Improving the Reliability of Bootstrap Confidence Intervals

by

Russell Davidson

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

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

russell@ehess.cnrs-mrs.fr

and

James G. MacKinnon

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

jgm@qed.econ.queensu.ca

Abstract

This paper investigates the relation between hypothesis testing and the construction of confidence intervals, with particular regard to bootstrap tests. In practice, confidence intervals are almost always based on Wald tests, and consequently are not invariant under nonlinear reparametrisations. Bootstrap percentile-t confidence intervals are an instance of this. However, the (asymptotically) pivotal functions of data and parameters on which likelihood ratio (LR) and Lagrange multiplier (LM) tests depend can be used to construct parametrisation-invariant confidence intervals. We show that, whenever an artificial regression can be used to find the restricted estimates needed for LR and LM tests, the nonlinear equations that define the limits of a confidence interval can be solved by an algorithm based on the same artificial regression. The algorithm involves roughly as much computation for each interval limit as is needed to find the restricted estimates.

Bootstrap tests are often more reliable when the bootstrap DGP is based on restricted estimates. Inverting such tests to find a confidence interval is computationally intensive, since many bootstrap samples must be generated for every set of restricted estimates considered. We show how to combine artificial regression based bootstrap testing with the algorithm for finding limits of confidence intervals.

This research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada.

May 2001

1. Introduction

It is commonly asserted that there is a duality between hypothesis testing and confidence intervals, and that, for every result applying to one of these, there is an analogous one applying to the other. This is true, of course, but it takes a certain amount of interpretation to be made operational. It is more useful to say that a confidence interval corresponds to a *family* of hypothesis tests, and that the P value of a hypothesis test corresponds to a family of confidence intervals. We are concerned here with the first part of this duality.

It is rare that, in practice, a confidence interval is constructed so as to correspond to anything other than a family of Wald tests. This is true not only of conventional asymptotic confidence intervals but also of conventional bootstrap confidence intervals. There are some exceptions: Dufour (1997) promotes the use of confidence sets based on Likelihood Ratio tests, and points out that similar procedures were suggested more than half a century ago by Anderson and Rubin (1949). Most often, though, the computational simplicity of confidence intervals based on families of Wald tests wins the day. This simplicity is bought at the price of non-invariance of the confidence intervals under reparametrisation.

In this paper, we consider, from an algorithmic point of view, how to construct confidence intervals based on other families of tests, especially Lagrange Multiplier tests, and, for contexts other than maximum likelihood, generalised score tests. These tests, like the Likelihood Ratio test, are invariant under reparametrisation, and so, therefore, are confidence intervals based on them. The key to our approach is the use of artificial regressions. It is very frequently possible to compute test statistics by use of these, and we have shown in Davidson and MacKinnon (1999b) that the time needed to compute bootstrap P values can be substantially reduced by their use. The main contribution of this paper is to extend existing artificial regression methods to the construction of confidence intervals.

In the following section, we look more closely at the connections between hypothesis tests and confidence intervals. Then, in Section 3, we define what we mean by an artificial regression, and construct algorithms that use them to compute confidence intervals corresponding to families of Lagrange Multiplier and Likelihood Ratio tests. The computational burden of these algorithms is comparable to that of the other algorithms that use artificial regressions, such as obtaining nonlinear estimates or computing Lagrange Multiplier test statistics. Section 4 moves on to bootstrap tests and confidence intervals, and discusses the use of the estimating function bootstrap (Hu and Kalbfleisch (2000)) for constructing them. This method generates bootstrap statistics in a way that does not depend on restricted parameter estimates, and so another method is developed, in which each test of the family used to construct the confidence interval is based on bootstrapping with restricted estimates. Artificial regression based algorithms are presented for both sorts of interval, and it is seen that the estimating function bootstrap is a good deal less computationally costly than the other method. In Section 6, simulation experiments are presented in which

different sorts of bootstrap confidence interval are compared for a parameter of a nonlinear regression model. Section 7 contains a few concluding remarks.

2. Relation between Hypothesis Tests and Confidence Intervals

Denote by θ the parameter for which it is desired to construct a confidence interval with nominal coverage of $1 - \alpha$. Suppose that, for each specific value θ_0 of the parameter, we can conduct a test of the hypothesis that $\theta = \theta_0$ with nominal level α. Denote by $T(\theta_0)$ the (binary) result of the test: $T(\theta_0) = 1$ if the test rejects the hypothesis that $\theta = \theta_0$, 0 if not. Then the confidence interval that corresponds to the family of hypothesis tests is the set of those values θ_0 for which the hypothesis that $\theta = \theta_0$ is not rejected by the appropriate test, that is, the set

$$\{\theta \mid T(\theta) = 0\}. \tag{1}$$

The coverage of this confidence interval is exactly $1 - \alpha$ if each test in the family rejects with probability exactly α under its own null.

A confidence interval for θ is most commonly constructed by computing an estimate $\hat{\theta}$ of θ and the standard error $\hat{\sigma}_\theta$ of the estimate. The confidence interval then takes the form

$$[\hat{\theta} - c_{1-\alpha/2}\,\hat{\sigma}_\theta,\ \hat{\theta} - c_{\alpha/2}\hat{\sigma}_\theta], \tag{2}$$

where $c_{1-\alpha/2}$ and $c_{\alpha/2}$ are respectively the $1-\alpha/2$ and $\alpha/2$ quantiles of the standard normal, or possibly the Student's-t distribution with appropriate degrees of freedom. Implicitly, such a confidence interval corresponds to a family of two-tailed tests where the test statistic for the null hypothesis that $\theta = \theta_0$ is

$$\tau(\theta_0) \equiv \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_\theta}. \tag{3}$$

Under the null, this ratio is asymptotically standard normal, although, in some circumstances, it may be preferable to use a t distribution, if it provides a better approximation to the finite-sample distribution. A two-tailed test rejects the null if the statistic τ is in one of the tails of the chosen distribution, where the tails are defined by the quantiles $c_{1-\alpha/2}$ and $c_{\alpha/2}$ in accordance with the chosen nominal level of the test. It follows immediately that the set (1) becomes the interval (2) for this family of tests, since the boundary points θ_- and θ_+ of (2) satisfy the equations

$$\tau(\theta_+) = c_{\alpha/2} \text{ and } \tau(\theta_-) = c_{1-\alpha/2} \tag{4}$$

which determine the values for which the test is at the margin between rejection and non-rejection. Note that the *upper* limit of the interval is attained when the test statistic reaches the *lower* critical value, and *vice versa*. This feature of the relation between hypothesis tests and confidence intervals requires special attention with

asymmetric intervals, and has given rise to confusion in the past with certain implementations of the bootstrap; witness the so-called *other* percentile method discussed, and condemned, by Hall (1992).

The statistic (3) is asymptotically *pivotal*, by which we mean that it has the property that the right-hand side of (3), which is a function not only of the data through the estimates $\hat{\theta}$ and $\hat{\sigma}_\theta$ but also directly of the parameter θ_0, has the same asymptotic distribution for all values of θ_0. That is, if the data used to compute $\hat{\theta}$ and $\hat{\sigma}_\theta$ are generated by a process for which the true parameter value is θ_0, then $\tau(\theta_0)$, with argument the same θ_0, is asymptotically standard normal whatever the value of θ_0. Almost all test statistics commonly used in econometrics are asymptotically pivotal, although it is rare that they should be exactly pivotal. Use of an at least asymptotically pivotal statistic is, of course, an important condition for asymptotic refinements when the bootstrap is used; see Beran (1988).

The family of tests based on the statistics (3) are Wald tests, based on estimates obtained under the *alternative* hypothesis. It is well known, see Gregory and Veall (1985), Lafontaine and White (1986), and Phillips and Park (1988), that Wald tests have a number of undesirable properties. In particular, they are not invariant under nonlinear reformulations of the restrictions being tested. The consequence of this for confidence intervals is that, if a new parameter ϕ is defined as a nonlinear function of θ, by the transformation $\phi = g(\theta)$ say, where $g(\cdot)$ is a nonlinear function, then the confidence interval based on the estimate $\hat{\phi} \equiv g(\hat{\theta})$ and the standard error

$$\hat{\sigma}_\phi \equiv |g'(\hat{\theta})| \hat{\sigma}_\theta,$$

computed conventionally by the delta method, is not the image under g of the interval (2) found for θ.

A useful way of expressing this failure of invariance is to observe that the asymptotic pivot implicitly used in the construction of a confidence interval for ϕ is not the same as that used for the interval for θ. Indeed, for ϕ, the asymptotic pivot is

$$\frac{\hat{\phi} - \phi_0}{\hat{\sigma}_\phi} = \frac{g(\hat{\theta}) - g(\theta_0)}{|g'(\hat{\theta})|\, \hat{\sigma}_\theta}. \tag{5}$$

It is clear that the right-hand side of (5), for given θ_0, is not deterministically related to (3), on account of the dependence of the derivative $g'(\hat{\theta})$ on $\hat{\theta}$, unless g is an affine function.

Bootstrap confidence intervals constructed using the percentile-t method are known to suffer from the same failure of invariance under nonlinear reparametrisations, and the reason for this is precisely that these confidence intervals, like conventional asymptotic confidence intervals, correspond to a family of Wald tests. Indeed, the only difference between an asymptotic confidence interval like (2) and a percentile-t bootstrap confidence interval is that the standard normal or t distribution of the asymptotic test is replaced by a simulated distribution generated by the bootstrap.

Specifically, bootstrap data sets are generated, by resampling the original data or otherwise, and, for each bootstrap data set, the value of the asymptotically pivotal statistic (3) is computed, using for θ_0 the true value of the parameter for the bootstrap samples. This true value is normally just the estimate $\hat{\theta}$ obtained from the original data. The empirical distribution function of the bootstrap statistics is then used in place of the standard normal distribution in constructing a confidence interval. In particular, the quantities $c_{1-\alpha/2}$ and $c_{\alpha/2}$ are now quantiles of the empirical distribution. Since this distribution is usually asymmetric, the bootstrap opens the possibility of constructing an asymmetric confidence interval, although this is by no means either necessary or desirable in all circumstances. Whether or not the bootstrap confidence interval is asymmetric, it is not invariant under a nonlinear transformation of θ, for the same reason that the asymptotic interval is not invariant, namely that the underlying asymptotic pivots (3) and (5) are random variables that are not related deterministically.

A confidence interval of type (1) can be based on any family of tests capable of testing the hypotheses $\theta = \theta_0$ for all values of θ_0 of interest. Rather than using a family of Wald tests, it is perfectly possible, at least theoretically, to use a family of Likelihood Ratio (LR) or Lagrange Multiplier (LM) tests. Since these tests are invariant under nonlinear reformulations of the restrictions under test, the confidence intervals they generate will also be invariant. There are reasons other than invariance for which one might prefer LR or LM tests. Often their finite-sample properties are closer to nominal than those of Wald tests; LR tests may be Bartlett correctable, *etc.*

A practical difficulty for the construction of a confidence interval based on, say, a family of LM tests is that the equations (4) cannot be solved as simply as for a family of Wald tests, for which the form of the statistic $\tau(\theta)$ as an affine function of θ leads to simple linear equations for the limits $\theta_\pm$ of the confidence interval. These linear equations are completely determined by $\hat{\theta}$ and $\hat{\sigma}_\theta$, whereas, more generally, we have to solve nonlinear equations that depend on more than just a parameter estimate and its standard error.

3. The Role of Artificial Regressions

In many cases, the restricted estimates needed for the construction of an LM or an LR test can be obtained by use of an iterative procedure based on an artificial regression. We use the term "artificial regression" in the sense exposited in Davidson and MacKinnon (2001), or earlier, but in a more limited context, in Davidson and MacKinnon (1990). Briefly, an artificial regression that can be written as

$$\boldsymbol{r}(\boldsymbol{\theta}) = \boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{b} + \text{residuals} \tag{6}$$

is said to correspond to a parametrised model, with parameters $\boldsymbol{\theta}$, and an estimator $\hat{\boldsymbol{\theta}}$ of those parameters, if three conditions are satisfied:

(i) The estimator $\hat{\boldsymbol{\theta}}$ is defined by the equations $\boldsymbol{R}(\hat{\boldsymbol{\theta}})^\top \boldsymbol{r}(\hat{\boldsymbol{\theta}}) = \boldsymbol{0}$;

(ii) for any root-n consistent $\acute{\boldsymbol{\theta}}$, a consistent estimate of $\mathrm{Var}(\operatorname{plim} n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0))$ is given by $n^{-1}\boldsymbol{R}(\acute{\boldsymbol{\theta}})^\top \boldsymbol{R}(\acute{\boldsymbol{\theta}})$. Formally,

$$\mathrm{Var}\left(\operatorname*{plim}_{n\to\infty} n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right) = \operatorname*{plim}_{n\to\infty} \left(n^{-1}\boldsymbol{R}^\top(\acute{\boldsymbol{\theta}})\boldsymbol{R}(\acute{\boldsymbol{\theta}})\right)^{-1};$$

(iii) if $\acute{\boldsymbol{b}}$ denotes the vector of estimates from the artificial regression (6) with regressand and regressors evaluated at $\acute{\boldsymbol{\theta}}$, then

$$\acute{\boldsymbol{\theta}} + \acute{\boldsymbol{b}} = \hat{\boldsymbol{\theta}} + o_p(n^{-1/2}).$$

For more information on artificial regressions, see also Orme (1995). The third condition, often called the *one-step property*, is what makes artificial regressions useful for computing estimates, restricted or unrestricted, and a theorem proved in Davidson and MacKinnon (2001) shows their utility in hypothesis testing. Since we make substantial use of it in this paper, we restate that theorem here. In the theorem statement, and subsequently, vectors and matrices written with an affix, like $\acute{\boldsymbol{r}}$ for instance, are to be interpreted as meaning $\boldsymbol{r}(\acute{\boldsymbol{\theta}})$, where the vector or matrix is evaluated at the value of the parameter vector distinguished by the appropriate affix.

Theorem 1: Let the parameter vector $\boldsymbol{\theta}$ of a parametrised model be partitioned as $\boldsymbol{\theta} = [\boldsymbol{\theta}_1 \vdots \boldsymbol{\theta}_2]$. Suppose that the artificial regression (6) corresponds to this model and to some root-n consistent, asymptotically normal, estimator of $\boldsymbol{\theta}$, and that the partition $\boldsymbol{R} = [\boldsymbol{R}_1 \quad \boldsymbol{R}_2]$ of the regressor matrix corresponds to the partition of $\boldsymbol{\theta}$. Then, in order to test the hypothesis that $\boldsymbol{\theta}_2 = \boldsymbol{0}$, one evaluates the variables of (6) at any root-n consistent estimate $\acute{\boldsymbol{\theta}} = [\acute{\boldsymbol{\theta}}_1 \vdots \boldsymbol{0}]$ of the parameters of the null hypothesis, and computes the statistic

$$\acute{\boldsymbol{r}}^\top \acute{\boldsymbol{M}}_1 \acute{\boldsymbol{R}}_2 (\acute{\boldsymbol{R}}_2^\top \acute{\boldsymbol{M}}_1 \acute{\boldsymbol{R}}_2)^{-1} \acute{\boldsymbol{R}}_2^\top \acute{\boldsymbol{M}}_1 \acute{\boldsymbol{r}}, \tag{7}$$

where $\acute{\boldsymbol{M}}_1 \equiv \mathbf{I} - \acute{\boldsymbol{R}}_1 (\acute{\boldsymbol{R}}_1^\top \boldsymbol{R}_1)^{-1} \acute{\boldsymbol{R}}_1^\top$ is the orthogonal projection on to the orthogonal complement of the space spanned by the columns of $\acute{\boldsymbol{R}}_1$. The statistic (7) is asymptotically distributed as $\chi^2(r)$ under the null hypothesis, where r is the dimension of $\boldsymbol{\theta}_2$, that is, the number of degrees of freedom of the test.

For models estimated by maximum likelihood, the statistic (7) is asymptotically equivalent, under the null and local alternatives, to any of the classical test statistics. More generally, (7) is asymptotically equivalent to more familiar test statistics commonly used in the context of extremum estimators of various sorts, such as GMM estimators. If the consistent estimate $\acute{\boldsymbol{\theta}}$ is in fact the ML estimate of the parameters of the restricted model, the statistic (6) is a version of the classical LM statistic. For other choices of $\acute{\boldsymbol{\theta}}$, it is a $C(\alpha)$ test; see Smith (1987) and Davidson and MacKinnon (1991).

For concreteness, consider a nonlinear regression model:

$$\boldsymbol{y} = \boldsymbol{x}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) + \boldsymbol{u}; \quad \boldsymbol{u} \sim \text{IID}(0, \sigma^2), \tag{8}$$

where the parameter vector $\boldsymbol{\beta}$ is partitioned as $\boldsymbol{\beta} = [\boldsymbol{\beta}_1 \vdots \boldsymbol{\beta}_2]$, and we wish to test the hypothesis that $\boldsymbol{\beta}_2 = \mathbf{0}$. An artificial regression that corresponds to the nonlinear least squares estimator of $\boldsymbol{\beta}$ is the well-known Gauss-Newton regression (GNR), which can be written as

$$\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta}) = \boldsymbol{X}_1(\boldsymbol{\beta})\boldsymbol{b}_1 + \boldsymbol{X}_2(\boldsymbol{\beta})\boldsymbol{b}_2 + \text{residuals}, \tag{9}$$

where the matrices $\boldsymbol{X}_1(\boldsymbol{\beta})$ and $\boldsymbol{X}_2(\boldsymbol{\beta})$ are the Jacobians of the vector $\boldsymbol{x}(\boldsymbol{\beta})$ of regression functions with respect to $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ respectively. Write the restricted NLS estimator of (8) as $\tilde{\boldsymbol{\beta}} \equiv [\tilde{\boldsymbol{\beta}}_1 \vdots \mathbf{0}]$. If the GNR (9) is evaluated at $\tilde{\boldsymbol{\beta}}$, then asymptotically equivalent versions of the LM statistic for the hypothesis $\boldsymbol{\beta}_2 = \mathbf{0}$ include n (the sample size) times the uncentred R^2 of (9), and the ratio of the explained sum of squares to the estimated error variance; see Davidson and MacKinnon (2001) for more details.

For the purposes of a confidence *interval*, rather than a confidence region in which more than one parameter is involved, the most straightforward statistic to use is just the t statistic associated with the column of $\boldsymbol{X}(\tilde{\boldsymbol{\beta}})$ that corresponds to what we can now write as the scalar parameter β_2. The very simplest case one can consider is that in which β is the only parameter of the model. In that case, the testing regression can be written as

$$\boldsymbol{y} - \boldsymbol{x}(\beta_0) = \boldsymbol{x}_\beta(\beta_0)b + \text{residuals}, \tag{10}$$

in what should be obvious notation, and the test statistic for testing the hypothesis that $\beta = \beta_0$ is the t statistic from this regression, which we may conveniently denote by $\tau(\beta_0)$. The limits of the confidence interval defined by the family of LM tests generated as β_0 varies are then the solutions to the equations

$$\tau(\beta_-) = c_{1-\alpha/2} \text{ and } \tau(\beta_+) = c_{\alpha/2}, \tag{11}$$

where the critical values $c_{1-\alpha/2}$ and $c_{\alpha/2}$ are quantiles of the standard normal distribution, or a t distribution, or a bootstrap distribution.

Just this approach to the construction of confidence intervals is proposed in a recent interesting paper by Hu and Kalbfleisch (2000), on what they call the estimating function bootstrap. Much of the material in this section is inspired by their approach, which can be seen to apply more generally than just to the estimating function bootstrap. Hu and Kalbfleisch (henceforth HK) show that confidence intervals defined by solving equations like (11) for suitable asymptotically pivotal statistics, and with critical values defined by a bootstrap distribution, can provide coverage much closer to nominal than can conventional bootstrap percentile-t confidence intervals.

Since the equations (11) for the artificial regression (10) are in general rather complicated nonlinear equations, some sort of nonlinear algorithm is needed to solve them.

The Gauss-Newton algorithm can be used to find the unrestricted estimate $\hat{\beta}$. With some suitable starting point $\beta^{(0)}$, the algorithm uses an iterative procedure with the following updating rule at step i:

$$\beta^{(i+1)} = \beta^{(i)} + b^{(i)},$$

where $b^{(i)}$ is the OLS estimate from the GNR (10), evaluated at $\beta^{(i)}$. Once $b^{(i)}$ is sufficiently close to zero, according to some stopping rule, the algorithm has considered to have converged. The standard error $\hat{\sigma}_\beta$ is just the standard error given by (10) at convergence.

It is not hard to modify this Gauss-Newton algorithm so as to converge to one of the limits $\beta_\pm$ of the desired confidence interval. For the upper limit, a sensible starting point, because it is trivial to compute, is the upper limit $\hat{\beta} - \hat{\sigma}_\beta c_{\alpha/2}$ of the interval (2) based on Wald tests. Then, at step i of the iterative procedure, the updating rule is

$$\beta^{(i+1)} = \beta^{(i)} + (t^{(i)} - c_{\alpha/2})\sigma_\beta^{(i)}, \tag{12}$$

where $t^{(i)}$ is the t statistic from the GNR (10) with variables evaluated at $\beta^{(i)}$. When the rule (12) converges with $\beta^{(i+1)} = \beta^{(i)}$, it must be the case that $t^{(i)} = c_{\alpha/2}$, since $\sigma_\beta^{(i)} \neq 0$. It then follows that $\beta^{(i)} = \beta_+$, since, when (10) is evaluated at $\beta^{(i)}$, the t statistic $t^{(i)}$ is the LM test for the hypothesis that $\beta = \beta^{(i)}$.

The rule (12) can be seen to be reasonable, in the sense that it can be expected to converge faster than other rules that converge to a condition in which $t^{(i)} = c_{\alpha/2}$, by considering the special case of a linear regression, for which (10) becomes

$$\boldsymbol{y} - \boldsymbol{x}\beta = \boldsymbol{x}b + \text{residuals}.$$

Note that the standard error $\hat{\sigma}_\beta$ from this regression is always the same, whatever the value of β at which the regressand is evaluated. The t statistic is the estimate $\hat{b}$ divided by $\hat{\sigma}_\beta$, and $\hat{b}$ is $\hat{\beta} - \beta$, where $\hat{\beta}$ is the OLS parameter estimator. Let $\beta^{(0)}$ be the starting point of the algorithm. Then rule (12) gives

$$\begin{aligned}
\beta^{(1)} &= \beta^{(0)} + (\hat{b}/\hat{\sigma}_\beta - c_{\alpha/2})\hat{\sigma}_\beta \\
&= \beta^{(0)} + \hat{b} - \hat{\sigma}_\beta c_{\alpha/2} \\
&= \beta^{(0)} + \hat{\beta} - \beta^{(0)} - \hat{\sigma}_\beta c_{\alpha/2} = \hat{\beta} - \hat{\sigma}_\beta c_{\alpha/2}.
\end{aligned}$$

In the linear case, the confidence interval corresponding to the family of LM tests is the same as that corresponding to the family of Wald tests, and so we see that the upper limit β_+ is reached in one single iteration from an arbitrary starting point. For the lower limit, the obvious starting point is the lower limit of the Wald interval, and, in the updating rule (12), $c_{\alpha/2}$ is replaced by $c_{1-\alpha/2}$. A convenient convergence criterion (in nonlinear cases) is just $t^{(i)} - c_{\alpha/2}$, or the analogous expression for the

lower limit. When the absolute value of this expression is smaller than some suitably small value, like 10^{-6}, the algorithm is terminated.

It turns out to be quite easy to extend the algorithm to cases in which there are other parameters than the one for which a confidence interval is desired. This is the case treated by Hu and Kalbfleisch (2000) under the heading of "Nuisance Parameters." Consider again the GNR (9) for the case in which β_2 is a scalar parameter. With suitable notational adjustments, (9) becomes

$$\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta}_1, \beta_2) = \boldsymbol{X}_1(\boldsymbol{\beta}_1, \beta_2)\boldsymbol{b}_1 + \boldsymbol{x}_2(\boldsymbol{\beta}_1, \beta_2)b_2 + \text{residuals}. \tag{13}$$

In order to obtain an LM test for the hypothesis $\beta_2 = \beta_2^0$, the variables of (13) should be evaluated at $\beta_2 = \beta_2^0$ and $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1(\beta_2^0)$, where this last expression denotes the restricted estimates of $\boldsymbol{\beta}_1$ under the assumption that $\beta_2 = \beta_2^0$. It is precisely the need to evaluate $\boldsymbol{\beta}_1(\beta_2)$ for each candidate value of β_2 that makes it harder to construct a confidence interval based on a family of tests other than Wald tests. The virtue of the artificial regression methodology is that it provides a single iterative procedure such that, at convergence, β_2 is equal to one of the limits of the desired confidence interval, and $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1(\beta_2)$.

The parameters $\boldsymbol{\beta}_1$ and β_2 are first estimated, using the GNR or otherwise, so as to obtain estimates $\hat{\boldsymbol{\beta}}_1$ and $\hat{\beta}_2$, along with the standard error $\hat{\sigma}_2$ of $\hat{\beta}_2$. The starting point for the algorithm is then $\boldsymbol{\beta}_1^{(0)} = \hat{\boldsymbol{\beta}}_1$, and $\beta_2^{(0)} = \hat{\beta}_2 - \hat{\sigma}_2 c_{\alpha/2}$ for the upper limit, with an analogous expression for the lower limit. Then, at step i of the iterative procedure, we update as follows. The GNR (13) is modified so that the last regressor is replaced by its orthogonal projection on to the orthogonal complement of the space spanned by the other regressors. This gives

$$\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta}_1^{(i)}, \beta_2^{(i)}) = \boldsymbol{X}_1(\boldsymbol{\beta}_1^{(i)}, \beta_2^{(i)})\boldsymbol{b}_1 + \boldsymbol{M}_1^{(i)}\boldsymbol{x}_2(\boldsymbol{\beta}_1^{(i)}, \beta_2^{(i)})b_2 + \text{residuals}, \tag{14}$$

where $\boldsymbol{M}_1^{(i)}$ projects on to the orthogonal complement of the columns of the matrix $\boldsymbol{X}_1(\boldsymbol{\beta}_1^{(i)}, \beta_2^{(i)})$. The easiest way to perform the modification is to evaluate the variables of (13) as usual, then to regress the last regressor on the others, and to use the residuals from this auxiliary regression as the last regressor in (14). After running (14) and obtaining the OLS estimates $\boldsymbol{b}_1^{(i)}$ and $b_2^{(i)}$, updating follows the rule

$$\begin{aligned} \boldsymbol{\beta}_1^{(i+1)} &= \boldsymbol{\beta}_1^{(i)} + \boldsymbol{b}_1^{(i)}, \quad \text{and} \\ \beta_2^{(i+1)} &= \beta_2^{(i)} + (t_2^{(i)} - c_{\alpha/2})\sigma_2^{(i)}, \end{aligned} \tag{15}$$

where $t_2^{(i)}$ and $\sigma_2^{(i)}$ are respectively the t statistic and standard error for the last regressor in (14). A convergence criterion can be based on the absolute value of $t_2^{(i)} - c_{\alpha/2}$ and the absolute values of the other t statistics in (14); either the sum of these absolute values or the sum of their squares would be appropriate.

When the procedure defined by (15) converges, the LM test statistic t_2 is equal to the critical value $c_{\alpha/2}$, as required. In addition, $\boldsymbol{b}_1 = \boldsymbol{0}$. Since the last regressor

in (14) is orthogonal to the others, the estimate of $\boldsymbol{b}_1$ is the same as the one we would obtain if the last regressor were to be omitted from the regression. If that estimate is zero, then this implies that the regressors $\boldsymbol{X}_1(\boldsymbol{\beta}_1, \beta_2)$ are orthogonal to the regressand $\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta}_1, \beta_2)$. These orthogonality conditions are exactly the first-order conditions defining the restricted estimates of $\boldsymbol{\beta}_1$. At convergence, therefore, we have $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1(\beta_2)$, again as required. Note that this result does not hold unless the last regressor in (14) is orthogonal to the others.

The procedure for basing a confidence interval on a family of LR tests is, for the case of a nonlinear regression, only slightly different from that for LM tests. If we denote by $\mathrm{SSR}(\boldsymbol{\beta}_1, \beta_2)$ the sum of squared residuals from the model (8), the asymptotically pivotal function used by LR tests is

$$\tau(\beta_2) \equiv n \log\Big(\frac{\mathrm{SSR}(\boldsymbol{\beta}_1(\beta_2), \beta_2)}{\mathrm{SSR}(\hat{\boldsymbol{\beta}}_1, \hat{\beta}_2)}\Big). \tag{16}$$

Since the nominal distribution of this statistic is $\chi^2(1)$ rather than standard normal, the equation defining the limits of the confidence interval is slightly different. It is

$$\tau(\beta_2) = c_{1-\alpha}, \tag{17}$$

where $c_{1-\alpha}$ is now the $1-\alpha$ quantile of the $\chi^2(1)$ distribution. The equation (17) now has two solutions, one the lower, the other the upper, bound of the interval. Using the explicit form of (16), (17) becomes

$$\mathrm{SSR}\big(\boldsymbol{\beta}_1(\beta_2), \beta_2\big) = \mathrm{SSR}(\hat{\boldsymbol{\beta}}_1, \hat{\beta}_2) \exp(c_{1-\alpha}/n).$$

In the notation of (14), the derivative of the left-hand side of this equation with respect to β_2 is

$$-2\boldsymbol{x}_2{}^{\top}(\boldsymbol{\beta}_1(\beta_2), \beta_2)\boldsymbol{M}_1\big(\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta}_1(\beta_2), \beta_2)\big),$$

and so the updating rule for β_2 in (15) can be replaced by an ordinary Newton step:

$$\beta_2^{(i+1)} = \beta_2^{(i)} + \frac{\mathrm{SSR}\big(\boldsymbol{\beta}_1(\beta_2^{(i)}), \beta_2^{(i)}\big) - \mathrm{SSR}(\hat{\boldsymbol{\beta}}_1, \hat{\beta}_2) \exp(c_{1-\alpha}/n)}{2\boldsymbol{x}_2{}^{\top}(\boldsymbol{\beta}_1(\beta_2^{(i)}), \beta_2^{(i)})\boldsymbol{M}_1^{(i)}\big(\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta}_1(\beta_2^{(i)}), \beta_2^{(i)})\big)},$$

while that for $\boldsymbol{\beta}_1$ remains unchanged. Note that $\mathrm{SSR}(\boldsymbol{\beta}_1(\beta_2^{(i)}), \beta_2^{(i)})$ is just the sum of squared residuals from the modified GNR (14) evaluated at $(\boldsymbol{\beta}_1(\beta_2^{(i)}), \beta_2^{(i)})$, that is, the GNR used at step i of the iterative procedure.

Although the example we have treated in detail in this section is the nonlinear regression model, the same algorithms can be applied to any model for which an artificial regression obeying the three conditions set out at the beginning of this section exists. For LM tests, the algorithm can be applied unaltered; for LR tests, the additional necessary information is the loglikelihood function and its derivative with respect to the parameter of interest.

4. The Bootstrap

The most straightforward application of the methodology developed in the preceding section to bootstrap confidence intervals is to the estimating function bootstrap of HK. They assume that the estimates of the k–vector $\boldsymbol{\theta}$ of parameters of a given model are defined by a set of *estimating equations*, which can be written as

$$\boldsymbol{s}(\boldsymbol{\theta}) = \mathbf{0}, \tag{18}$$

where the k–vector $\boldsymbol{s}(\boldsymbol{\theta})$ of estimating functions depends implicitly on the data. The estimator $\hat{\boldsymbol{\theta}}$ is defined as the solution of the equations (18). Suppose that $\boldsymbol{\theta}$ is partitioned as $\boldsymbol{\theta} = [\boldsymbol{\theta}_1 \vdots \boldsymbol{\theta}_2]$, as in the statement of Theorem 1. Then it is further supposed that $\boldsymbol{s}(\boldsymbol{\theta})$ can be partitioned as $\boldsymbol{s}(\boldsymbol{\theta}) = [\boldsymbol{s}_1(\boldsymbol{\theta}) \vdots \boldsymbol{s}_2(\boldsymbol{\theta})]$, conformably with the partition of $\boldsymbol{\theta}$. For fixed $\boldsymbol{\theta}_2$, the restricted estimates of $\boldsymbol{\theta}_1$ are defined by the estimating equations $\boldsymbol{s}_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{0}$, of which the solution can be written as $\tilde{\boldsymbol{\theta}}_1 = \boldsymbol{\theta}_1(\boldsymbol{\theta}_2)$. Then the vector $\boldsymbol{s}_2\big(\boldsymbol{\theta}_1(\boldsymbol{\theta}_2), \boldsymbol{\theta}_2\big)$ of the remaining estimating functions is, under very mild regularity conditions, of expectation zero if evaluated at the true $\boldsymbol{\theta}_2$, and can be combined with a suitable covariance matrix estimate to provide an asymptotically pivotal statistic with an asymptotic χ^2 distribution. If $\boldsymbol{\theta}_2$ is scalar, then an asymptotically standard normal asymptotic pivot can be constructed in the usual manner.

HK give general instructions as to how to obtain bootstrap estimates of the distribution of such asymptotic pivots. It is then possible to construct a confidence region for a vector $\boldsymbol{\theta}_2$, or a confidence interval for a scalar θ_2, by means of an equation setting the asymptotic pivot to the appropriate critical value of the bootstrap distribution. They show that the coverage error of such regions or intervals is $O(n^{-3/2})$ as the sample size n tends to infinity, this being the usual order of magnitude of the error made by the bootstrap when used for a two-tailed test with an asymptotic pivot.

If the estimating functions are the components of the score vector for a model estimated by maximum likelihood, then it is clear that HK's pivot is just an LM test statistic. For other estimators, the pivot is a test statistic analogous to the classical LM statistic; in the rest of the paper, we will not distinguish between classical LM statistics and their analogues in contexts other than that of maximum likelihood.

HK recommend a particular way of obtaining the bootstrap distribution of the statistic. They assume that the components of the estimating functions $\boldsymbol{s}(\boldsymbol{\theta})$ are sums of independent contributions from the observations contained in the sample: $\boldsymbol{s}(\boldsymbol{\theta}) = n^{-1/2} \sum_{t=1}^{n} \boldsymbol{g}_t(\boldsymbol{\theta})$. Let $\boldsymbol{z}_t \equiv \boldsymbol{g}_t(\hat{\boldsymbol{\theta}})$. Then a bootstrap sample $(\boldsymbol{z}_1^*, \dots, \boldsymbol{z}_n^*)$ is obtained by resampling the $\boldsymbol{z}_t$. The bootstrap distribution is the empirical distribution of the quantities τ^* defined as

$$\tau^* = (\boldsymbol{s}^*)^\top (\boldsymbol{U}^*)^- \boldsymbol{s}^*,$$

where $\boldsymbol{s}^* \equiv n^{-1/2} \sum_{t=1}^{n} \boldsymbol{z}_t^*$ and $(\boldsymbol{U}^*)^-$ is the Moore-Penrose generalised inverse of a covariance matrix based on the bootstrap covariance matrix

$$\boldsymbol{V}^* = n^{-1} \sum_{t=1}^{n} (\boldsymbol{z}_t^* - \bar{\boldsymbol{z}}^*)(\boldsymbol{z}_t^* - \bar{\boldsymbol{z}}^*)^\top,$$

with $\bar{\boldsymbol{z}}^* \equiv n^{-1} \sum_{t=1}^n \boldsymbol{z}_t^*$; see HK for details. In some circumstances, HK suggest that it may be desirable for reasons of efficiency to use a different definition of $\boldsymbol{V}^*$.

If HK's recommendation regarding the bootstrap distribution is followed, the method of the preceding section for constructing confidence intervals can be applied to it unchanged. For the case of a scalar θ_2, the critical values $c_{\alpha/2}$ and $c_{1-\alpha/2}$ are quantiles of the bootstrap distribution of an asymptotically standard normal statistic rather than those of the standard normal or t distributions. These need to be evaluated only once, and, that done, the iterative procedure can be used as described, since the statistic is just an LM statistic. The possibility of determining the critical values of the bootstrap distribution just once, without the need to perform a possibly costly nonlinear estimation for each bootstrap sample, is one of the great virtues of HK's estimating function bootstrap.

However, although the confidence interval does indeed correspond to a family of tests each with its own null hypothesis, the *bootstrap DGP*, that is, the data-generating process for the bootstrap statistics, is the same for each null. At first glance, this seems to violate the principle by which the bootstrap DGP must satisfy the null hypothesis, but since the bootstrap estimating functions $\boldsymbol{s}^*$ are computed by resampling from a set which by construction has mean zero, the violation is only apparent.

Despite this, it is in general desirable for the bootstrap DGP, to the extent that it depends on parameters, to be based on parameter estimates computed under the null hypothesis that is to be tested. Davidson and MacKinnon (1999a) show that this can in many cases lead to an additional bootstrap refinement. Implementing this suggestion may however be computationally costly. To see why, observe that it is an analogous requirement that makes it more computationally demanding to use a family of LM tests than a family of Wald tests to construct a confidence interval. For each limit of the interval, it is necessary to consider a statistic that depends on the *restricted* estimates of the other parameters; these depend on what restriction is to be imposed, and so are different, not only for the two limits of the confidence interval, but also for any candidates for the limit that an algorithm may consider. If we want our bootstrap DGPs to use restricted estimates, this seems to imply that an iterative procedure for finding the limits of a bootstrap confidence interval needs to employ at least as many different bootstrap DGPs as the number of iterations, something that could be an exceedingly costly undertaking in terms of computing time.

However that may be, it is clear that, if we use different bootstrap DGPs for the different limits, and also along the path we take in order to find these limits, then there is no single bootstrap distribution, as there is with HK, of which we can use the quantiles as critical values. Instead, the appropriate concept is the *bootstrap P value*, which, in the construction of a bootstrap confidence interval, is used instead of the asymptotic pivot $\tau(\theta)$ used for an asymptotic interval. Indeed, if θ_0 is the true value of the parameter, the bootstrap P value for the hypothesis that $\theta = \theta_0$ is asymptotically uniformly distributed on the interval $[0, 1]$, and is therefore asymptotically pivotal.

If we denote this bootstrap P value as $P^*(\theta_0)$, the equation defining the limits of the confidence interval is

$$P^*(\theta_\pm) = \alpha, \tag{19}$$

for an interval of nominal coverage $1 - \alpha$. As with the LR test, the equation (19) will have two solutions.

We can now describe conceptually an algorithm for constructing a confidence interval based on a family of bootstrap tests for which the bootstrap DGP is based on restricted estimates. Suppose that we plan to use an artificial regression that we write as

$$\boldsymbol{r}(\boldsymbol{\theta}_1, \theta_2) = \boldsymbol{R}_1(\boldsymbol{\theta}_1, \theta_2)\boldsymbol{b}_1 + \boldsymbol{r}_2(\boldsymbol{\theta}_1, \theta_2)b_2 + \text{residuals}; \tag{20}$$

this is a partitioned version of (6), in which we separate out the parameter θ_2 for which we wish to construct a confidence interval. This artificial regression must correspond to the estimator $\hat{\boldsymbol{\theta}}$ defined by the estimating equations (18). We intend to base the bootstrap tests on the LM statistics (7) generated by this artificial regression evaluated at $(\boldsymbol{\theta}_1(\theta_2), \theta_2)$ for each hypothesised value of θ_2. First, suitable starting values of $\boldsymbol{\theta}_1$ and θ_2 can be found exactly as for an asymptotic interval. Then, at step i of the subsequent iterative procedure, we update as follows.

(i) Run the artificial regression (20) evaluated at $\boldsymbol{\theta}^{(i)} \equiv (\boldsymbol{\theta}_1^{(i)}, \theta_2^{(i)})$, obtaining the estimates $\boldsymbol{b}^{(i)}$ and the t statistic, $\tau(\theta_2^{(i)})$ say, corresponding to the regressor $\boldsymbol{r}_2(\boldsymbol{\theta}^{(i)})$, which would be the LM statistic if $\boldsymbol{\theta}_1^{(i)}$ were equal to $\boldsymbol{\theta}_1(\theta_2^{(i)})$.

(ii) Update $\boldsymbol{\theta}_1$ by the formula

$$\boldsymbol{\theta}_1^{(i+1)} = \boldsymbol{\theta}_1^{(i)} + \boldsymbol{b}_1^{(i)}.$$

(iii) Use a bootstrap DGP for which the parameters take on the values $\boldsymbol{\theta}_1^{(i+1)}$ and $\theta_2^{(i)}$ to generate B bootstrap samples, for each of which we compute the bootstrap LM statistic τ_j^*, $j = 1, \ldots, B$.

(iv) Compute the bootstrap P value $P^*(\theta_2^{(i)})$ as the proportion of the τ_j^* that are greater than $\tau(\theta_2^{(i)})$.

(v) Define $\theta_2^{(i+1)}$ in some suitable manner so as to make $P^*(\theta_2^{(i+1)})$ closer to α than was $P^*(\theta_2^{(i)})$.

This conceptual layout is lacking essential details, and can be expected to be computationally very inefficient. However, it is clear that, if it can be implemented in some way, then at convergence θ_2 satisfies the defining equation (19) for a limit of the confidence interval. The next section is devoted to supplying the missing details and improving efficiency.

5. Details of a More Efficient Algorithm

Steps (i) and (ii) of the conceptual algorithm above involve no difficulty, and entail only trivial computational cost. The same is by no means true of step (iii). We spend no time discussing what *sort* of bootstrap DGP to use, because the choice should be specific to the model. It may be parametric or semiparametric, that is, using resampling, and it may impose much or little structure. The important point is just that its true parameter values should be $(\boldsymbol{\theta}_1^{(i+1)}, \theta_2^{(i)})$. The use of $\boldsymbol{\theta}_1^{(i+1)}$ rather than $\boldsymbol{\theta}_1^{(i)}$ is motivated by two considerations: First, updating $\boldsymbol{\theta}_1$ is easy, and, second, doing so makes the bootstrap DGP closer to using the genuine restricted estimates of $\boldsymbol{\theta}_1$ for the value of $\theta_2^{(i)}$. Formally, this point is just an implementation detail, rather than a point of principle.

It is in general costly to compute the actual LM statistic for each of the bootstrap samples, since it is necessary to compute the restricted estimates of $\boldsymbol{\theta}_1$ for each of them. However, it has been shown by Davidson and MacKinnon (1999b) that the artificial regression can be used to economise significantly on computing time by performing only a fixed number of iterations in the direction of the restricted estimates. See also Andrews (2001) for an exposition of the theoretical underpinnings of the method. The trick is that the fixed number of iterations use as starting point the known true parameter values of the bootstrap DGP. The method can be used, not only for bootstrapping an LM statistic, but also for an LR statistic.

A crucial point at this stage is that the *same* random numbers must be used in successive iterations when generating the bootstrap samples. Unless the number B of bootstrap samples is truly immense, simulation error will be enough to swamp the information necessary for the overall procedure to converge. One possible approach is to generate all the needed random numbers once and for all and to save them for subsequent use; another is simply to save the random number seed at the beginning, and to reset it to the same value for each iteration. On modern computers, the second approach is probably preferable, since it makes virtually no use of computer memory, and the time needed to generate random numbers is not much greater than the time needed to fetch them from memory.

Step (iv) once again appears straightforward and almost free of cost. But it must be noted that the resulting P values take on only a discrete set of values, namely i/B, $i = 0, 1, \ldots, B$. It is well known that, if a nominal level of α is of particular interest, size distortion of a bootstrap test is minimised if B is chosen so that $\alpha(B+1)$ is an integer; see for instance Davison and Hinkley (1997). If such a choice is made, no P value can be exactly equal to α. Rather, as θ_2 varies in the direction from the interior to the exterior of the confidence interval, there is a discontinous jump from a value greater than α to a value less than α. The value of θ_2 for which the jump occurs is the limit of the confidence interval.

It is not possible to use Newton's method, or indeed any of the usual methods of finding zeros of functions, when the function is discontinuous. It is therefore desirable

to perform a linear interpolation. For given θ_2, let τ_-^* be the greatest bootstrap statistic that is less than the actual statistic $\tau(\theta_2)$, and let τ_+^* be the smallest one that is greater than or equal to τ. Further, let there be exactly m bootstrap statistics greater than or equal to τ. Then we replace the integer m by the real number

$$m + \frac{\tau_+^* - \tau(\theta_2)}{\tau_+^* - \tau_-^*}, \tag{21}$$

which varies continuously with θ_2, since $\tau(\theta_2)$ does so, since, for constant m, τ_-^* and τ_+^* do so also, and since, for values of θ_2 for which m jumps by an integer, the fractional part of (21) jumps by an integer in the other direction. The limit of the confidence interval in then the locally unique θ_2 for which (21) is equal to the integer $\alpha(B+1)$.

Step (v) is left completely unspecified in the layout of the preceding section. In principle, Newton's method could be used, or some approximation to it, in order to find θ_2 such that the expression (21) is equal to $\alpha(B+1)$. The derivative of (21) with respect to $\tau(\theta_2)$ is well approximated by $-1/(\tau_+^* - \tau_-^*)$, and the derivative of $\tau(\theta_2)$ with respect to θ_2 is easy to approximate using the data provided by the artificial regression. However, attempts to implement this method proved very numerically unstable, on account of the discontinuities of the derivative of (21) at points at which m jumps.

There exist many algorithms for finding the zeros of scalar functions of a scalar argument, often called *root-finding* algorithms,and most of them perform very well when the function is monotonic in the neighbourhood of the zero. In all but the most pathological cases, (21) will be a monotonic function of θ_2, decreasing as θ_2 moves towards the exterior of the confidence interval. A number of available root-finding methods were tried for the implementation of step (v), and one in particular, Brent's method, for which see Press, Teukolsky, Vetterling, and Flannery (1992), section 9.3, was particularly effective in our simulation experiments. The method combines the method of bisection with higher-order interpolation methods in an attempt to minimise the number of function evaluations needed in the search for the zero.

When a root-finding algorithm like Brent's method is used, step (v) becomes an iteration of that method. Such methods normally suppose that the function of which the zero is sought can be evaluated exactly at trial values of the argument. That is not the case here, because, at each iteration of the overall algorithm, $\boldsymbol{\theta}_1$ is updated as well as θ_2. Neither Brent's method nor any other of the methods we tried seemed to be affected by this point. In addition, convergence of the overall algorithm seems to be governed by convergence of the root-finding method. In principle, it is necessary to be sure that the updates of $\boldsymbol{\theta}_1$ have converged as well as the updates of θ_2, but, in practice, it was found that the convergence of θ_2 was slower than that of $\boldsymbol{\theta}_1$, so that convergence of θ_2 is enough for overall convergence. Since this result is by no means a mathematical theorem, it is no doubt prudent in practice to check convergence of $\boldsymbol{\theta}_1$ once θ_2 has converged. This is easy to do using the artificial regression.

We conclude this section by respecifying steps (iii)-(v) of the procedure of the preceding section.

(iii) Use a bootstrap DGP for which the parameters take on the values $\boldsymbol{\theta}_1^{(i+1)}$ and $\theta_2^{(i)}$ to generate B bootstrap samples, using the same random numbers for each iteration of the overall procedure. For each bootstrap sample, perform a fixed number ($2, 3, 4$, in typical cases) of iterations of the artificial regression (20), starting from $(\boldsymbol{\theta}_1^{(i+1)}, \theta_2^{(i)})$. The bootstrap statistic τ_j^*, $j = 1, \ldots, B$, is typically the absolute value of the t statistic associated with the regressor $\boldsymbol{r}_2$ on the last artificial regression iteration.

(iv) Compute the expression (21) that represents a smoothed version of the number of bootstrap statistics greater than $\tau(\theta_2^{(i)})$.

(v) Do one iteration of Brent's method interpreted as a method for finding the zero of the function of θ_2 given by (21) minus $\alpha(B + 1)$.

Many root-finding methods, Brent's in particular, essentially need two starting points, in order to bracket the zero. If the starting point turns out to be inside the confidence interval, then the other starting point should be far enough away from the first to be outside, and *vice versa*. This requirement is easy to satisfy in practice.

At this point, it is reasonable to ask how much longer it takes to compute a confidence interval of the type we are considering compared with a conventional percentile-t interval or an interval of the sort recommended by HK based on the estimating function bootstrap. A second question is then whether the extra computing time is worth it in terms of improved coverage accuracy. An answer to the second question is given in one particularly simple case by Davidson (2000), who considers the coverage accuracy of a confidence interval for the autoregressive parameter ρ in the simple autoregression

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

based on a family of parametric bootstrap tests where the bootstrap DGP is defined using restricted estimates. Accuracy is seen to be substantially superior to that of a percentile-t interval; indeed, even with $100,000$ simulations, it was impossible to reject a null of perfect coverage. In the next section, we investigate computing time and accuracy in the context of a nonlinear regression.

6. Illustrative Example

Simulation experiments were carried out in the context of the regression model

$$y_t = \alpha + \delta x_{t1}/\gamma + \gamma x_{t2} + u_t, \quad u_t \sim \text{IID}(0, \sigma^2). \tag{22}$$

This model appears to be nonlinear, but it can be reparametrised to become linear, so that the parameters α, δ, and γ can be estimated by OLS. It is desired to construct a confidence interval for δ, and it is clear that, if the value of δ is fixed, the regression

function is genuinely nonlinear with respect to α and γ. If $\delta = 1$, the model (22) is the model considered by Gregory and Veall (1985), known to yield particularly badly behaved Wald statistics for large or small values of γ.

The GNR that corresponds to (22) can be written as

$$y_t - \alpha - \delta x_{t1}/\gamma - \gamma x_{t2} = b_\alpha + b_\gamma (x_{t2} - \delta x_{t1})/\gamma^2 + b_\delta x_{t1}/\gamma + \text{residual}. \tag{23}$$

We use this GNR, and the restricted version of it in which the last regressor is absent, as the artificial regression for constructing confidence intervals.

We consider three types of bootstrap confidence interval. The first is the percentile-t interval, for which (22) is estimated unrestricted (by OLS), and an asymptotic pivot

$$\tau_t(\delta) = \frac{\hat{\delta} - \delta}{\hat{\sigma}_\delta}$$

of the form (3), that is, an asymptotic t statistic, bootstrapped using the bootstrap DGP

$$y_t^* = \hat{\alpha} + \hat{\delta} x_{t1}/\hat{\gamma} + \hat{\gamma} x_{t2} + u_t^*, \tag{24}$$

where $\hat{\alpha}$, $\hat{\delta}$, and $\hat{\gamma}$ are the OLS estimates from (22), and where the u_t^* are resampled from the residuals of (22), multiplied by $(n/(n-3))^{1/2}$, a degrees-of-freedom correction. The regressors x_{t1} and x_{t2}, treated as exogenous, are used unchanged in the bootstrap DGP. The bootstrap statistics τ_t^* are computed as

$$\tau_t^* = \frac{\delta^* - \hat{\delta}}{\sigma_\delta^*},$$

where δ^* and σ_δ^* are respectively the OLS estimate of δ and its standard error, from (24).

The second type of confidence interval we consider is the HK interval for models with IID errors. The asymptotic pivot, $\tau_{\text{HK}}(\delta)$, is the t statistic on the last regressor of the GNR (23), with variables evaluated at δ, and $\alpha(\delta)$ and $\gamma(\delta)$, where these are restricted parameter estimates with fixed δ. The bootstrap DGP does not generate samples y_t^*, but just bootstrap statistics according to the formula

$$\tau_{\text{HK}}^* = \frac{1}{\sigma^*} \boldsymbol{a}^\top \boldsymbol{u}^*,$$

where $\boldsymbol{u}^*$ is a vector of residuals resampled as for the percentile-t method, and $\boldsymbol{a}$ is a vector obtained by projecting the last regressor of the GNR (23), with variables evaluated at $(\hat{\alpha}, \hat{\delta}, \hat{\gamma})$, on to the orthogonal complement of first two regressors, and then standardising the result to have unit Euclidean length. The estimate σ^* is defined as the sample variance of the components of $\boldsymbol{u}^*$.

Finally, the third type of confidence interval is of the sort discussed in the preceding section. The GNR (23) is used as the artificial regression, and Brent's method is used to update the parameter δ. The asymptotic pivot is the same as for the HK method, and the bootstrap DGP for given δ is

$$y_t^* = \alpha(\delta) + \delta x_{t1}/\gamma(\delta) + \gamma(\delta)x_{t2} + u_t^*,$$

with the bootstrap residuals u_t^* defined as for the other two methods.

7. Concluding Remarks

The algorithms developed in this paper can be applied more generally than we have done here. Sometimes one-sided confidence intervals are of interest, based on a family of one-tailed tests. There is of course no difficulty in applying our algorithms to such intervals. If a confidence set is desired for more than one parameter, then things are more complicated. HK maintain that their method can be used with little difficulty, but it is still necessary to solve an equation, not for a point, but for a contour or higher-dimensional manifold. We know of no computationally efficient way of doing this, and have therefore limited ourselves in this paper to confidence intervals.

There is plenty of scope for incorporating techniques of variance reduction into the algorithms of this paper. Control variates and importance functions have been suggested for reducing simulation error in bootstrapping. It seems likely that artificial regressions can incorporate these techniques. A method of improving the reliability of bootstrap tests is proposed in Davidson and MacKinnon (2000), involving the computation of two statistics for each bootstrap sample. This too can easily be incorporated into the artificial regression methodology.

We hope to have shown in this paper that bootstrap confidence intervals can benefit from the numerous refinements that have been shown to exist for bootstrap tests, but are seldom exploited for confidence intervals. Bootstrapping with restricted estimates is a prime example of a technique that has been shown to improve the accuracy of bootstrap P values, and that can now be made use of for bootstrap confidence intervals of improved coverage accuracy.

References

Anderson, T. W., and H. Rubin (1949). "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Mathematical Statistics*, **20**, 46–63.

Andrews, D. W. K. (2001). "Higher-order Improvements of a Computationally Attractive $k-$step Bootstrap for Extremum Estimators," *Econometrica*, forthcoming.

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

Davidson, R. (2000). "Comment on *Recent Developments in Bootstrapping Time Series*, by Berkowitz and Kilian," *Econometric Reviews*, **19**, 49-54.

Davidson, R., and J. G. MacKinnon (1990). "Specification tests based on artificial regressions," *Journal of the American Statistical Association*, **85**, 220–27.

Davidson, R., and J. G. MacKinnon (1991). "Artificial Regressions and $C(\alpha)$ Tests," *Economics Letters*, **35**, 149–153.

Davidson, R., and J. G. MacKinnon (1999a). "The Size Distortion of Bootstrap Tests," *Econometric Theory*, **15**, 361–376.

Davidson, R., and J. G. MacKinnon (1999b). "Bootstrap testing in Nonlinear Models," *International Economic Review*, **40**, 487–508.

Davidson, R., and J. G. MacKinnon (2000). "Improving the Reliability of Bootstrap Tests," Queen's Institute for Economic Research, DP #995.

Davidson, R., and J. G. MacKinnon (2001). "Artificial Regressions," in *A Companion to Theoretical Econometrics*, ed. B. H. Baltagi, Blackwell, Oxford, pp 16–37.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and Their Application*, Cambridge, Cambridge University Press.

Dufour, J.-M. (1997). "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, **65**, 1365–1387.

Gregory, A. W., and M. R. Veall (1985). "On formulating Wald tests for nonlinear restrictions," *Econometrica*, **53**, 1465–68.

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

Hu, F., and J. D. Kalbfleisch (2000). "The estimating function bootstrap," *Canadian Journal of Statistics*, **28**, 449–499.

Lafontaine, F., and K. J. White (1986). "Obtaining any Wald statistic you want," *Economics Letters*, **21**, 35–40.

Orme, C. (1995). "On the use of artificial regressions in certain microeconometric models," *Econometric Theory*, **11**, 290–305.

Phillips, P. C. B., and J. Y. Park (1988). "On the formulation of Wald tests of nonlinear restrictions," *Econometrica*, **56**, 1065–83.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1992). *Numerical Recipes in C*, Cambridge University Press, Cambridge.

Smith, R. J. (1987). "Alternative asymptotically optimal tests and their application to dynamic specification," *Review of Economic Studies*, **54**, 665–80.