Computational Economics **12:** 79–95, 1998.

Implementing the Double Bootstrap

B.D. McCULLOUGH
Federal Communications Commission, 2000 M St. NW Room 533, Washington, DC 20554, USA, Internet: bmccullo@fcc.gov

H.D. VINOD
Economics Department, Fordham University NY, NY, 10458-5158, USA, Internet: vinod@murray.fordham.edu

Abstract. The single bootstrap already is popular in economics, though the double bootstrap has better convergence properties. We discuss the theory and implementation of the double bootstrap, both with and without the pivotal transformation, and give detailed examples of each. One example is a nonlinear double bootstrap of a Cobb-Douglas production function, and explains the use of Gauss-Newton Regressions as a device to decrease computational time. Another example is double bootstrapping elasticities from a translog production function.

Key words: Cobb-Douglas, confidence intervals, nonlinear regression, translog

1. Introduction

In economics, perhaps the most widespread example of the substitution of computational power for analytic theory has been the use of the single bootstrap, which requires the economist to run hundreds of regressions. Recent advances in statistical theory and ever-faster computers have opened up a still more powerful computational substitute for analytic theory: the double bootstrap, which requires the economist to run hundreds of thousands of regressions. The purpose of the present paper is to help add the double bootstrap to the economist's toolkit.

Efron's (1982) bootstrap has been well-explored in several statistical texts: Efron and Tibshirani (1993); Hall (1992); LePage and Billard (1992); Mammen (1992); Shao and Tu (1995); and Davison and Hinkley (1997). The bootstrap possesses many attractive features when compared to classical asymptotic normal theory. It is applicable under milder conditions and can handle problems which classical methods cannot solve. Indeed, the bootstrap is preferable to classical asymptotics even when the classical assumptions are valid since the bootstrap is 'always at least as good as, and in some cases better than, the classical normal approximation' (Navidi, 1989, p. 1472). In economics the single bootstrap has been fruitfully applied to a wide variety of problems, including: Stein-Rule estimation (Brownstone, 1990); confidence intervals for translog estimates of elasticities (Eakin, McMillen and Buono, 1990) and supply and demand elasticities (Vinod and McCullough, 1994); extreme bounds analysis (McAleer and Veall, 1989); ARIMA identification (Aczel and Josephy, 1992); AR forecasting (Thombs and Schuchany, 1990; McCullough, 1994); linear regression forecasting with fixed regressors (Stine, 1985; Breiman, 1992), stochastic regressors (McCullough, 1996), and random coefficients (Beran, 1995); and model selection (Veall, 1992). See also the survey by Vinod (1993).

Articles on double bootstrap methods regularly appear in the statistical literature (Beran, 1987, 1988, 1990; Kuk, 1989; Martin, 1990, 1992; Shi, 1992; Scholz, 1994), though we are aware of only two applications of the double bootstrap in economics. The first concerns the simple Keynesian multiplier based on the marginal propensity to consume (MPC). Vinod and McCullough (1995) estimate confidence intervals for the multiplier, 1/(1-MPC), from a cointegrating regression of consumption on income, and introduce the use of plots based on the double bootstrap as a diagnostic tool for econometric modeling. The second application concerns ridge regression. While it is well-known that ridge regression provides better slope estimates in the presence of multicollinearity, obtaining accurate confidence intervals was problematic due to the stochastic nature of the biasing parameter. Vinod (1995) uses the double bootstrap to obtain consistent confidence intervals in a ridge regression, a problem that cannot be solved by either classical methods or the single bootstrap.

Section 2 motivates the double bootstrap and presents some basic theory along with an algorithm. While the theory may appear complex, the implementation is not much more difficult than the single bootstrap, and is much more simple, from a programming perspective, than some of the more sophisticated single bootstraps involving bias-correction or acceleration. Section 3 presents our first example, double bootstrapping a test of constant returns-to-scale in a Cobb-Douglas production function with additive error. If J is the number of first-stage resamples and K is the number of second-stage resamples, the double bootstrap calls for JK nonlinear regressions, a not insubstantial demand on computing time. Davidson and MacKinnon (1997) show that this double bootstrap can be executed much more quickly and with no loss of accuracy using only a single nonlinear regression, aided by mJK linear regressions, where m is a small integer. Section 4 discusses Shi's (1992) double bootstrap, a method for double bootstrapping when no pivot is available. The usual double bootstrap is predicated on the existence of a pivoting transformation, which could be effected in the first example. In economics there are many applications when such a transformation cannot be made. Supply and demand elasticities, and substitution elasticities from a translog production function are but two such examples. In Section 5 we apply Shi's method to the problem of obtaining confidence intervals for substitution elasticities from a translog production function. We also describe an artifice for substantially reducing the amount of computer time necessary. Section 6 offers the conclusions.

2. Theory of the Double Bootstrap

The chief advantage of the double bootstrap over the single bootstrap is that the double bootstrap confidence interval typically has a higher order of accuracy. For a sample of size n, an estimator $\hat{\theta}$ of a parameter θ is said to be k-th order accurate if its convergence rate is $O(n^{-k/2})$. For example, suppose a population is non-normal so that asymptotic arguments are necessary. Let μ be the parameter of interest (say, the population mean) with standard deviation σ and let $\bar{x}$ and s be estimators thereof, calculated from the original sample. In the usual situation, $|\bar{x} - \mu| = O(1/\sqrt{n})$, so the sample mean is first-order accurate. A confidence set C_n of θ is said to be k-th order accurate if

$$|P\{\theta \in C_n\} - (1 - \alpha)| = O(n^{-k/2}) \tag{1}$$

where typically $\alpha = 0.05$. Continuing the above example, an asymptotic normal upper one-sided interval for μ, given by $(-\infty, \bar{x} + t_\alpha s]$, is first-order accurate. For large samples, $t_\alpha = 1.64$.

The *level error* of a confidence set is the difference between the nominal level, typically $(1 - \alpha)$, and the actual probability that the confidence set contains the true parameter. The double bootstrap works by adjusting the level error of a single bootstrap confidence set. We assume the reader is familiar with the single bootstrap, computational details of which are presented in McCullough and Vinod (1993). Theoretical reviews are given by DiCiccio and Romano (1988) and Manteiga, Sánchez, and Romo (1994).

Consider the *percentile bootstrap*. In the usual fashion, we can bootstrap $\hat{\theta}$ a large number of times, say $J = 999$, and obtain bootstrap estimates $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_J^*$. These bootstrap estimates may be sorted to form the empirical cdf with order statistics 25 and 975 representing lower and upper limits for a 95% confidence interval for θ. Let us call these two quantities $\hat{\theta}_L^*$ and $\hat{\theta}_U^*$. Beran (1987) has shown that the asymptotic normal and naive bootstrap one-sided intervals are both first-order accurate. The accuracy of the percentile bootstrap can be increased via a pivoting transformation.

Let $x_n = \{X_1, X_2, \ldots, X_n\}$ be a sample from a distribution with true cdf F, which has a parameter θ. Suppose $\hat{\theta}$ is a normally distributed statistic. Then the distribution of $\hat{\theta} - \theta$ is normal with mean zero and unknown variance σ^2; i.e., the distribution of $\hat{\theta} - \theta$ depends upon F through the unknown σ. By contrast, the distribution of $(\hat{\theta} - \theta)/\sigma$ is standard normal, i.e., it does not depend upon F. The idea, then, is to use $R^* = (\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$ to approximate $R = (\hat{\theta} - \theta)/\hat{\sigma}$. To this end, define

$$\hat{t}_\alpha \equiv \sup\{t : P(R^* \leq t) \leq \alpha\} \tag{2}$$

whence the one-sided upper and lower, and two-sided $100(1 - \alpha)\%$ intervals are respectively given by

$$I_1 \equiv (-\infty, \hat{\theta} - \hat{t}_\alpha \hat{\sigma}] \tag{3}$$

$$I_2 \equiv [\hat{\theta} - \hat{t}_{1-\alpha}\hat{\sigma}, \infty) \tag{4}$$
$$I_3 \equiv [\hat{\theta} - \hat{t}_{1-\alpha/2}\hat{\sigma}, \hat{\theta} - \hat{t}_{\alpha/2}\hat{\sigma}] \tag{5}$$

The various quantities $\hat{t}$ are computed from the empirical distribution of R^* and we can expect them to be better than similar quantities obtained by looking up Student's-t or standard normal tables. As a practical matter, determining the confidence interval for θ is a matter of 'unraveling' the root. On the j-th bootstrap resample define the root

$$\hat{R}_j^*(x_n, \theta^*) = (\hat{\theta}_j^* - \hat{\theta})/\hat{\sigma}_j^* \tag{6}$$

$j = 1, 2, \ldots, J$ where $\hat{\theta}$ is calculated from the original sample and $\hat{\theta}_j^*$ and $\hat{\sigma}_j^*$ are calculated from the j-th bootstrap resample. Let the ordered values, from smallest to largest, be denoted $\hat{R}^*_{(1)}, \hat{R}^*_{(2)}, \ldots, \hat{R}^*_{(J)}$. Thus the estimates of the critical values are

$$\hat{t}_\alpha = \hat{R}^*_{([(J+1)\alpha])} \tag{7}$$
$$\hat{t}_{1-\alpha} = \hat{R}^*_{([(J+1)(1-\alpha)])} \tag{8}$$
$$\hat{t}_{\alpha/2} = \hat{R}^*_{([(J+1)(\alpha/2)])} \tag{9}$$
$$\hat{t}_{1-\alpha/2} = \hat{R}^*_{([(J+1)(1-\alpha/2)])} \tag{10}$$

where $[\cdot]$ is the greatest integer function. For example, if we seek a 90% lower one-sided interval and $J = 999$ then $\hat{t}_{1-\alpha} = \hat{R}^*_{(100)}$. Similarly, a two-sided 90% interval would be $[\hat{\theta} - \hat{R}^*_{(950)}\hat{\sigma}, \hat{\theta} - \hat{R}^*_{(50)}\hat{\sigma}]$. It might seem that the interval should be $[\hat{\theta} + \hat{R}^*_{(50)}\hat{\sigma}, \hat{\theta} + \hat{R}^*_{(950)}\hat{\sigma}]$ (recall that $\hat{R}^*_{(50)} < 0$), but Hall (1988) makes it clear that such an interval is incorrect.

Continuing this line of thought, consider use of the studentizing transformation to form the studentized functional

$$R(x_n, \theta) = \frac{\hat{\theta} - \theta}{\hat{\sigma}} \tag{11}$$

where we now emphasize the dependence of R on both the true parameter value and the sample at hand, as we proceed to the double bootstrap. If the distribution $H(\cdot)$ of this R does not depend on θ then R is said to be a pivot, otherwise it is a *root* in the terminology of Beran (1987). Even if the transformation (11) does not completely eliminate the dependence of R on θ, it can decrease the dependence and yield improved confidence intervals. Henceforth, we refer to such a transformed statistic as a root, since (11) allows us not to worry whether the dependence has been completely eliminated. We use an asterisk to denote a theoretical single bootstrap construct, superscripted for statistics and subscripted for distributions and confidence sets. We add a circumflex to denote its empirical counterpart. Similarly we use two asterisks for double bootstrap quantities. The root (11) implies a one-sided upper bootstrap confidence set of the form

$$B_*(\alpha, x_n) = \{\theta \in \Theta : \hat{R}^*(x_n, \theta) < \hat{H}_*^{-1}(1-\alpha, \hat{F}_n)\} \tag{12}$$

where Θ is the range of θ, $H_*^{-1}(\alpha, F)$ is the α-quantile of H_*, and $\hat{F}_n$ is the empirical distribution of F.

Beran's (1987) key insight was to observe that if pivoting once can improve the order of accuracy of the bootstrap estimates, then pivoting twice can improve it more. In the same way that the single bootstrap can use $\hat{R}^* - \hat{R}$ to approximate $\hat{R} - R$, the double bootstrap uses $\hat{R}^{**} - \hat{R}^*$ to approximate $\hat{R}^* - \hat{R}$. The double bootstrap improves the accuracy of a single bootstrap by estimating an error, e.g., the coverage error of a confidence interval, and then uses this estimate to adjust the single bootstrap and in doing so reduce its error, thus achieving a higher order of accuracy as defined in (1). The double bootstrap achieves these goals by determining a better estimate of $\hat{t}$.

Let $T(F)$ be some parameter with range **T** and let T_n be an estimator based on n observations with standard deviation σ_n. Consider the true root $R^*(x_n, T(F)) = (T_n - T(F))/\sigma_n$, which has cdf $H_*(x, F)$ whose α-quantile is given by $H_*^{-1}(\alpha, F) = \inf\{x : H_n(x, F) \geq \alpha\}$. The empirical counterparts are $\hat{R}^*(x_n, t)$, $\hat{H}_*(x_n, \hat{F})$, and $\hat{H}_*^{-1}(\alpha, \hat{F})$. The one-sided single bootstrap confidence set for $T(F)$ is given by

$$B_*(\alpha, x_n) = \{t \in \mathbf{T} : \hat{R}^*(x_n, t) < \hat{H}_*^{-1}(1-\alpha, \hat{F}_n)\} \tag{13}$$

The confidence interval for θ is obtained by unraveling the pivot.

At the second stage a new root $R^{**}(\cdot)$ with cdf $H_{**}(\cdot, F)$ is formed by using the first-stage distribution $\hat{H}_*(\cdot)$ of $R^*(\cdot)$ as follows. We define

$$\hat{R}^{**}(x_n, T(F)) = [\hat{R}^* - R^*] = \hat{H}_*\{\hat{R}^*(x_n, T(F)), \hat{F}\} \tag{14}$$

Hence, (13) can be rewritten as

$$B_*(\alpha, x_n) = \{t \in \mathbf{T} : \hat{R}^{**}(x_n, t) < 1 - \alpha\} \tag{15}$$

which, in effect, approximates H_{**} by the the uniform distribution. This leads to a double bootstrap confidence set based on the new root R^{**}

$$B_{**}(\alpha, x_n) = \{t \in \mathbf{T} : \hat{R}^{**}(x_n, t) < H_{**}^{-1}(1-\alpha, \hat{F}_n)\} \tag{16}$$

Beran's (1987) argument is that R^{**} is closer to being pivotal than R^*, i.e., the distribution H_{**} of R^{**} is less dependent on F than is the distribution H_* of the root R^*. Indeed, by the probability integral transform, when R^* is exactly pivotal, R^{**} has cdf H_{**} which is uniform on $[0, 1]$, with no unknown parameters. If the root $R^*(\cdot)$ is not exactly pivotal but is asymptotically pivotal, then the cdf of $R^{**}(\cdot)$ is asymptotically uniform.

The probability that the true $T(F)$ falls in the confidence set $B_*(\alpha, x_n)$ is given by $H_{**}(1-\alpha, F)$ which is unknown since F is unknown. However, an estimate of this probability is $\hat{H}_{**}(1-\alpha, \hat{F}_n)$. Following Beran (1987, p. 461) we estimate $\hat{H}_{**}$ by defining a new variable Z_j as the fraction of values $\{\hat{R}^{**}_{jk}, 1 \leq k \leq K\}$ which are less than $\hat{R}^*_j$. For sufficiently large J and K the empirical distribution function of $Z_j, 1 \leq j \leq J$ approximates H_{**}. There will be J such estimates Z_j which will be uniformly distributed on $[0,1]$ if the root is a pivot having a continuous distribution and is free of the problems of bias, lattice distribution, etc. We will use this information from the second-stage bootstrap to refine the first stage intervals. The problem now is to choose α_1 so that $\hat{R}^{**}$ satisfies

$$\hat{H}_{**}(1-\alpha_1, \hat{F}_n) = 1-\alpha \tag{17}$$

where α_1 is a new nominal level needed to accomplish the original α level. In particular, we find that $\hat{t}$ quantities are now given by

$$\hat{t}_{\alpha} = \hat{R}^*_{([(J+1)Q_\alpha])} \tag{18}$$

$$\hat{t}_{1-\alpha} = \hat{R}^*_{([(J+1)Q_{1-\alpha}])} \tag{19}$$

$$\hat{t}_{\alpha/2} = \hat{R}^*_{([(J+1)Q_{\alpha/2}])} \tag{20}$$

$$\hat{t}_{1-\alpha/2} = \hat{R}^*_{([(J+1)Q_{1-\alpha/2}])} \tag{21}$$

where Q_α is the α percentile of the vector Q.

Unraveling a confidence interval for $T(F)$ from the confidence set (16) is notationally tedious. Fortunately, it is computationally facile. Some heuristics will make this critical idea clear. In the same way that pivoting a single bootstrap is used to refine the naive bootstrap, in the double bootstrap R^{**} is used to adjust R^* which, in turn, yields a confidence interval for θ. Suppose that $J = K = 999$ and we seek an upper 95% limit for θ. If the 95th quantile of $\hat{H}_{**}$ is 0.91, say, then we use the 91st quantile of $\hat{H}_*$ as the upper 95% limit for R. Of course, to the 91st quantile of $\hat{H}_*$ there corresponds some $\hat{\theta}^*_j$ which is then used for the upper 95% limit of θ. Lower limits are similarly computed.

While the theory of the double bootstrap may at times seem daunting, its implementation is actually quite easy as the following algorithm makes clear.

1. Let y be the dependent variable. Based on the original sample, calculate an estimate $\hat{\theta}$ and standard error $\sigma(\hat{\theta})$ for the parameter of interest, θ and a vector of fitted values $\hat{y}$. Form vector of residuals $e = y - \hat{y}$. Where $(n-p)$ is the number of degrees of freedom, rescale the residuals by $\sqrt{n/(n-p)}$, still denoting them by e.
2. Check to ensure that e is consistent with the true errors being i.i.d., perhaps by using the Durbin-Watson test. Autocorrelated or heteroscedastic errors can impair the bootstrap to the point of inconsistency.

3. A large number of times $j = 1, 2, \ldots, J$, 'shuffle' the residuals by randomly resampling with replacement from e to form a vector of resampled residuals e_j^*. Form a pseudo-x vector as $y_j^* = \hat{y} + e_j^*$.
4. Calculate a bootstrap estimate of the parameter of interest $\hat{\theta}_J^*$ and its standard error $\sigma(\hat{\theta}_j^*)$, and fitted values $\hat{y}_j^*$. Form the root $\hat{R}_j^* = (\hat{\theta}_j^* - \hat{\theta})/\sigma(\hat{\theta}_j^*)$.
5. For each first-stage bootstrap resample j, a large number of times $k = 1, 2, \ldots, K$ shuffle e_j^* to form e_{jk}^{**}. Form a pseudo-y vector as $y_{jk}^{**} = \hat{y}_j^* + e_{jk}^{**}$ and calculate $\hat{\theta}_{jk}^{**}$ and $\sigma(\hat{\theta}_{jk}^{**})$. Form the root $\hat{R}_{jk}^{**} = (\hat{\theta}_{jk}^{**} - \hat{\theta}_j^*)/\sigma(\hat{\theta}_{jk}^{**})$.
6. On the j-th first-stage resample, after all K second-stage operations are complete, let Z_j be the proportion of times that $\hat{R}_{jk}^{**} \leq \hat{R}_j^*$, i.e., $Z_j = \#(\hat{R}_{jk}^{**} \leq \hat{R}_j^*)/K \;\; k = 1, 2 \ldots, K$. It is this variable Z_j which will be used to adjust the first-stage intervals.
7. After all bootstrapping operations are complete, we have estimates $\hat{\theta}_j^*, \hat{R}_j^*$ and Z_j, $j = 1, 2, \ldots J$. Recall that by construction $Z_j \in [0, 1]$. To determine the upper 95% limit for θ, sort the Z_j and choose the 95-th quantile; suppose it is 0.91. The 91-st quantile of the sorted $\hat{R}_j^*$ is the adjusted upper 95% limit for the true root, R^*. Similar operations determine a lower limit.

There remains the matter of choosing J and K, the number of first- and second-stage resamples, respectively. Booth and Hall (1994) provide needed guidance on the proper choices. For given J, the accuracy of the approximation decreases when K is too small or too large. For a one-sided interval of level $(1 - \alpha)$ the asymptotic mean square error is proportional to

$$M_1(J, K) = 2\alpha(1 - 2\alpha)J^{-1} + \left(\frac{1}{2} - \alpha\right)^2 K^{-2} \tag{22}$$

the two components of which are in balance is K is of size $J^{1/2}$. Minimizing M_1 subject to $JK = L$ yields $J = \gamma_1 L^{2/3}$ and $K = \gamma_1^{-1} L^{1/3}$ where $\gamma_1 = \{\alpha(1 - 2\alpha)(1/2 - \alpha)^{-2}\}^{1/3}$. For a two-sided interval the asymptotic mean square error is proportional to

$$M_2(J, K) = \alpha\left(\frac{5}{4} - \alpha\right) J^{-1} + (1 - \alpha)^2 K^{-2} \tag{23}$$

where the two components are in balance if K is of size $J^{1/2}$. Minimizing M_2 subject to $JK = L$ yields $J = \gamma_2 L^{2/3}$ and $K = \gamma_2^{-1} L^{1/3}$ where $\gamma_2 = \{(1/2)(1 - \alpha)^{-2}\alpha(5/4 - \alpha)\}^{1/3}$. For accurate approximation, the product $JK (= L)$ should be at least size n^3 where n is the sample size, and preferably an order of magnitude larger. Due to the discreteness of the empirical distribution, it is extremely desirable that $(J+1)\alpha$, $(J+1)/K$ and $K/2$ all be integers. Failure to take into proper account this discreteness can affect the quality of the approximation. For example, while entry 975 marks the upper 2.5% tail if $J = 999$, when $J = 499$ there is no quantile which marks the upper 2.5% tail: entry 488 marks the upper 2.4% while entry 487 marks the upper 2.6% tail. Since there is much latitude in choosing $L(= JK)$, the latter requirements for integer product and divisors can easily be met.

In the next section we show how the above algorithm can be employed. We illustrate the use of the double bootstrap to derive a confidence interval in the situation where the assumptions of classical asymptotic normal theory do not hold. In particular, the sample size is not large enough to sustain appeal to a central limit theorem for asymptotic normality.

3. Double Bootstrapping a Nonlinear Model

The Cobb-Douglas production function with additive error is given by

$$Q_t = \alpha_1 {L_t}^{\beta} {K_t}^{\gamma} + \epsilon_t \tag{24}$$

and we are interested in testing whether constant returns to scale holds, i.e., $H_0 : \theta = 1.0$ against $H_A : \theta \neq 1.0$ where $\theta = \beta + \gamma$. Using the 30 observations in Table 12.3 of Judge et al. (1988, p. 512) we run a linear regression on the logarithms, as if the model had multiplicative errors, to obtain initial estimates. Using these initial estimates, the nonlinear estimates of the parameters (with standard errors in parentheses) are easily found to be $\hat{\alpha} = 1.330$ $(0.1289), \hat{\beta} = 0.723$ $(0.1316), \hat{\gamma} = 0.687$ (0.1148) and cov$(\hat{\beta}, \hat{\gamma}) = 0.0003$ so an estimate of the returns to scale parameter is $\hat{\theta} = 1.409$ with standard error $\hat{\sigma}(\hat{\theta}) = 0.1732$. Thus an asmptotic normal 95% confidence interval for θ is [1.070, 1.749].

From this initial nonlinear regression we obtain a vector of fitted values for the dependent variable $\hat{Q}$ and vector of (rescaled) residuals. The residuals are shuffled to form e_j^* and added to $\hat{Q}$ to form a pseudo-dependent variable $Q_j^*, j = 1, 2 \ldots J$. As initial estimates for this nonlinear regression we use $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$, and obtain a bootstrap estimate of the returns to scale parameter $\hat{\theta}_j^*$ and its standard error, $\sigma(\theta_j^*)$, and fitted values for the pseudo-dependent variable, $\hat{Q}^*$. Form $R_j^* = (\hat{\theta}_j^* - \hat{\theta})/\sigma(\hat{\theta}_j^*)$. If we were to repeat this procedure 1999 times the percentile bootstrap confidence interval would be $[\hat{\theta}^*_{(50)}, \hat{\theta}^*_{(1950)}]$, or [1.112,1.804]. The percentile-t interval would be $[\hat{\theta} - R^*_{(1950)}\sigma(\hat{\theta}), \hat{\theta} - R^*_{(50)}\sigma(\hat{\theta})]$, or [1.061, 1.764].

To implement the double bootstrap, for each first-stage resample j we shuffle e^* to obtain e^{**} and form $Q^{**} = \hat{Q}^* + e^{**}$. Again we use $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$ as initial values and obtain second-stage estimates of the parameter of interest and its standard error, $\hat{\theta}_{jk}^{**}$ and $\sigma(\hat{\theta}_{jk}^{**})$ to form a pivot $\hat{R}_{jk}^{**}$, for $k = 1, 2, \ldots K$. Count the number of times that $\hat{R}_{jk}^{**}$ is less than $\hat{R}_j^*$ and divide this number by K to obtain Z_j. After all bootstrapping operations are complete, sort the Z_j in ascending order. Since we seek a 95% interval and $J = 1999$, we find that $Z_{(50)} = 0.036$ and $Z_{(1950)} = 0.968$. Therefore appropriate quantiles of $\hat{R}^*$ are $0.036 \cdot 2000 = 72$ and $0.968 \cdot 2000 = 1936$, and the double bootstrap interval $[\hat{\theta} - \hat{R}^*_{(1936)}\sigma(\hat{\theta}), \hat{\theta} - \hat{R}^*_{(72)}\sigma(\hat{\theta})]$ = [1.078, 1.719]. The histogram of Z_j can be a useful diagnostic graphic, as advocated by Vinod and

McCullough (1995). If Z is not uniform, this constitutes evidence that the model is incorrect. Vinod (1995) gives two examples of non-uniform Z. A test statistic for uniformity is $-2\sum \ln(Z_j)$ which is distributed as χ^2 with $2J$ degrees of freedom under the null (Stuart and Ord, 1991, §23.11). In the present example, the value of the test statistic is 3942.6 which has a marginal significance level of 0.965; hence we do not reject the hypothesis of uniform Z.

This above program, written in RATS *v*4.2 took 11 hours to run on a 90MHz Pentium. However, a recent advance in bootstrap methodology enabled us to reduce this to four hours. Davidson and MacKinnon (1997) describe the use of many different artificial regressions, such as the Gauss-Newton Regression (GNR), as a means of reducing computational time in nonlinear regression without sacrificing accuracy. In essence, each nonlinear estimation of bootstrap estimates is replaced by m linear estimations, where m is a small integer. Usually m linear estimations take less time than one nonlinear estimation, and subject to regularity conditions the GNR bootstrap with one nonlinear and JKm linear regressions is as accurate as the bootstrap with JK nonlinear regressions. To see this surprising result, from Section 2 we know that the bootstrap is accurate to some order k, in particular the nonlinear answer approximates the true interval to $O(n^{-k/2})$. Davidson and MacKinnon (1997) prove that the difference between the nonlinear estimate and the estimate obtained after m rounds of the GNR is $O(n^{-(m+1)/2})$. Since $O(n^{-k/2}) + O(n^{-(m+1)/2}) = O(n^{-\min(k,m+1)/2})$, by choosing $m = k - 1$, the GNR bootstrap will approximate the true interval to the same order of accuracy as the full nonlinear bootstrap. While the GNR for the single bootstrap is covered in Davidson and MacKinnon (1997), the use of the GNR in the double bootstrap merits brief exposition. We assume the reader is familiar with the principles of the GNR, which is explained in great detail in Davidson and MacKinnon (1993, ch. 6).

We need the partials of the production function, which are:

$$\partial Q/\partial\alpha = K^\beta L^\gamma = f_1(\alpha,\beta,\gamma) \tag{25}$$

$$\partial Q/\partial\beta = \alpha \ln(K) K^\beta L^\gamma = f_2(\alpha,\beta,\gamma) \tag{26}$$

$$\partial Q/\partial\gamma = \alpha K^\beta \ln(L) L^\gamma = f_3(\alpha,\beta,\gamma) \tag{27}$$

each of which becomes a vector of length 30 when the observed values of K and L are plugged in.

We are concerned with GNR estimates $\tilde\alpha, \tilde\beta$, and $\tilde\gamma$ of the parameters α, β, and γ in a regression with Q^* (for the first stage) or Q^{**} (for the second stage) as the dependent variable. We also need fitted values of the dependent variable computed using $\tilde\alpha, \tilde\beta$, and $\tilde\gamma$, denoted $\tilde Q$. Let $r = Q - \tilde Q$. The GNR is a regression of r on f_1, f_2, and f_3. For a first stage regression, the dependent variable is Q_j^*. The following algorithm produces bootstrap estimates of the coefficients and the standard errors of the coefficients, so that a pivot can be formed.

- Initialize the procedure by using values from the nonlinear regression, setting $\tilde\alpha = \hat\alpha, \tilde\beta = \hat\beta, \tilde\gamma = \hat\gamma$, and $\tilde Q = \hat Q$.

Table I. Bootstrap intervals for returns-to-scale parameter using nonlinear and GNR methods.

Interval	Nonlinear	GNR
Percentile	[1.1122, 1.8044]	[1.1122, 1.8044]
Percentile-t	[1.0609, 1.7644]	[1.0615, 1.7650]
Double	[1.0772, 1.7191]	[1.0778, 1.7194]

- do i = 1 to m
- compute the three series $f_1(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$, $f_2(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$, and $f_1(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$.
- compute $r = Q_j^* - \tilde{Q}$
- regress r on f_1, f_2 and f_3 yielding estimates $\alpha_i, \beta_i, \gamma_i$.
- update the coefficient estimates: $\tilde{\alpha} = \tilde{\alpha} + \alpha_i, \tilde{\beta} = \tilde{\beta} + \beta_i, \tilde{\gamma} = \tilde{\gamma} + \gamma_i$.
- update the fitted values: $\tilde{Q} = \tilde{\alpha} K^{\tilde{\beta}} L^{\tilde{\gamma}}$
- end do

After the loop set $\hat{\theta}_j = \tilde{\beta} + \tilde{\gamma}$. The standard errors for the coefficients are those from the final regression of r on f_1, f_2, and f_3, and so $\sigma(\hat{\theta}_j^*)$ and R^* can be computed. For the second-stage merely substitute Q^{**} in place of Q^*. Table I, in which we chose $m = 4$ shows the bootstrap intervals calculated using nonlinear regression and the GNR. The agreement is quite good.

4. The Double Bootstrap with No Pivot

As Hall (1992, p. 128) notes, pivoting can be difficult to sustain when the scale parameter cannot be stably estimated, i.e., in the case of an estimate which is formed via a ratio, examples of which are elasticities from a translog function or flexibilities. While the single bootstrap can be applied without a pivotal transformation, not so the usual double bootstrap. Since iterating the bootstrap principle yields better confidence intervals, we would like to be able to double bootstrap when no pivot is available. How do we do this? The question was answered by Shi (1992).

Let F be the cdf of X with parameter θ and let x_n denote the sample. A $100\alpha\%$ upper confidence limit $\bar{T}$ can be constructed from bootstrap resamples by choosing α_1 and $\bar{T} = \bar{T}(x_n, \alpha_1) = \hat{G}^{-1}(\alpha_1)$ such that

$$P\{\theta \leq \bar{T}(x_n, \alpha_1)|F\} = \alpha \tag{28}$$

where

$$\hat{G}^{-1}(z) \equiv P(\hat{\theta}^* \leq z|\hat{F}_n) \tag{29}$$

For $j = 1, 2, \ldots J$ first-stage resamples consider the ordered bootstrap estimates of $\theta : \hat{\theta}^*_{(1)}, \hat{\theta}^*_{(2)}, \ldots \hat{\theta}^*_{(J)}$. The percentile bootstrap uses $\hat{\theta}^*_{[(J+1)\alpha]}$ as an estimate of the upper limit.

Suppose we seek a $100\alpha\%$ upper limit $\bar{T}$. Consider the system of equations

$$P\{\theta \leq \bar{T}(x_n, \alpha_1)|F\} = \alpha \tag{30}$$
$$P\{\hat{\theta}^* \leq \bar{T}(x_n, \alpha_1)|\hat{F}_n\} = \alpha_1 \tag{31}$$

where $\hat{F}_n$ is the empirical distribution of x_n. Note that in (30), $\bar{T}(x_n, \alpha_1)$ is random and depends upon F and α_1, while in (31) it is fixed. Similarly, in (30), θ is fixed while in (31) $\hat{\theta}^*$ varies. Obviously, once α_1 is determined, the correct upper limit $\bar{T}(x_n, \alpha_1)$ can be calculated as $\hat{\theta}^*_{([(J+1)\alpha_1])}$. However, α_1 is unknown and therefore we seek an estimate $\hat{\alpha}_1$ of α_1.

To obtain such an estimate, analagous to (28) for the original sample, for the bootstrap resample x_n^* define

$$P\{\hat{\theta} \leq \bar{T}^*(x_n^*, \hat{\alpha}_1)|\hat{F}_n\} = \alpha \tag{32}$$

and analagous to (29) define

$$\hat{G}^{*-1}(z) \equiv P(\hat{\theta}^{**} \leq z|\hat{F}_n^*) \tag{33}$$

where $\hat{F}_n^*$ is the empirical distribution of x^* and $\hat{\theta}^{**}$ is calculated from x_n^{**}, a second-stage bootstrap resample of the first-stage bootstrap resample x_n^*.

Thus we have another system of equations

$$P\{\hat{\theta} \leq \bar{T}^*(x_n^*, \hat{\alpha}_1)|\hat{F}_n\} = \alpha \tag{34}$$
$$P\{\hat{\theta}^{**} \leq \bar{T}^*(x_n^*, \hat{\alpha}_1)|\hat{F}_n^*\} = \hat{\alpha}_1 \tag{35}$$

which is equivalent to choosing $\hat{\alpha}_1$ so that

$$P[\hat{\theta} \leq \hat{G}^{*-1}(\hat{\alpha}_1)|\hat{F}_n] = \alpha \tag{36}$$

Upon rewriting, we have

$$P[P(\hat{\theta}^{**} \leq \hat{\theta}|\hat{F}_n^*) \leq \hat{\alpha}_1|\hat{F}_n] = \alpha \tag{37}$$

which can be solved for $\hat{\alpha}_1$. We note that the inner probability $Q(x_n^*) = P(\hat{\theta}^{**} \leq \hat{\theta}|\hat{F}_n^*)$ is a function of the first-stage bootstrap resample. Before we formed Z_j for the root, now instead we compute

$$Q_j = \#(\hat{\theta}_{jk}^{**} \leq \hat{\theta})/K \tag{38}$$

i.e., for each first stage, Q_j is the proportion of second stage estimates less than or equal to the initial estimate.

Shi shows that the one-sided confidence interval $I_1 = (-\infty, \bar{T}(x_n, \hat{\alpha}_1))$ has coverage probability $O(n^{-1})$, which is the same as the accelerated bias-corrected method, and faster than the percentile bootstrap. His Monte Carlo study shows that his method outperforms the various single bootstrap methods, including accelerated bias-correction. In the regression context, a Monte Carlo study by Dorfman, Kling, and Sexton (1990) estimated confidence intervals for Waugh's (1964) flexibility measure, which is a ratio of random variables. They applied the percentile bootstrap, bias-corrected, and accelerated bootstraps; all yielded approximately the same coverage, and all failed to achieve nominal coverage. The usual double bootstrap cannot be applied to this problem because a stable pivot does not exist. Letson and McCullough (1997) use Shi's method on the Waugh data, and their Monte Carlo study comparing Shi's method to the percentile bootstrap shows that the former achieves nominal coverage while the latter does not. Next we apply Shi's Method to the problem of bootstrapping substitution elasticities from a translog production function.

5. Double Bootstrapping Translog Elasticities

The single bootstrap has been applied to translog productions functions by Green, Rocke, and Hahn (1987), Eakin, Kelly, and Buono (1990), and Krinsky and Robb (1986, 1991), among others. Since both the translog methodology and the Berndt and Wood (1975) study are well-known (see Berndt, 1991, ch. 9), we immediately consider the share equations

$$\begin{aligned} S_K &= \alpha_K + \gamma_{KK}\ln(P_K/P_M) + \gamma_{KL}\ln(P_L/P_M) + \\ &\quad \gamma_{KE}\ln(P_E/P_M) + \epsilon_K \end{aligned} \tag{39}$$

$$\begin{aligned} S_L &= \alpha_L + \gamma_{KL}\ln(P_K/P_M) + \gamma_{LL}\ln(P_L/P_M) + \\ &\quad \gamma_{LE}\ln(P_E/P_M) + \epsilon_L \end{aligned} \tag{40}$$

$$\begin{aligned} S_E &= \alpha_E + \gamma_{KE}\ln(P_K/P_M) + \gamma_{LE}\ln(P_L/P_M) + \\ &\quad \gamma_{EE}\ln(P_E/P_M) + \epsilon_E \end{aligned} \tag{41}$$

$$\begin{aligned} S_M &= \alpha_M + \gamma_{KM}\ln(P_K/P_M) + \gamma_{LM}\ln(P_L/P_M) + \\ &\quad \gamma_{EM}\ln(P_E/P_M) + \epsilon_M \end{aligned} \tag{42}$$

The substitution elasticities are obtained by the formulae:

$$\hat{\sigma}_{mn} = \frac{\hat{\gamma}_{mn} + \hat{S}_m\hat{S}_n}{\hat{S}_m\hat{S}_n} \tag{43}$$

$$\hat{\sigma}_{mm} = \frac{\hat{\gamma}_{mm} + \hat{S}_m^2 - \hat{S}_m}{\hat{S}_m^2} \tag{44}$$

where $m, n = K, L, E, M$. We use σ by convention of the translog literature; since standard errors are not used in this section, there is no danger of confusion. Iterated Seemingly Unrelated Regressions (ItSUR) applied to the system (39-41) yields initial estimates of the coefficients and residuals, both of which are denoted by a circumflex. Coefficients for (42) are obtained via the 'adding-up' conditions. Fitted shares values in (43-44) are evaluated at the mid-point of the series. For complete details see Berndt (1991, ch. 9), with whose results our initial estimates agree. Before the bootstrap resampling by random uniform draws, the residuals are rescaled by $\sqrt{N/(N-k)}$ where N is the number of observations and k is the number of right-hand-side variables plus one. As is typical of bootstrapping systems of equations, in order to preserve the covariance structure of the errors, each vector of residuals is permuted the same way, e.g., the entries of $\hat{\epsilon}_K$ which constitute $\hat{\epsilon}^*_K$ are the same entries of $\hat{\epsilon}_L$ which constitute $\hat{\epsilon}^*_L$.

The first-stage replicates of the dependent variables are

$$\begin{aligned} S^*_K &= \hat{\alpha}_K + \hat{\gamma}_{KK}\ln(P_K/P_M) + \hat{\gamma}_{KL}\ln(P_L/P_M) + \\ &\quad \hat{\gamma}_{KE}\ln(P_E/P_M) + \hat{\epsilon}^*_K \qquad (45) \\ S^*_L &= \hat{\alpha}_L + \hat{\gamma}_{KL}\ln(P_K/P_M) + \hat{\gamma}_{LL}\ln(P_L/P_M) + \\ &\quad \hat{\gamma}_{LE}\ln(P_E/P_M) + \hat{\epsilon}^*_L \qquad (46) \\ S^*_E &= \hat{\alpha}_E + \hat{\gamma}_{KE}\ln(P_K/P_M) + \hat{\gamma}_{LE}\ln(P_L/P_M) + \\ &\quad \hat{\gamma}_{EE}\ln(P_E/P_M) + \hat{\epsilon}^*_E \qquad (47) \end{aligned}$$

Replacing S_K, S_L, and S_E in (39-41) with the replicated series S^*_K, S^*_L, and S^*_E and applying ItSUR yields first-stage bootstrap coefficient estimates $\hat{\alpha}^*_K$, $\hat{\gamma}^*_{KK}$, etc. and first-stage bootstrap estimates of the elasticities $\hat{\sigma}^{j*}_{mn}$ and $\hat{\sigma}^{j*}_{mm}$. Random sampling of $\hat{\epsilon}^*_K$, $\hat{\epsilon}^*_L$, and $\hat{\epsilon}^*_E$ to preserve covariance structure yields $\hat{\epsilon}^{**}_K$, $\hat{\epsilon}^{**}_L$, and $\hat{\epsilon}^{**}_E$. The second-stage replicates of the dependent variables are

$$\begin{aligned} S^{**}_K &= \hat{\alpha}^*_K + \hat{\gamma}^*_{KK}\ln(P_K/P_M) + \hat{\gamma}^*_{KL}\ln(P_L/P_M) + \\ &\quad \hat{\gamma}^*_{KE}\ln(P_E/P_M) + \hat{\epsilon}^{**}_K \qquad (48) \\ S^{**}_L &= \hat{\alpha}^*_L + \hat{\gamma}^*_{KL}\ln(P_K/P_M) + \hat{\gamma}^*_{LL}\ln(P_L/P_M) + \\ &\quad \hat{\gamma}^*_{LE}\ln(P_E/P_M) + \hat{\epsilon}^{**}_L \qquad (49) \\ S^{**}_E &= \hat{\alpha}^*_E + \hat{\gamma}^*_{KE}\ln(P_K/P_M) + \hat{\gamma}^*_{LE}\ln(P_L/P_M) + \\ &\quad \hat{\gamma}^*_{EE}\ln(P_E/P_M) + \hat{\epsilon}^{**}_E \qquad (50) \end{aligned}$$

Replacing the dependent variables in (39-41) with the second-stage replicates and applying ItSUR yields second-stage estimates of the coefficients $\hat{\alpha}^{**}_K$, $\hat{\gamma}^{**}_K$, etc. and second-stage estimates of the elasticities $\hat{\sigma}^{jk**}_{mn}$ and $\hat{\sigma}^{jk**}_{mm}$. As to the choice of J and K, a natural choice for J is 999. However, for a 95% confidence interval the there is no nearby K which meets the integer-divide requirements. Another natural choice is $J = 1999$, for which optimal K is 245.76. Therefore, we use $J = 1999$ and $K = 250$, whose product more than suffices for the order of magnitude requirement.

To ease the notational burden, we henceforth suppress identification of particular elasticities and use the subscripts to indicate the first (j) and second (k) stage resamples. For each elasticity there exists a J-vector Q, the j-th element of which is determined as follows: $Q_j = \#(\hat{\sigma}^{**}_{jk} \leq \hat{\sigma})/K,\ k = 1, 2, \ldots, K$ where $\#(\cdot)$ indicates the number of times the condition in parentheses holds. After all bootstrapping operations are complete, for each elasticity we have J

Table II. Single and double bootstrap confidence intervals (* indicates that the double bootstrap reverses a single bootstrap conclusion).

Substitution elasticity	Single bootstrap		Double bootstrap		Q	
					$\hat{\alpha}_{1L}$	$\hat{\alpha}_{1U}$
	Lower	Upper	Lower	Upper	Lower	Upper
σ_{KK}	–11.87	–3.60	–8.52	–5.73	0.272	0.800
σ_{KL}	0.47	1.53	0.79	1.06	0.236	0.616
σ_{KE}	–6.36	–0.01	–4.61	–0.42	0.180	0.956
$\sigma_{KM}*$	–0.12	1.09	0.17	0.63	0.164	0.736
σ_{LL}	–1.84	–1.44	–1.71	–1.61	0.252	0.648
σ_{LE}	0.18	1.07	0.38	0.94	0.132	0.916
σ_{LM}	0.45	0.72	0.55	0.65	0.272	0.884
$\sigma_{EE}*$	–18.55	–5.92	–13.99	0.30	0.264	1.000
$\sigma_{EM}*$	0.20	1.57	–0.32	1.06	0.000	0.736
σ_{MM}	–0.48	–0.24	–0.39	–7.44	0.308	0.928

single bootstrap estimates for both $\hat{\sigma}_j^*$ and and Q_j. Let the ordered values be $\hat{\sigma}_{(1)}^*, \hat{\sigma}_{(2)}^*, \ldots, \hat{\sigma}_{(J)}^*$ and $Q_{(1)}, Q_{(2)}, \ldots Q_{(J)}$. The single bootstrap confidence interval is $[\hat{\sigma}^*_{([(J+1)(1-\alpha)])}, \hat{\sigma}^*_{([(J+1)\alpha])}]$. The double bootstrap interval is given by $[\hat{\sigma}^*_{([(J+1)\hat{\alpha}_{1L}])}, \hat{\sigma}^*_{([(J+1)\hat{\alpha}_{1U}])}]$ where the lower and upper estimates of α_1 comes from the Q vector as follows: $\hat{\alpha}_{1L} = Q_{([(J+1)(1-\alpha)])}$ and $\hat{\alpha}_{1U} = Q_{([(J+1)\alpha])}$. We illustrate the precise steps involved in determining the double bootstrap interval by examining the case for σ_{KK}.

We seek a two-sided confidence interval with $\alpha = 0.05$. The upper and lower single bootstrap limits for σ_{KK} are $\hat{\sigma}^*_{([2000\cdot 0.025])} = \hat{\sigma}^*_{(50)} = -11.87$ and $\hat{\sigma}^*_{([2000\cdot 0.975])} = \hat{\sigma}^*_{(1950)} = -3.60$. To obtain the double bootstrap confidence limits we need the estimates $\hat{\alpha}_{1L}$ and $\hat{\alpha}_{1U}$. We have $\hat{\alpha}_{1L} = Q_{([2000\cdot 0.025])} = Q_{(50)} = 0.272$ and $\hat{\alpha}_{1U} = Q_{([2000\cdot 0.975])} = Q_{(1950)} = 0.800$. The lower double bootstrap limit is $\hat{\sigma}^*_{([2000\cdot 0.272])} = \hat{\sigma}^*_{(544)} = -8.52$ and the upper double bootstrap limit is $\hat{\sigma}^*_{([2000\cdot 0.800])} = \hat{\sigma}^*_{(1600)} = -5.73$. Similar calculations produce confidence intervals for the other substitution elasticities, the results of which are presented in Table II.

The program, written in RATS v4.2 and run on a 90MHz Pentium, takes approximately six hours, and results are presented in Table 1. If the double offered no improvement then $\hat{\alpha}_{1L}$ would be 0.025 and the $\hat{\alpha}_{1U}$ would be 0.975 and the double bootstrap interval would coincide with the single bootstrap interval. We see that the corrected percentiles, given in the Q column, can be far from the nominal levels of 0.025 and 0.975 used by the single bootstrap. Indeed, the differences can be so pronounced as to reverse qualitative conclusions based on the single bootstrap intervals. For example, the single bootstrap indicates that σ_{KM} is not significantly different from zero, while the double bootstrap indicates that it is positive, and conversely for σ_{EM}. The single bootstrap indicates that σ_{EE} is negative, while the double bootstrap indicates that it is not significantly different from zero.

We mention here a useful artifice for substantially decreasing computing time when Shi's method is applied to linear regression. Since there is no pivoting, each regression requires only estimation of the slope coefficients, not the standard errors. The typical regression command computes many unneeded statistics. If the econometrics package has a matrix facility, letting X be the matrix of independent variables, define $P = (X'X)^{-1}X'$. Then the first-stage and second-stage coefficients can be obtained by $b_j^* = Py_j^*$ and $b_{jk}^{**} = Py_{jk}^{**}$, respectively. Using this device, Letson and McCullough (1997) reduced the time needed for a double bootstrap from almost fifty to less than nine minutes, and were able to complete their Monte Carlo study in less than two months instead of almost a year.

6. Conclusions

The single bootstrap already is popular in applied economics. The double bootstrap, due to its superior convergence properties, is likely to be of more practical value to economists, especially since median bias, nonlinearity, nuisance parameters, etc. are commonly encountered in economics. Further, just as the single bootstrap can succeed in situations where classical asymptotics fail (e.g., confidence intervals for complicated nonlinear statistics), so the double bootstrap can succeed in cases where the single bootstrap fails (e.g., obtaining consistent confidence intervals for a ridge regression). The double bootstrap also yields a new graphical method for diagnostic checking. By virtue of the fact that the doubly pivoted root converges to a uniform distribution on [0,1] if the model holds, failure of the model to hold can be assessed by examining the distribution of the doubly pivoted root. We explain the double bootstrap and give an algorithm for implementing it. Details on the proper choice of the numbers J of first-stage and K of second-stage iterations are provided. We apply this algorithm to the nonlinear estimation problem of testing returns to scale in a Cobb-Douglas production function with additive error. We also describe a recently developed method based on the Gauss-Newton Regression for reducing computational time when bootstrapping a nonlinear model.

The usual double bootstrap requires a root, which does not exist in all situations. Moreover, when scale cannot be estimated with stable variance, pivoting is difficult to sustain, as in the estimation of a correlation coefficient or some ratios of means. The variance of a ratio of means, for example, when calculated using the delta method, often is insufficiently stable to sustain pivoting. Therefore, we explain Shi's method for double bootstrapping when a pivot is not available. This method is applied to the problem of estimating confidence intervals for substitution elasticities from a translog production function. It is shown that the double bootstrap can reverse conclusions reached by a single bootstrap. We also describe a method for reducing computation time for this double bootstrap.

Acknowledgements

Thanks to J. MacKinnon, M. Veall and an anonymous referee for comments, and to S. Shi for useful discussions.

References

Aczel, A. and Josephy, N. (1992). Improved ARIMA identification with the bootstrap. *Journal of Forecasting*, **11**, 71–80.

Beran, R. (1987). Prepivoting to reduce level error in confidence sets. *Biometrika*, **74**, 457–468.

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

Beran, R. (1990). Refining bootstrap simultaneous confidence sets. *Journal of the American Statistical Association*, **85**, 417–426.

Beran, R. (1995). Prediction in Random Coefficient Regression. *Journal of Statistical Planning and Inference*, **43**, 205–213.

Berndt, E. (1991). *The Practice of Econometrics*. Addison-Wesley, Reading, MA.

Berndt, E. and Wood, D. (1975). Technology, prices, and the derived demand for energy. *Review of Economics and Statistics*, **57**, 259–268.

Booth, J.G. and Hall, P. (1994). Monte Carlo approximation and the iterated bootstrap. *Biometrika*, **81**, 331–340.

Breiman, L. (1992). The little bootstrap and other methods for dimensionality seclection in regression: X-fixed prediction error. *Journal of the American Statistical Association*, **87**, 738–754.

Brownstone, D. (1990). Bootstrapping improved estimators for linear regression models. *Journal of Econometrics*, **44**, 171–188.

Davidson, R. and MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press, New York.

Davidson, R. and MacKinnon, J.G. (1997). Bootstrap testing in nonlinear models. Discussion Paper No. 944, Queen's University Institute for Economic Research, Ontario.

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

DiCiccio, T. and Romano, J. (1988). A review of bootstrap confidence intervals. *Journal of the Royal Statistical Society, Series B*, **50**, 338–354.

Dorfman, J., Kling, C. and Sexton, R.J. (1990). Confidence intervals for elasticities and flexibilities: reevaluating the ratio of normals case. *American Journal of Agricultural Economics*, **72**, 1006–1017.

Eakin, B.K., McMillen, D.P. and Buono, M.J. (1990). Constructing confidence intervals using the bootstrap: an application to a multiproduct cost function. *Review of Economics and Statistics*, **72**, 339–344.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. CBMS-NSF Regional Conference Series in Applied Mathematics No. 38, SIAM, Philadelphia.

Efron, B. and Tibshirani, R. (1993). *Introduction to the Bootstrap*, Chapman and Hall, New York.

Green, R., Rocke, D. and Hahn, W. (1987). Standard errors for elasticities: a comparison of bootstrap and asymptotic standard errors. *Journal of Business and Economic Statistics*, **5**, 145–149.

Hall, P. (1988). Theoretical comparisons of bootstrap confidence intervals. *Annals of Statistics*, **16**, 927–953.

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

Judge, G., Hill, R., Griffiths, W., Lüketphol, H. and Lee, T.-C. (1988). *Introduction to the Theory and Practice of Econometrics, 2e*. J. Wiley and Sons, New York.

Krinsky, I. and Robb, L. (1991). Three methods for calculating the statistical properties of elasticities: a comparison. *Empirical Economics*, **16**, 199–209.

Krinsky, I. and Robb, L. (1986). On approximating the statistical properties of elasticities. *Review of Economics and Statistics*, **68**, 715–719.

Kuk, A. (1989). Double bootstrap estimation of variance under systematic sampling with probability proportional to size. *Journal of Statistical Computing and Simulation*, **31**, 73–82.

LePage, R. and Billard, L. (eds.) (1992). *Exploring the Limits of the Bootstrap*. Wiley Interscience, New York.

Letson, D. and McCullough, B.D. (1997). Better confidence intervals: the double bootstrap with no pivot. Mimeo, Department of Economics, University of Miami.

McAleer, M. and Veall, M.R. (1989). How fragile are fragile inferences? *Review of Economics and Statistics*, **71**, 99–106.

Mammen, E. (1992). *When Does Bootstrap Work?*, Springer-Verlag, New York.

Manteiga, W., Sánchez, J. and Romo, J. (1994). The bootstrap – a review. *Computational Statistics*, **9**, 165–205.

Martin, M. (1990). On bootstrap iteration for coverage correction in confidence intervals. *Journal of the American Statistical Association*, **85**, 1105–1118.

Martin, M. (1992). On the double bootstrap. In C. Page and R. LePage (eds.), *Computing Science and Statistics: The Interface*, 73–78. Springer-Verlag, New York.

McCullough, B.D. (1994). Bootstrapping forecast intervals: an application to AR(p) models. *Journal of Forecasting*, **13**, 51–66.

McCullough, B.D. (1996). Estimating forecast intervals when the exogenous variable is stochastic. *Journal of Forecasting*, **15**, 293–304.

McCullough, B.D. and Vinod, H. (1993). Implementing the single bootstrap: some computational considerations. *Computational Economics*, **6**, 1–15.

Navidi, W. (1989). Edgeworth expansions for bootstrapping regression models. *The Annals of Statistics*, **17**, 1472–1478.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag, New York.

Scholz, F.W. (1994). On exactness of the parametric double bootstrap. *Statistica Sinica*, **4**, 477–492.

Shi, S.G. (1992), Accurate and efficient double bootstrap confidence intervals. *Computational Statistics and Data Analysis*, **13**, 21–32.

Stine, R. (1985). Bootstrap prediction intervals for regression. *Journal of the American Statistical Association*, **80**, 1026–1031.

Stuart, A. and Ord, J.K. (1991). *Kendall's The Advanced Theory of Statistics, Vol. 2*, Oxford University Press, New York.

Thombs, L. and Schuchany, W. (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, **85**, 486–492.

Veall, M. (1992). Bootstrapping the process of model selection: an econometric example. *Journal of Applied Econometrics*, **7**, 93–99.

Vinod, H.D. (1993). Bootstrap, jackknife resampling and simulation: applications in econometrics. In G.S. Maddala, C.R. Rao and H.D. Vinod (eds.), *Handbook of Statistics: Econometrics, Vol. 11*, 629–661. North-Holland, New York.

Vinod, H.D. (1995). Double bootstrap for shrinkage estimators. *Journal of Econometrics*, **68**, 287–302.

Vinod, H.D. and McCullough, B.D. (1994). Estimating demand and supply elasticities. *Journal of Asian Economics*, **5**, 367–379.

Vinod, H.D. and McCullough, B.D. (1995). Estimating cointegrating parameters: an application of the double bootstrap. *Journal of Statistical Planning and Inference*, **43**, 147–156.

Waugh, F.V. (1964). Demand and price analysis: some estimates from agriculture. U.S. Department of Agriculture/Economic Research Service. Technical Bulletin 1316.