{X}^\top\boldsymbol{U}\boldsymbol{H}^*\boldsymbol{K}\boldsymbol{H}^*\boldsymbol{U}\boldsymbol{X} = \boldsymbol{Z}^\top\boldsymbol{H}^*\boldsymbol{K}\boldsymbol{H}^*\boldsymbol{Z}, \tag{14}$$

which gives

$$\tau^* = \boldsymbol{\eta}^{*\top}\boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{H}^*\boldsymbol{K}\boldsymbol{H}^*\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{\eta}^* = \boldsymbol{\iota}^\top\boldsymbol{H}^*\boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{H}^*\boldsymbol{K}\boldsymbol{H}^*\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{H}^*\boldsymbol{\iota}, \tag{15}$$

where $\boldsymbol{\iota}$ is a vector each element of which is one.

The bootstrap procedure outlined above is an extension of the wild bootstrap considered in Section 3. To see this, note that the wild bootstrap version of the statistic (12) appropriate for testing $\boldsymbol{\beta} = \boldsymbol{0}$ is

$$\boldsymbol{u}^{*\top}\boldsymbol{X}(\boldsymbol{X}^\top\hat{\boldsymbol{\Omega}}^*\boldsymbol{X})^\top\boldsymbol{X}^\top\boldsymbol{u}^*, \tag{16}$$

where $\hat{\boldsymbol{\Omega}}^* = (\boldsymbol{U}^*)^2$, the diagonal matrix with elements the $\tilde{u}_t^{*2}$. Since for the wild bootstrap $\boldsymbol{u}^* = \boldsymbol{U}\boldsymbol{\varepsilon}^*$, we can write $\boldsymbol{U}^* = \boldsymbol{U}\boldsymbol{E}^*$, where $\boldsymbol{E}^*$ is a diagonal matrix with elements those of $\boldsymbol{\varepsilon}^*$. Because $\boldsymbol{U}\boldsymbol{X} = \boldsymbol{Z}$, we see that the wild bootstrap statistic is

$$\boldsymbol{\varepsilon}^{*\top}\boldsymbol{Z}(\boldsymbol{Z}^\top(\boldsymbol{E}^*)^2\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{\varepsilon}^*. \tag{17}$$

Now replace the kernel $\boldsymbol{K}$ of the bootstrap statistic of this section by an identity matrix, which means that $\boldsymbol{L}$ is also an identity matrix. Then also $\boldsymbol{\eta}^* = \boldsymbol{\varepsilon}^*$, and $\boldsymbol{H}^* = \boldsymbol{E}^*$. From the middle expression in (15), the bootstrap statistic is

$$\boldsymbol{\varepsilon}^{*\top}\boldsymbol{Z}(\boldsymbol{Z}^\top(\boldsymbol{E}^*)^2\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{\varepsilon}^*,$$

which is identical to (17).

The statistic τ of (12) is one possible realisation of the wild bootstrap statistic (16), obtained by setting $\boldsymbol{\varepsilon}^*$ equal to $\boldsymbol{\iota}$. With a kernel $\boldsymbol{K}$ different from the identity matrix, this property no longer holds. A way that may improve this situation is to modify the statistic τ itself, so that it takes the same form as τ^*. If we define $\boldsymbol{\eta} = \boldsymbol{L\iota}$, and let $\boldsymbol{H}$ be the diagonal matrix with elements those of $\boldsymbol{\eta}$, we may construct a statistic with the same form as the rightmost expression in (15), so as to get

$$\boldsymbol{\iota}^\top \boldsymbol{HZ}(\boldsymbol{Z}^\top \boldsymbol{HKHZ})^{-1}\boldsymbol{Z}^\top \boldsymbol{H\iota}.$$

Now

$$\boldsymbol{Z}^\top \boldsymbol{H\iota} = \boldsymbol{X}^\top \boldsymbol{UH\iota} = \boldsymbol{X}^\top \boldsymbol{HU\iota} = \boldsymbol{X}^\top \boldsymbol{Hu},$$

and

$$\boldsymbol{Z}^\top \boldsymbol{HKHZ} = \boldsymbol{X}^\top \boldsymbol{HUKUHX} = \boldsymbol{X}^\top \boldsymbol{H\hat{\Omega}HX}.$$

Thus the new statistic can be written as

$$\boldsymbol{u}^\top \boldsymbol{HX}(\boldsymbol{X}^\top \boldsymbol{H\hat{\Omega}HX})^{-1}\boldsymbol{X}^\top \boldsymbol{Hu}. \tag{18}$$

This differs from the original statistic (12) by replacing $\boldsymbol{X}$ by $\boldsymbol{HX}$.

The procedure that is suggested by the above is not a standard bootstrap. It is purely a question of potential power loss whether the new statistic (18) can reasonably replace the original statistic (12). But the bootstrap statistic (15) is *not* the bootstrap version of (18), at least in the conventional sense, according to which τ^* would be obtained from the τ of (18) by replacing $\boldsymbol{u}$ by $\boldsymbol{u}^*$, $\boldsymbol{H}$ by $\boldsymbol{H}^*$, and $\boldsymbol{\hat{\Omega}}$ by $\boldsymbol{\hat{\Omega}}^*$. But it is the case that the statistic (18) is one realisation of (15), with $\boldsymbol{\varepsilon}^* = \boldsymbol{\iota}$.

Asymptotic validity

There is nothing particular involved in showing that either of the bootstrap procedures introduced here is asymptotically valid. Since, under regularity conditions that it is unnecessary to make explicit here, the statistics (12) and (18) both tend in distribution to χ^2_k as $n \to \infty$, it is enough to show that the bootstrap statistics also do so in probability.

The main property necessary for a HAC covariance estimator to be valid asymptotically is that, for the asymptotic construction considered,

$$\operatorname*{plim}_{n\to\infty} n^{-1}\boldsymbol{X}^\top \boldsymbol{\hat{\Omega}X} = \lim_{n\to\infty} n^{-1}\boldsymbol{X}^\top \boldsymbol{\Omega X}. \tag{19}$$

We assume therefore that (19) holds. Next, we need to be able to use a central-limit theorem to show that

$$n^{-1/2}\boldsymbol{X}^\top \boldsymbol{u} \to_d \mathrm{N}\big(\boldsymbol{0}, \lim_{n\to\infty} n^{-1}\boldsymbol{X}^\top \boldsymbol{\Omega X}\big). \tag{20}$$

Assumptions (19) and (20) together imply that the τ of (12) tends in distribution to χ^2_k. A conventional asymptotic construction supposes that the matrix $n^{-1}\boldsymbol{X}^\top \boldsymbol{X}$ converges, either deterministically or in probability, to a nonrandom finite positive definite matrix. In order to obtain the same limiting distribution for the statistic (18), we must assume that the kernel $\boldsymbol{K}$ is chosen so that $\boldsymbol{HX}$ is also such that $n^{-1}\boldsymbol{X}^\top(\boldsymbol{H})^2\boldsymbol{X}$ tends to a nonrandom finite positive definite matrix. This requirement has consequences for the choice of the lag truncation parameter of the Bartlett kernel, or for the bandwidth of the quadratic spectral kernel, but we do not look too closely at this matter.

For the bootstrap statistic, we start by working conditionally on the original data, which for present purposes means conditionally on $\boldsymbol{u}$ and $\boldsymbol{U}$. Note that $\boldsymbol{X}^\top\boldsymbol{u}^* = \boldsymbol{X}^\top\boldsymbol{U}\boldsymbol{\eta}^* = \boldsymbol{Z}^\top\boldsymbol{\eta}^*$. Provided that $\boldsymbol{Z} = \boldsymbol{UX}$ satisfies the usual condition that $n^{-1}\boldsymbol{Z}^\top\boldsymbol{Z}$ tends to a finite positive definite matrix that is random only through $\boldsymbol{U}$ (if at all), then

$$n^{-1/2}\boldsymbol{Z}^\top\boldsymbol{\eta}^* \to_{d^*} \mathrm{N}\big(\boldsymbol{0}, \lim_{n\to\infty} n^{-1}\boldsymbol{Z}^\top\boldsymbol{KZ}\big),$$

where the notation $\to_{d^*}$ means convergence in distribution conditional on the original data (as one might say, in the bootstrap world); since $\mathrm{E}(\boldsymbol{\eta}^*\boldsymbol{\eta}^{*\top}) = \boldsymbol{LL}^\top = \boldsymbol{K}$. But $\boldsymbol{Z}^\top\boldsymbol{KZ} = \boldsymbol{X}^\top\boldsymbol{UKUX} = \boldsymbol{X}^\top\hat{\boldsymbol{\Omega}}\boldsymbol{X}$, and so, under the assumption (19), the limiting covariance is $\lim_{n\to\infty} n^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X}$, independently of $\boldsymbol{U}$. By (20), this is also the limiting covariance of $n^{-1/2}\boldsymbol{X}^\top\boldsymbol{u}$.

It remains to consider the limit of $n^{-1}\boldsymbol{X}^\top\hat{\boldsymbol{\Omega}}^*\boldsymbol{X}$, the covariance matrix in the bootstrap statistic (13). From (14), we have

$$\boldsymbol{X}^\top\hat{\boldsymbol{\Omega}}^*\boldsymbol{X} = \boldsymbol{Z}^\top\boldsymbol{H}^*\boldsymbol{K}\boldsymbol{H}^*\boldsymbol{Z} = \sum_{t=1}^{n}\sum_{s=1}^{n}\boldsymbol{Z}_t\eta_t^*K_{ts}\eta_s^*\boldsymbol{Z}_s.$$

(K_{ts} is element (t,s) of $\boldsymbol{K}$, η_t^* is element t of $\boldsymbol{\eta}^*$.) The conditional expectation of the term in the double sum above is $\boldsymbol{Z}_tK_{ts}^2\boldsymbol{Z}_s$. Let $\boldsymbol{K}_2$ be the matrix of which element (t,s) is K_{ts}^2. Then we see that

$$\mathrm{E}^*(\boldsymbol{X}^\top\hat{\boldsymbol{\Omega}}^*\boldsymbol{X}) = \boldsymbol{Z}^\top\boldsymbol{K}_2\boldsymbol{Z} = \boldsymbol{X}^\top\boldsymbol{U}\boldsymbol{K}_2\boldsymbol{UX},$$

where E^* stands for the conditional expectation. Except for the presence of $\boldsymbol{K}_2$ instead of $\boldsymbol{K}$, this is the same as $\boldsymbol{X}^\top\boldsymbol{UKUX} = \boldsymbol{X}^\top\hat{\boldsymbol{\Omega}}\boldsymbol{X}$. The question is whether, as $n \to \infty$, the unconditional limit of $n^{-1}\boldsymbol{X}^\top\boldsymbol{U}\boldsymbol{K}_2\boldsymbol{UX}$ is the same as the limit of $n^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X}$. It is if $\boldsymbol{K}_2$ is a suitable kernel for a HAC estimator.

A suitable kernel must have all of its diagonal elements equal to 1, and be positive definite. Since all the diagonal elements of $\boldsymbol{K}$ are one, the former condition is satisfied trivially. For the latter, let the original kernel $\boldsymbol{K}$ have the spectral decomposition

$$\boldsymbol{K} = \sum_{i=1}^{n}\lambda_i\boldsymbol{v}_i\boldsymbol{v}_i^\top,$$

where the eigenvalues λ_i are all positive because $\boldsymbol{K}$ is positive definite. The (t,s) element of $\boldsymbol{K}$ is thus

$$K_{ts} = \sum_{i=1}^{n}\lambda_i v_{ti}v_{si},$$

so that

$$\boldsymbol{K}_2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j(\boldsymbol{v}_i \odot \boldsymbol{v}_j)(\boldsymbol{v}_i \odot \boldsymbol{v}_j)^\top, \tag{21}$$

where $\odot$ denotes the element-wise product. But (21) expresses $\boldsymbol{K}_2$ as the sum of positive semi-definite rank-one matrices, which demonstrates that $\boldsymbol{K}_2$ is itself positive definite. Thus the asymptotic validity of our bootstrap procedures is confirmed.

4. An Interesting Failure: the Maximum-Entropy Bootstrap

The principle of maximum entropy was propounded by Jaynes (1957) as an interpretation of statistical mechanics that treated the problems of thermodynamics as problems of statistical inference on the basis of extremely limited information. One application of the principle was proposed by Theil and Laitinen (1980), for the estimation from a random IID sample of the density of the underlying distribution, under the assumption that the distribution is continuous and is almost everywhere differentiable. For a brief discussion of the method, see the more accessible Theil and Fiebig Denzil (1982). For a sample of size n, with order statistics $x_{(i)}$, $i = 1, \ldots, n$, the estimated distribution has, except in the tails, a continuous piecewise linear CDF that assigns probability mass $1/n$ to each interval $I_i \equiv [(x_{(i-1)} + x_{(i)}/2, (x_{(i)} + x_{(i+1)})/2]$, for $i = 2, \ldots, n-1$. The distribution is exponential in the tails, defined as the intervals I_1 from $-\infty$ to $(x_{(1)} + x_{(2)})/2$, and I_n from $(x_{(n-1)} + x_{n)})/2$ to $+\infty$. Each of the infinite intervals receives a probability mass of $1/n$, and the lower interval is constructed to have an expectation of $0.75x_{(1)} + 0.25x_{(2)}$, the upper an expectation of $0.25x_{(n-1)} + 0.75x_{(n)}$.

This way of estimating a distribution was picked by Vinod (2006), who bases a technique for bootstrapping time series on it. He modifies the procedure described above so as to allow for the possibility of a bounded rather than an infinite support, but I cannot follow the details (I think they are wrong). Aside from this, his method proceeds as follows:

1. Define an $n \times 2$ sorting matrix S_1 and place the index set $T_0 = \{1, 2, \ldots, n\}$ in the first column and the observed time series x_t in the second column.

2. Sort the matrix S_1 with respect to the numbers in its second column while carrying along the numbers in the first column. This yields the order statistics $x_{(i)}$ in the second column and a vector $I_{\rm rev}$ of sorted T_0 in the first column. From the $x_{(i)}$ construct the intervals I_i defined above.

3. Denote by $\hat{F}$ the CDF of the maximum-entropy distribution defined above. Generate n random numbers p_i, $i = 1, \ldots, n$ distributed uniformly on $[0, 1]$. Obtain a resample x_i^* as the p_i quantiles of $\hat{F}$, $i = 1, \ldots, n$.

5. Define another $n \times 2$ sorting matrix S_2. Sort the x_i^* in increasing order and place the result in column 1 of S_2. Place the vector $I_{\rm rev}$ in column 2.

6. Sort the S_2 matrix with respect to the second column to restore the order $\{1, 2, \ldots, n\}$ there. Redefine the x_i^* as the elements of the jointly sorted column 1 of S_2.

The idea is clearly to preserve as much of the correlation structure of the original series as possible. It is a pity that Vinod went on directly to apply his method to real data, as it turns out that altogether too many of the specific properties of the original series are retained in each bootstrap sample, so that there is not enough variability in the bootstrap DGP.

I have documented this method in full because, although it does not work, it shows up a number of interesting things. First, resampling from the continuous distribution $\hat{F}$ can very well be employed instead of resampling from the discrete empirical distribution. Rescaling, and other operations that specify higher moments, can easily be incorporated into the maximum entropy algorithm. Although in most cases one may expect there to be little difference relative to conventional resampling, there are situations in which it may be necessary to impose the continuity of the bootstrap distribution.

The other reason for my dwelling on this method is that it allows me to demonstrate a technique for analysing bootstrap success or failure. Consider the following model, which I will use as a test case for this and the other bootstrapping methods considered here:

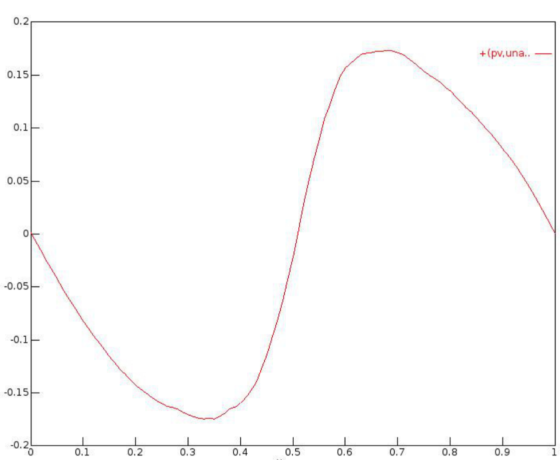
$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}, \quad u_t = \rho u_{t-1} + v_t. \tag{22}$$

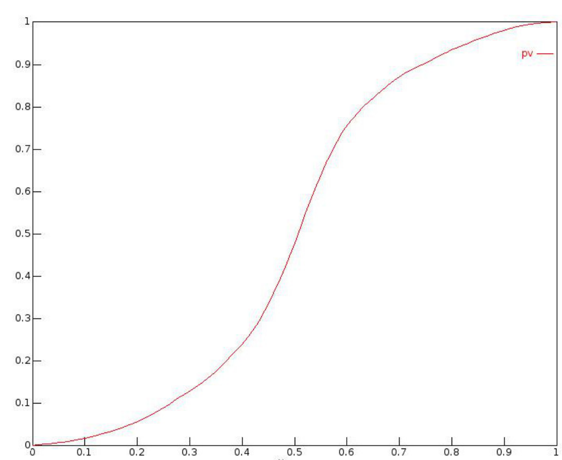
The regressor matrix $\boldsymbol{X}$ includes a constant and three other variables, constructed so that they are serially correlated with autocorrelation coefficient ρ_1. The disturbances follow an AR(1) process. The null hypothesis is that the full coefficient vector $\boldsymbol{\beta} = \boldsymbol{0}$. The test statistic is the asymptotic chi-squared statistic, with four degrees of freedom:

$$\tau = \boldsymbol{y}^\top \boldsymbol{X}(\boldsymbol{X}^\top \hat{\boldsymbol{\Omega}} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y},$$

where $\hat{\boldsymbol{\Omega}}$ is the Newey-West HAC covariance matrix estimator.

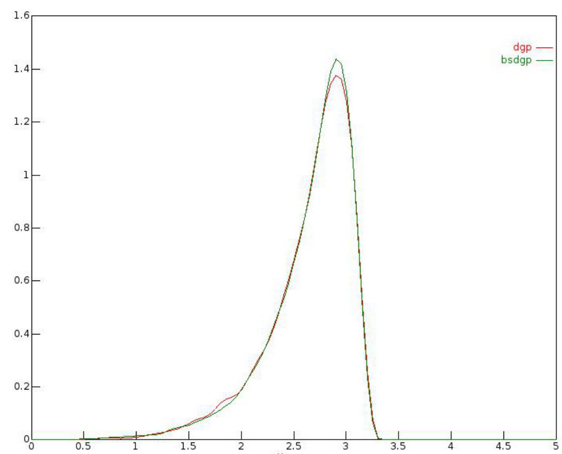
Below are the P value discrepancy and P value plots for $n = 50$, $\rho = 0.9$, $\rho_1 = 0.8$, and a lag-truncation parameter $p = 20$ for $\hat{\boldsymbol{\Omega}}$. There are $9,999$ replications with 399 bootstrap repetitions each.





It is quite clear that something is badly wrong! There is severe underrejection for small α, and equally severe overrejection for large α. There are at least two possible reasons for this. The first is that, if the distribution of the bootstrap statistic is on average more dispersed than that of the statistic itself, then the mass in the bootstrap distribution to the right of τ is too great for large values of τ, so that the P value is also too great, and it is too small when τ is small, so that the the P value is also too small. A second possible explanation is that, for each replication, the bootstrap statistics are strongly positively correlated with τ. In that event, when τ is large, the bootstrap distribution is shifted right, and conversely.

In order to diagnose the bootstrap failure we see in the graphs above, I plotted the density of the actual statistic using a kernel estimate based on the 9,999 realisations of τ, and the density of 9,999 bootstrap statistics, one from each replication. The result can be seen in the next figure, the density of τ in red, that of the bootstrap statistics in green. Clearly, the distributions are almost identical, which rules out the first possible explanation.



Next, for each replication, I saved the realisation of τ and the last of the 399 bootstrap statistics, and regressed the latter on a constant and the former. If the bootstrap statistic is denoted τ^*, the result of the regression was

$$\tau^* = 0.51 + 0.81\tau, \quad \text{centred } R^2 = 0.659.$$

Both coefficients are highly significant. Thus this is clear evidence of the second possible explanation: τ^* is indeed strongly positively correlated with τ. What this shows is that the attempt to make the bootstrapped time series mimic the real series is too successful, and so there is too little variation in the bootstrap distribution.

This analysis is closely related to the rationale behind the fast double bootstrap; see later. It can be applied to the other techniques used here, with quite different results.

5. The Fast Double Bootstrap

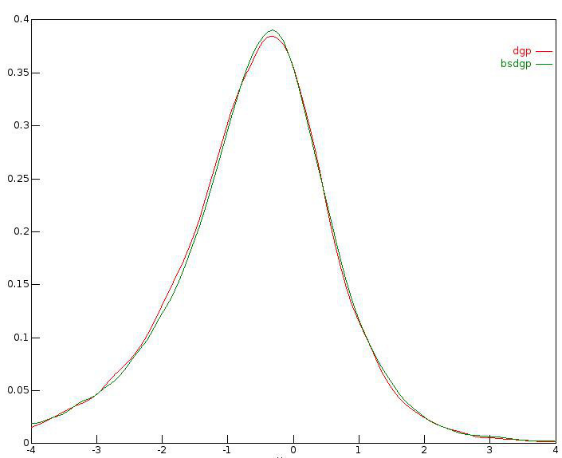
The fast double bootstrap (FDB) of Davidson and MacKinnon (2007) is based on two approximations. The first is to assume that, for any DGP in the null hypothesis, the statistic τ and the bootstrap DGP are independent. The assumption is of course false except in special circumstances, but it holds asymptotically in many commonly encountered situations. Combined with this assumption, the second assumption leads to an approximate expression for the CDF of the bootstrap P value, as follows. Let p_1 denote the bootstrap P value considered as a random variable defined on $[0, 1]$. Then the approximation can be written as

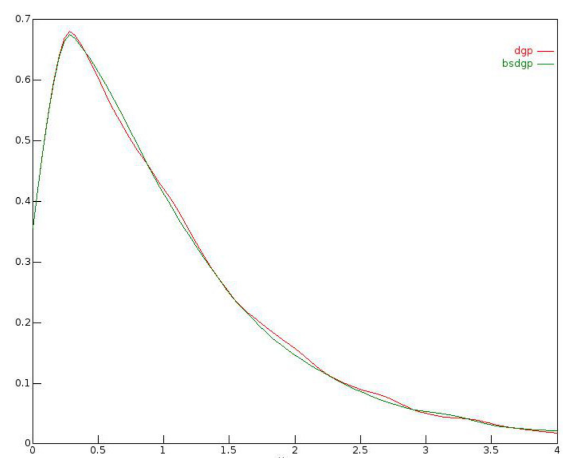
$$\Pr(p_1 \leq \alpha) = R_0\big(Q^1(\alpha)\big), \tag{23}$$

where $R_0(\alpha)$ is the probability that τ is in the α-rejection region according to some, perhaps asymptotic, nominal distribution, and $R^1(\alpha)$ is the same probability but for the bootstrap statistic τ^*. Q^1 is the quantile function that is inverse to R^1.

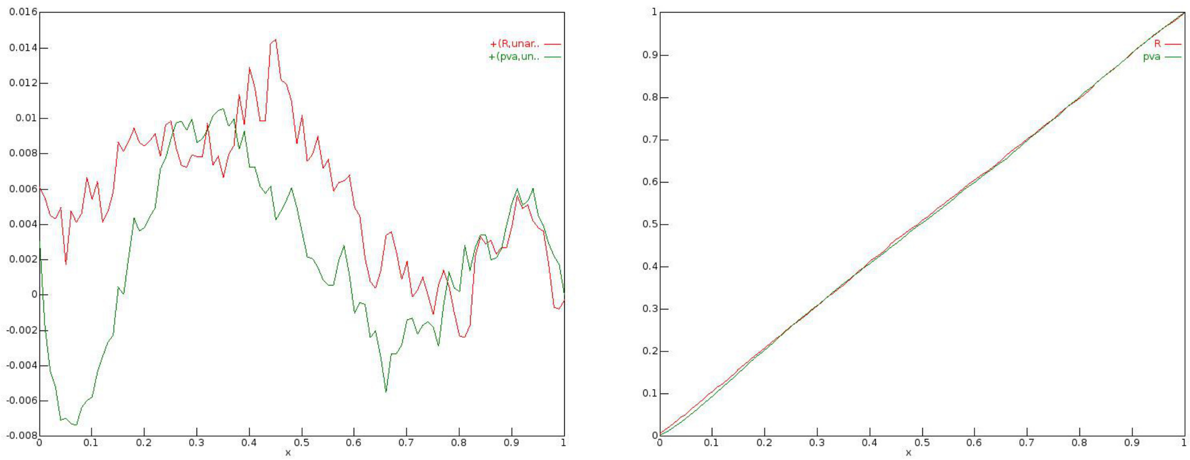
It can be seen that the idea behind the FDB is behind the diagnostic analysis applied to the maximum-entropy bootstrap in Section 5. Clearly the regression of the realisations of τ^* on those of τ allows one to see to what extent the first approximation is reasonable. The applicability of both assumptions jointly can be tested by comparing the approximation (23) with the actual distribution of the bootstrap P value, as estimated by simulation. For the case of the maximum-entropy bootstrap, (23) gives a distribution not far removed from the uniform, a conclusion to be expected given the near coincidence of the distributions of τ and τ^*. The true distribution of the bootstrap P value is very different from the uniform, as shown graphically in the figure in Section 5.

Here, the same analysis is applied to some of the other bootstraps we have considered. For the wild bootstrap in the presence of heteroskedasticity alone, since we considered a one-degree-of-freedom test, it was possible to look separately at a one-tailed test that rejects to the right and the two-tailed test. Below are plotted the kernel density estimates of the distributions of the statistic and the bootstrap statistic for both cases.



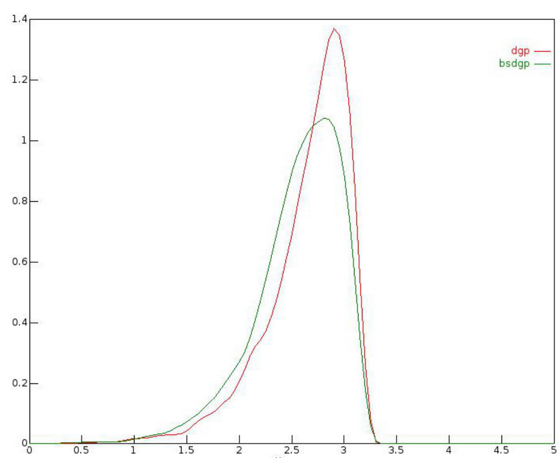


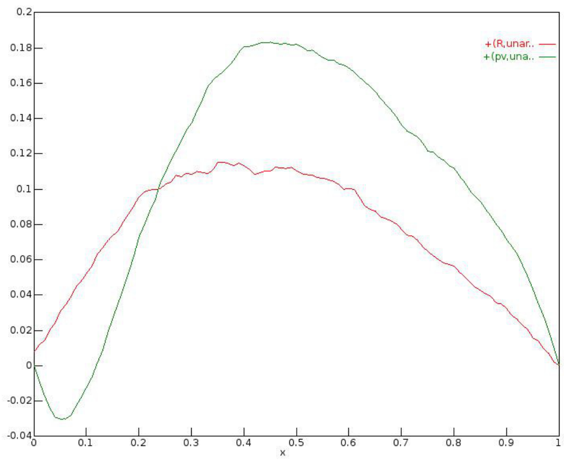
For the one-tailed test, the regression of the bootstrap statistic τ^* on a constant and τ gives a barely marginally significant coefficient for τ, with a t ratio of 2.03 and a centred R^2 of 0.004. For the two-tailed test, the t ratio for the coefficient of τ is 3.18, and the centred R^2 is 0.010.

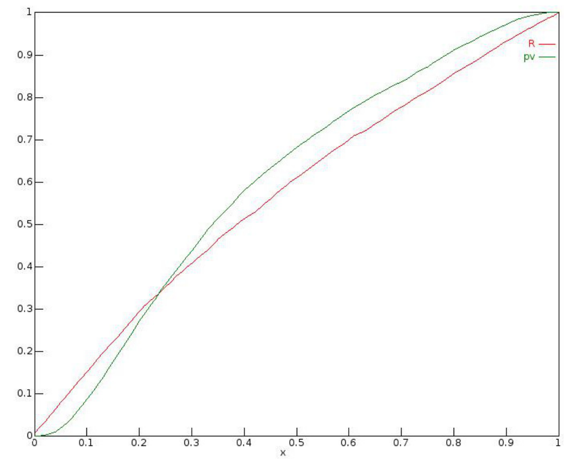


The figure above illustrates the comparison of (23) with the simulated distribution of the bootstrap P value, as measured by discrepancies from uniform to the left, and by actual probabilities to the right. The curves for (23) are in red, those for the P value distribution in green. Although the discrepancies from uniform are very small, one cannot reject the hypothesis that (23) captures those that exist rather well.

Next I look at the extended DWB, again for the standard setup.







The regression of τ^* on τ gives a coefficient of 0.25 and t ratio of 25 and centred R^2 of 0.06.

A number of regularities appear in the simulations with the extended DWB. When the bandwidth is small, that is, p is small relative to the sample size n, the correlation between τ^* and τ is smaller than for larger p, and the agreement between (23) and the actual distribution of the P value better. On the other hand, the brute bootstrap discrepancy at nominal 5% is minimised for some larger p.

6. Simulation Evidence

Since the main focus of this paper is the HAC wild bootstrap, in this section we give more detailed simulation results for it. We also consider how to apply the FDB to it, in order to see whether the performance enhancement suggested by the results in the previous section are realised. The table below gives the bootstrap discrepancies, estimated by simulation, for a variety of choices of the parameters of the standard setup.

n	p	ρ	ρ_1	bootstrap original	discrepancy modified
20	2	0.9	0.0	0.263	0.250
20	6	0.9	0.0	0.016	-0.029
20	6	0.9	0.8	-0.008	-0.024
20	8	0.9	0.0	-0.016	-0.040
20	8	0.9	0.8	-0.037	-0.042
50	10	0.9	0.0	0.084	-0.024
50	10	0.9	0.8	0.098	0.051
50	20	0.9	0.0	0.009	-0.027
50	20	0.9	0.8	-0.007	-0.012
50	30	0.9	0.0	-0.043	-0.049
50	30	0.9	0.8	-0.038	-0.023
100	10	0.9	0.0	0.098	0.083
100	10	0.9	0.8	0.107	0.099
100	25	0.9	0.0	0.038	-0.012
100	25	0.9	0.8	0.030	-0.011
100	40	0.9	0.0	-0.007	-0.045
100	40	0.9	0.8	-0.016	-0.041
100	60	0.9	0.0	-0.046	-0.050
100	60	0.9	0.8	-0.046	-0.047

7. Conclusions

Bootstrapping in the presence of heteroskedasticity and autocorrelation of unknown form has revealed that the wild bootstrap is able to handle any problems associated with heteroskedasticity by itself, but that autocorrelation presents a more severe challenge. The dependent wild bootstrap, extended here to what we have called the HAC wild bootstrap, is capable of giving performance as good as, and often better than, other methods that take autocorrelation into account. It seems to hold a good deal of promise, although its reliability depends sensitively on the choice of lag-truncation parameter. In combination with the fast double bootstrap, it can yield quite satisfactory inference. We also developed an interesting diagnostic technique related to the fast double bootstrap, for analysing bootstrap success or failure.

References

Andrews, D. W. K. and J. C. Monahan (1992). "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator", *Econometrica* 60, 953–66.

Davidson, R. and E. Flachaire (2008). "The wild bootstrap, tamed at last", *Journal of Econometrics*, 146, 162–9.

Davidson, R. and J. G. MacKinnon (2007). "Improving the Reliability of Bootstrap Tests with the Fast Double Bootstrap", *Computational Statistics and Data Analysis*, **51**, 3259–3281.

Davidson, R. and J. G. MacKinnon (2010). "Wild bootstrap tests for IV regession", *Journal of Business and Economic Statistics*, 28, 128–44.

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.

Fiebig Denzil, G. and H. Theil (1982). Comment, *Econometric Reviews* 1:2, 263–269

Freedman, D. A., (1981). "Bootstrapping regression models", *Annals of Statistics* 9, 1218–1228.

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

Jaynes. E. T. (1957). "Information Theory and Statistical Mechanics", *Physical Review* 106, 620–630.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*, New York: Springer.

Liu, R. Y. (1988). "Bootstrap procedures 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.

Newey, W. K. and K. D. West (1987). "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica* 55, 703–8.

Shao, X. (2010). "The Dependent Wild Bootstrap", *Journal of the American Statistical Association* 105, 218–235.

Theil, H. and K. Laitinen (1980). "Singular moment matrices in applied econometrics", in *Mu1tivariate Analysis - V* (P.R. Krishnaiah, Ed.). Amsterdam: North-Holland Pub1ishing Co., 629–649.

Vinod, H.D. (2006). "Maximum entropy ensembles for time series inference in economics", *Journal of Asian Economics* 17, 955–978.

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.