

A better way to bootstrap pairs

Emmanuel Flachaire

GREQAM - Université de la Méditerranée
CORE - Université Catholique de Louvain

April 1999

Abstract

In this paper we are interested in heteroskedastic regression models, for which an appropriate bootstrap method is bootstrapping pairs, proposed by Freedman (1981). We propose an ameliorate version of it, with better numerical performance.

JEL classification : C1

Key-words : bootstrap, heteroskedasticity.

1 Introduction

Bootstrap methods can be of great use in econometrics. They provide reliable inference in small sample sizes, and permit the use of statistics with properties that cannot be calculated analytically. Efron (1979) first introduced these statistical methods. In econometrics, see Horowitz (1994, 1997), Hall and Horowitz (1996), Li and Maddala (1996), and Davidson and MacKinnon (1996, 1999). Correctly used, the bootstrap can often yield large asymptotic refinements.

Theoretical developments show that, if the test statistic is a **pivot**, the bootstrap yields exact inference and, if it is an **asymptotic pivot**, bootstrap inference is more reliable than asymptotic inference. If we consider an econometric model and a null hypothesis, a statistic is said to be pivotal if its distribution is the same for all data generating processes (DGPs) that satisfy the null. It is an asymptotic pivot if its asymptotic distribution is the same for all such DGPs. In econometrics, since the vast majority of tests are asymptotically pivotal, the bootstrap has a wide range of potential application.

In this paper we are interested in heteroskedastic regression models, for which an appropriate version of the bootstrap is **bootstrapping pairs**, proposed by Freedman (1981). In theory, bootstrapping pairs seems to be preferable to other classical bootstrap implementations, because this procedure is valid even if error terms are heteroskedastic. In practice, Monte Carlo studies show that this method has poor numerical performance, see Horowitz (1997).

We propose a **better way** to bootstrap pairs. In section 2, we present the classical bootstrap. In section 3, we present bootstrapping pairs. In section 4, we present a new implementation. Finally, in section 5, we investigate Monte Carlo simulations.

2 Classical bootstrap

Consider the non-linear regression model with independent and identically distributed (IID) error terms,

$$\mathbf{y}_t = \mathbf{x}_t(\boldsymbol{\beta}, \boldsymbol{\gamma}) + \mathbf{u}_t \quad \mathbf{u}_t \sim \mathbf{F}(\mathbf{0}, \boldsymbol{\sigma}^2) \quad (1)$$

where \mathbf{y}_t is the dependent variable, $\mathbf{x}_t(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is a regression function that determines the mean value of \mathbf{y}_t conditional on $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and on some exogenous regressors \mathbf{Z}_t , and \mathbf{u}_t is an error term drawn from an unknown distribution \mathbf{F} with mean zero and variance $\boldsymbol{\sigma}^2$. The statistic $\boldsymbol{\tau}$, which tests the null hypothesis $\mathbf{H}_0 : \boldsymbol{\gamma} = \mathbf{0}$, is supposed to be an asymptotic pivot.

The bootstrap principle is to construct a data generating process, called the bootstrap DGP, based on estimates of the unknown parameters and probability distribution. If the DGP is completely specified up to unknown parameters, as is the case if we know the function \mathbf{F} , we use what is called a **parametric bootstrap**. If the function \mathbf{F} is unknown, we use the empirical distribution function (EDF) of the residuals, which is a consistent estimator of the cumulative distribution function (CDF) of the error terms. This is called a **non-parametric bootstrap**. The distribution of the test statistic under this artificial DGP is called the **bootstrap distribution**, and a P -value based on it

is called a **bootstrap P -value**. We can rarely calculate this distribution analytically, and so we approximate it by simulations. That is why the bootstrap is often considered a computer-intensive method. There are many ways to specify the bootstrap DGP. The key requirement is that it should satisfy the restrictions of the null hypothesis.

A **first refinement** is ensured if the test statistic and the bootstrap DGP are asymptotically independent. Davidson and MacKinnon (1999) show that, in this case, the error of the bootstrap P -value is in general of the same order as $n^{-3/2}$, compared with the asymptotic P -value, of which the error is of the same order as $n^{-1/2}$. This state of affairs is often achieved by using the fact that parameter estimates under the null are asymptotically independent of the statistics associated with tests of that null. The proof of Davidson and MacKinnon (1987) for the case of classical test statistics based on maximum likelihood estimation can be extended in regular cases to NLS, GMM and other forms of extremum estimation. Thus, for the parametric bootstrap, the condition is always satisfied if the bootstrap DGP is constructed with parameter estimates under the null. In the non-parametric case, the bootstrap statistic has to be asymptotically independent of the regression parameters and of the resampling distribution too. Davidson and MacKinnon (1999) show that we respect this hypothesis whether we use the EDF of the residuals under the null or under the alternative. Monte-Carlo experiments show that residuals under the null provide a substantial efficiency gain, see van Giersbergen and Kiviet (1994).

A **second refinement** can be obtained, with the non-parametric bootstrap, if we use modified residuals, such that variance of the resampling distribution is an unbiased estimate of the variance of the error terms. We use the fact that, in the linear regression model $\mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta} + \mathbf{u}_t$ with IID error terms, $E(\hat{\mathbf{u}}_t^2) = (\mathbf{1} - \mathbf{h}_t)\sigma^2$, where $\hat{\mathbf{u}}_t$ is a residual and $\mathbf{h}_t = \mathbf{X}_t(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}_t^\top$. To correct the implied bias, we transform the residuals as

$$\tilde{\mathbf{u}}_t^{(2)} = \frac{\hat{\mathbf{u}}_t}{(\mathbf{1} - \mathbf{h}_t)^{1/2}} - \frac{1}{n} \sum_{s=1}^n \frac{\hat{\mathbf{u}}_s}{(\mathbf{1} - \mathbf{h}_s)^{1/2}} \quad (2)$$

We divide $\hat{\mathbf{u}}_t$ by a factor proportional to the square root of its variance. The raw residuals $\hat{\mathbf{u}}_t$ do not all have the same variance: there is an artificial heteroskedasticity. The modified residuals have the same variance and are centered. In the non-linear case these developments are still valid asymptotically, and, to leading order, we use $\hat{\mathbf{h}}_t \equiv \hat{\mathbf{X}}_t(\hat{\mathbf{X}}_t^\top \hat{\mathbf{X}}_t)^{-1} \hat{\mathbf{X}}_t^\top$, where $\hat{\mathbf{X}}_t = \mathbf{X}_t(\hat{\boldsymbol{\beta}})$ is the vector of partial derivatives of $\mathbf{x}_t(\boldsymbol{\beta})$ with respect of $\boldsymbol{\beta}$, evaluated at $\hat{\boldsymbol{\beta}}$; see for example Davidson and MacKinnon(1993, pp 167 et 179). Another way to make the variance of the resampling distribution equal to the unbiased estimator of the variance of the error terms is to use the fact that $E(\hat{\mathbf{u}}^\top \hat{\mathbf{u}}) = \sigma^2(\mathbf{n} - \mathbf{k})$, where \mathbf{k} is the number of regressors. We can correct the bias by multiplying the residuals by the square root of $\mathbf{n}/(\mathbf{n} - \mathbf{k})$ and recentering them. An inconvenient aspect of this transformation is that, in a lot of cases, for example with an F -statistic, this transformation has no effect, because the statistic is invariant with respect to the scale of the variance of the error terms.

Finally, the **classical bootstrap** is based on the bootstrap DGP

$$\mathbf{y}_t^* = \mathbf{x}_t(\tilde{\boldsymbol{\beta}}, \mathbf{0}) + \mathbf{u}_t^* \quad (3)$$

where $\tilde{\beta}$ is the vector of parameter estimates under the null, and \mathbf{u}_t^* a random drawing from $\tilde{\mathbf{u}}_t^{(2)}$, the residuals under the null transformed by (2).

3 Bootstrapping pairs

Consider the non-linear regression model with heteroskedastic error terms,

$$\mathbf{y}_t = \mathbf{x}_t(\beta, \gamma) + \mathbf{u}_t \quad \mathbf{E}(\mathbf{u}_t) = \mathbf{0}, \quad \mathbf{E}(\mathbf{u}_t^2) = \sigma_t^2 \quad (4)$$

The difference with respect to the model (1) is just that the error terms may have different variances in the model (4). If the heteroskedasticity is of unknown form, we cannot use the classical bootstrap. The heteroskedasticity could depend on some elements of the function \mathbf{x}_t , such as the exogenous variables \mathbf{Z} , and so we cannot resample residuals independently of these elements. Freedman (1981) proposed to resample directly from the original data: that is, to resample the couple *dependent variable* and *regressors* (\mathbf{y}, \mathbf{Z}), this is called **bootstrapping pairs**. Note that this method is not valid if \mathbf{Z} contains lagged values of \mathbf{y} . The original data set satisfied the relation $\mathbf{y}_t = \mathbf{x}_t(\hat{\beta}, \hat{\gamma}) + \hat{\mathbf{u}}_t$ where $\hat{\beta}$ and $\hat{\gamma}$ are the non-linear least squares (NLS) parameter estimates, and $\hat{\mathbf{u}}_t$ the residuals. Thus resampling (\mathbf{y}, \mathbf{Z}) is equivalent to resampling $(\mathbf{Z}, \hat{\mathbf{u}})$ and then generating the dependent variable with the bootstrap DGP $\mathbf{y}_t^* = \mathbf{x}_t^*(\hat{\beta}, \hat{\gamma}) + \mathbf{u}_t^*$, where $\mathbf{x}_t^*(\cdot)$ is defined using the resampled \mathbf{Z} , and \mathbf{u}_t^* is the resampled $\hat{\mathbf{u}}$. The bootstrap statistic has to test a true hypothesis, and so the null hypothesis has to be modified to become $\mathbf{H}'_0 : \gamma = \hat{\gamma}$. A detailed discussion can be found in the introduction of the book of Hall (1992).

This method has two major drawbacks. The first is that the bootstrap DGP is not constructed with parameter estimates under the null hypothesis: the extra refinement from Davidson and MacKinnon (1999) is not ensured. We correct this in the new implementation proposed in the next section. The second is that we draw the dependent variable and the regressors at the same time: the regressors in the bootstrap DGP are therefore not exogenous. This second drawback is intrinsic to the nature of this bootstrap method, and if we wish to correct it, we should use another method proposed by Wu (1986), called the wild bootstrap.

4 New implementation

We propose to use the bootstrap DGP

$$\mathbf{y}_t^* = \mathbf{x}_t^*(\tilde{\beta}, \mathbf{0}) + \mathbf{u}_t^* \quad (5)$$

where $\tilde{\beta}$ is the vector of parameter estimates under the null. To generate the bootstrap sample $(\mathbf{y}^*, \mathbf{Z}^*)$, we first resample from the set of couples $(\mathbf{Z}, \hat{\mathbf{u}}^{(2)})$, where $\hat{\mathbf{u}}^{(2)}$ are the residuals under the alternative, transformed by (2), and the dependent variable \mathbf{y}^* is calculated using (5). With this choice, we obtain the two refinements of the classical bootstrap.

The **first refinement** is obtained because the bootstrap DGP respects the null hypothesis, and so the statistic is asymptotically independent of the bootstrap DGP. The extra refinement from Davidson and MacKinnon (1999) is ensured.

The **second refinement** is due to the use of the transformed residuals $\hat{\mathbf{u}}_t^{(2)}$ from estimating the alternative hypothesis. Let us show that we cannot use residuals under the null as in the classical bootstrap. Consider the linear regression model: $\mathbf{y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{u}$ and the null hypothesis $\mathbf{H}_0 : \boldsymbol{\gamma} = \mathbf{0}$. In this case, the bootstrap DGP we proposed is $\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\gamma}_0 + \mathbf{u}^*$, where $\boldsymbol{\gamma}_0 = \mathbf{0}$ and $(\mathbf{X}^*, \mathbf{u}^*)$ is drawn from $(\mathbf{X}, \ddot{\mathbf{u}})$. The bootstrap parameter estimate is equal to:

$$\hat{\boldsymbol{\gamma}}^* = (\mathbf{X}^{*\top} \mathbf{X}^*)^{-1} \mathbf{X}^{*\top} \mathbf{y}^* = \boldsymbol{\gamma}_0 + (\mathbf{X}^{*\top} \mathbf{X}^*)^{-1} \mathbf{X}^{*\top} \mathbf{u}^*. \quad (6)$$

It is a consistent estimator if

$$\text{plim}_{n \rightarrow \infty} (\hat{\boldsymbol{\gamma}}^* - \boldsymbol{\gamma}_0) = \text{plim}_{n \rightarrow \infty} \frac{1}{n} (\mathbf{X}^{*\top} \mathbf{X}^*)^{-1} \text{plim}_{n \rightarrow \infty} \frac{1}{n} (\mathbf{X}^{*\top} \mathbf{u}^*) = \mathbf{0} \quad (7)$$

where

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} (\mathbf{X}^{*\top} \mathbf{u}^*) = \text{plim}_{n \rightarrow \infty} \frac{1}{n} (\mathbf{X}^\top \ddot{\mathbf{u}}) \quad (8)$$

This expression is equal to $\mathbf{0}$ if $\ddot{\mathbf{u}}$ and \mathbf{X} are asymptotically orthogonal. The residuals $\hat{\mathbf{u}}$ computed under the alternative and \mathbf{X} are orthogonal by construction, but this is not true of the residuals under the null. Consequently, if $\ddot{\mathbf{u}} = \hat{\mathbf{u}}$, the bootstrap parameter estimates are consistent, but not otherwise. When n tends to infinity, $\mathbf{h}_t = \mathbf{X}_t (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}_t$ tends to $\mathbf{0}$. If $\ddot{\mathbf{u}}$ corresponds to the residuals under the alternative transformed by (2), $\ddot{\mathbf{u}}$ and \mathbf{X} are asymptotically orthogonal and the bootstrap estimate is consistent.

5 Simulations

We consider a small Monte Carlo experiment designed to study the performance of the new implementation. The Monte Carlo design is the same as in Horowitz (1997). We consider a linear regression model: $\mathbf{y} = \boldsymbol{\alpha} + \mathbf{x}\boldsymbol{\beta} + \mathbf{u}$, with an intercept and one regressor sampled from $\mathcal{N}(\mathbf{0}, \mathbf{1})$ with probability $\mathbf{0.9}$ and from $\mathcal{N}(\mathbf{2}, \mathbf{9})$ with probability $\mathbf{0.1}$. The sample size is small $n = \mathbf{20}$ and the number of simulations large $\mathbf{S} = \mathbf{20,000}$, $\boldsymbol{\alpha} = \mathbf{1}$, $\boldsymbol{\beta} = \mathbf{0}$ and the null hypothesis is $\mathbf{H}_0 : \boldsymbol{\beta} = \mathbf{0}$. The variance of error term t is either $\mathbf{1}$ or $\mathbf{1} + \mathbf{x}^2$, according to whether the error terms are IID or heteroskedastic. We use a t -test using the heteroskedasticity-consistent covariance matrix estimator (HCCME) of Eicker (1963) and White (1980). In fact, we use two HCCMEs. The first is called the *naive HCCME*, and is constructed with the residuals $\hat{\mathbf{u}}_t$. The second is calculated with the transformed residuals $\hat{\mathbf{u}}_t / (\mathbf{1} - \mathbf{h}_t)^{-1}$. This transformation improves the numerical performance of the t -test, being an approximation to the “jackknife” estimator, see MacKinnon and White (1985) and Davidson and MacKinnon (1993, pp 554). We call it *jack-approx HCCME*.

	<i>naive HCCME</i>			<i>jack-approx HCCME</i>		
	<i>asympt</i>	<i>boot pairs</i>	<i>boot new</i>	<i>asympt</i>	<i>boot pairs</i>	<i>boot new</i>
<i>IID</i>	0.173	0.115	0.093	0.082	0.087	0.073
<i>HET</i>	0.247	0.121	0.098	0.118	0.090	0.075

Table 1: Empirical level of t -test using HCCME, at nominal 0.05 level

In all the cases considered the new implementation improves the reliability of the t -test. In the presence of heteroskedasticity, the empirical level of the asymptotic t -test using the naive HCCME equals **0.247**. The use of transformed residuals and of the new implementation of the bootstrap by pairs, corrects the empirical level to **0.075**.

Acknowledgements

I am specially indebted to Russell Davidson for his helpful comments. All remaining errors are mine.

References

- Davidson, R. and J. G. MacKinnon (1987). “Implicit alternatives and the local power of test statistics”. *Econometrica* 55, 1305–29.
- Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- Davidson, R. and J. G. MacKinnon (1996). “The power of bootstrap tests”. Queen’s University Institute for Economic Research, Discussion paper 937.
- Davidson, R. and J. G. MacKinnon (1999). “The size distortion of bootstrap tests”. *Econometric Theory* 15, 361–376.
- Efron, B. (1979). “Bootstrap methods: another look at the jackknife”. *Annals of Statistics* 7, 1–26.
- Eicker, B. (1963). “Limit theorems for regression with unequal and dependant errors”. *Annals of Mathematical statistics* 34, 447–456.
- Freedman, D. A. (1981). “Bootstrapping regression models”. *Annals of Statistics* 9, 1218–1228.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer Series in Statistics. New York: Springer Verlag.
- Hall, P. and J. L. Horowitz (1996). “Bootstrap critical values for tests based on generalized method of moments estimators”. *Econometrica* 64, 891–916.
- Horowitz, J. L. (1994). “Bootstrap-based critical values for the information matrix test”. *Journal of Econometrics* 61, 395–411.

- Horowitz, J. L. (1997). “Bootstrap methods in econometrics: theory and numerical performance”. In *Advances in Economics and Econometrics: Theory and Application*, Volume 3, pp. 188–222. David M. Kreps and Kenneth F. Wallis (eds), Cambridge, Cambridge University Press.
- Li, H. and G. S. Maddala (1996). “Bootstrapping time series models”. *Econometric Reviews* 15, 115–158.
- MacKinnon, J. G. and H. L. White (1985). “Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties”. *Journal of Econometrics* 21, 53–70.
- van Giersbergen, N. P. A. and J. F. Kiviet (1994). “How to implement bootstrap hypothesis testing in static and dynamic regression model”. Discussion paper TI 94-130, Amsterdam: Tinbergen Institute. Paper presented at *ESEM*'94 and *EC*²'93.
- 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.