Better Confidence Intervals: The Double Bootstrap with No Pivot

David Letson and B.D. McCullough

The double bootstrap is an important advance in confidence interval generation because it converges faster than the already popular single bootstrap. Yet the usual double bootstrap requires a stable pivot that is not always available, e.g., when estimating flexibilities or substitution elasticities. A recently developed double bootstrap does not require a pivot. A Monte Carlo analysis with the Waugh data finds the double bootstrap achieves nominal coverage whereas the single bootstrap does not. A useful artifice dramatically decreases the computational time of the double bootstrap.

Key words: confidence interval, convergence, elasticity, flexibility, iterated bootstrap, pivot.

When reporting point estimates that are nonlinear functions of estimated parameters, researchers frequently fail to report confidence intervals. Examples of such point estimates are elasticities from a translog function or flexibilities. In large part this omission occurs because analytic calculation of such confidence intervals can be technically difficult, as in the case of a Taylor expansion to approximate the standard error of a nonlinear function of estimated parameters (the delta method) or the construction of Fieller intervals. However, the era of reporting only point estimates should be over (Dorfman, Kling, and Sexton; hereafter DKS). Even when derivation of a confidence interval is analytically intractable, bootstrap methods can produce reliable confidence intervals with minimal difficulty. These confidence intervals can accommodate asymmetric intervals when the data so indicate, whereas the usual asymptotic normal approximation always imposes the assumption of a symmetric sampling distribution. Efron and Tibshirani provide an excellent introduction to the single bootstrap.[1] For a more sophisticated nontechnical treatment, see Davison and Hinkley.

The single bootstrap has been applied to obtain confidence intervals in a number of situations: flexibilities (DKS), translog production functions (Eakin, McMillan, and Buono); supply and demand elasticities (Vinod and McCullough 1995b); intervals for autoregressive processes (Thombs and Schuchany 1990, McCullough 1994); prediction for random coefficient models (Beran 1995); prediction for stochastic independent variables (McCullough 1996); forecasting demand (Veall) and supply (Prescott and Stengos); welfare estimates in recreational demand models (Kling); and benefits estimates in dichotomous choice contingent valuation (Cooper). See the survey by Vinod (1993) for additional applications in economics.

The reliability of a confidence interval is gauged by its coverage; i.e., the proportion of times that the estimated interval covers the true parameter. A 90% confidence interval has nominal coverage of 90%, but its actual coverage might be either higher or lower. In a Monte Carlo study using the Waugh data, DKS compare Fieller intervals, the delta method, and three types of the single bootstrap, and find that the single bootstrap does not achieve nominal coverage. This result is dispiriting because the single bootstrap can be easier to implement than the other methods.

The drawback to the single bootstrap is its slow rate of convergence, sometimes manifesting itself by achieving less than nominal

---

David Letson is assistant professor of marine affairs and economics at the University of Miami. B.D. McCullough is senior economist at the Federal Communications Commission in Washington, DC.

We thank S. Shi for illuminating conversations, H.D. Vinod for useful suggestions, and two anonymous referees for detailed and valuable reports. The views expressed herein are those of the authors and do not necessarily reflect those of the Federal Communications Commission.

[1] Because the bootstrap is a class of methods, we distinguish between single and double bootstraps and note that there is a variety of each type.

Amer. J. Agr. Econ. 80 (August 1998): 552–559

coverage (see Hall 1992, and Shao and Tu). Defining coverage error as the difference between nominal coverage and actual coverage, for a one-sided interval, the asymptotic normal and percentile bootstrap intervals both have coverage error $O(1/\sqrt{n})$ where n is the sample size. The percentile-t and bias-corrected single bootstrap intervals have coverage error $O(1/n)$. Thus, as sample size increases, the difference between nominal coverage and actual coverage will be less for a bias-corrected single bootstrap than for a percentile single bootstrap. However, DKS's study shows that for small samples the difference in coverage error might not be noticeable. Converging even faster than the single bootstrap, the usual double bootstrap interval based on a pivot has coverage error $O(n^{-3/2})$.

The requirement of a pivot can be an obstacle to implementing the double bootstrap, however. A pivot is simply a function of the data and an unknown parameter that has the same distribution for all values of the unknown parameter. For example, the recentered sample mean, $\bar{x} - \mu$, is not a pivot, because its distribution depends on the variance. By contrast, in the case of sampling from a normal population where the population standard deviation (σ) is known, the usual z-statistic, $z = (\bar{x} - \mu)/\sigma$, is a pivot because its distribution, standard normal, is the same for all values of μ. If the sample is not normal or σ is estimated by s, then $z = (\bar{x} - \mu)/s$ is asymptotically pivotal because its distribution is asymptotically standard normal no matter what the value of the unknown parameter, μ. Sometimes a pivot is readily available, and the usual double bootstrap can be applied in a straightforward fashion, as in the case of cointegrating regressions (Vinod and McCullough 1995a). At other times, a natural pivot might not be available but the researcher can, with some difficulty, construct one, as in the case of ridge regression (Vinod 1995) or estimation of Euler equations (Vinod 1998). For many problems, however, a stable pivot is not available and one cannot be constructed, such as for translog models or the estimation of flexibilities. A stable pivot means that the scale parameter in the denominator of the pivot should have a small variance (Hall 1992, p. 18). If it has a large variance then the ratio will be unstable and the bootstrap will fail, a situation encountered by McCullough and Waldon.

The usual double bootstrap operates as an additive correction, either to the interval endpoint (Hall 1986) or to the nominal coverage level (Beran 1987) and requires the existence of a stable pivot to effect the necessary pivotal transformation. When a pivot either does not exist or cannot be stably estimated, the usual double bootstrap cannot be applied. A standard example is the case of determining a confidence interval for a ratio of estimated parameters, such as elasticities from a translog function or flexibilities. Shi has developed a double bootstrap requiring only a consistent estimator and does not require a pivoting transformation and therefore is applicable in situations where the usual double bootstrap is not.

Shi's double bootstrap is proposed as a method for improving interval generation. We outline bootstrap theory, emphasizing the role of the pivotal transformation. We then introduce Shi's double bootstrap, explaining how to apply it and how it differs from the usual double bootstrap. To facilitate comparisons with previous work (e.g., Miller, Capps, and Wells; DKS), double bootstrap confidence intervals are derived for flexibilities based on Waugh's classic study. Finally, we report on a Monte Carlo study in which the single bootstrap interval fails to achieve nominal coverage while the double bootstrap interval achieves nominal coverage. To our knowledge, this is the first Monte Carlo of the double bootstrap in the economics literature.

The Bootstrap

To fix ideas, five types of confidence intervals are presented for the univariate case: classical asymptotic, single bootstrap with no pivot (percentile bootstrap), single bootstrap with a pivot (percentile-t bootstrap), double bootstrap with pivot (usual double bootstrap), and double bootstrap with no pivot (Shi's double bootstrap). Let $\{x_n\}$ be a sample of size n on a random variable X with finite variance. In deriving a two-sided 95% confidence interval for the mean, μ, the classical asymptotic interval is known to be $\bar{x} \pm t_{0.025}s$, where $\bar{x}$ is the sample mean, s is the sample standard deviation, and $t_{0.025}$ is the upper 2.5% tail of Student's-t distribution. In the case that X is normal, this interval is exact; otherwise it has coverage error $O(1/\sqrt{n})$.

The single bootstrap interval is constructed as follows [see McCullough and Vinod (1993) for details]. Form the residuals $e_i = x_i - \bar{x}$, $i = 1, 2, \dots, n$ and denote the collection **e**. Rescale the residuals by $\sqrt{n/(n-k)}$ where $n - k$ is the number of degrees of freedom. Shuffle the residuals by random draws with replacement from a uniform distribution, sampling n of the rescaled $\mathbf{e}$ to form a vector of bootstrap residuals e_j^*, and repeat this process J times, yielding $e_1^*, e_2^*, \ldots, e_J^*$. For each j form a bootstrap resample x_j^* and calculate its mean, $\bar{x}_j^*$. Denote the ordered means by $\bar{x}_{(1)}^*, \bar{x}_{(2)}^*, \ldots, \bar{x}_{(J)}^*$. If $J = 999$, say,[2] then a 95% percentile confidence interval for μ is $[\bar{x}_{(25)}^*, \bar{x}_{(975)}^*]$. This confidence interval has coverage error $O(1/n)$. The one-sided 95% percentile confidence intervals, $[-\infty, \bar{x}_{(950)}^*]$ and $[\bar{x}_{(50)}^*, \infty]$, have coverage error $O(1/\sqrt{n})$.

The accuracy of the single bootstrap confidence interval can be improved by employing a pivotal transformation. For each bootstrap resample x_j^* calculate its standard deviation $\hat{\sigma}_j^*$ and form the pivotal quantity

$$\text{(1)} \quad \hat{R}_j^* = (\bar{x}_j^* - \bar{x})/\hat{\sigma}_j^*.$$

A 95% confidence interval for the true value of R is $[\hat{R}_{(25)}^*, \hat{R}_{(975)}^*]$. A test of the null hypothesis H_0: $\mu = \mu_0$ may be conducted by comparing $\hat{R} = (\bar{x} - \mu_0)/\hat{\sigma}$ to the confidence interval. The percentile-t bootstrap confidence interval for μ can be recovered by unraveling the pivot as follows[3]

$$\text{(2)} \quad [\bar{x} - \hat{R}_{(975)}\hat{\sigma}, \bar{x} - \hat{R}_{(25)}^*\hat{\sigma}].$$

This confidence interval has accuracy $O(n^{-3/2})$. The increase in accuracy is due to the fact that the distribution of R is more nearly pivotal (depends on no unknown parameters) than the distribution of X [see Hall (1992) for details]. A one-sided bootstrap confidence interval produced with this pivoting transformation has accuracy $O(1/n)$.

However, as Hall (1992, p. 128) notes, pivoting can be difficult to sustain when the scale parameter cannot be stably estimated (i.e., if the variance of the scale parameter is large). This is evident from equation (1): if $\hat{\sigma}_j^*$fluctuates greatly from one bootstrap resample to another, then $\hat{R}_j^*$ will wildly fluctuate, destroying the bootstrap's accuracy. Examples of cases in which this can occur are elasticities from a translog function or flexibilities, where there is no natural standard error, and a crude approximation such as a Taylor series expansion must be used. While the single bootstrap can be applied without a pivotal transformation, this is not so for the usual double bootstrap.

To better motivate the double bootstrap with no pivot, the usual double bootstrap is described, which is constructed as follows [see McCullough and Vinod (1998) for details]. Let $\hat{\sigma}$ be the standard deviation of the original sample, and let $\hat{\sigma}_j^*$ be the standard deviation of the jth bootstrap resemble. For the original sample there exists the root $R = (\bar{x} - \mu)/\hat{\sigma}$. For the first-stage bootstrap calculate J roots $R_j^* = (\bar{x}_j^* - \bar{x})/\hat{\sigma}_j^*$. Proceeding to the second stage, shuffle e_j^* to form e_{jk}^{**}. Repeat this procedure a large number of times, K. On each kth second-stage replication, form the root

$$\text{(3)} \quad R_{jk}^{**} = (\bar{x}_{jk}^{**} - \bar{x}_j^*)/\hat{\sigma}_{jk}^{**}$$

For each j define

$$\text{(4)} \quad Z_j = \#(R_{jk}^{**} \le R_j^*)/K, \quad k = 1, 2, \ldots, K$$

where $\#(\cdot)$ is the number of times the condition in parentheses is true. Thus, Z_j is the proportion of times that the second-stage root is less than first-stage root. If a pivot cannot be estimated stably, then this method cannot be applied.

Note that $Z_j \in [0,1]$. An important result is that if R is a pivot having a continuous distribution, then Z_j converges to a uniform distribution. This justifies the use of the Z_j as a diagnostic tool, as advocated by Vinod and McCullough (1995). If the Z_j are found to be nonuniform, then the model may be misspecified. Vinod (1995) gives two examples of nonuniform Z_j.

Let the ordered values be $Z_{(1)}, Z_{(2)}, \ldots, Z_{(J)}$ and suppose $\alpha = 0.05$ and $J = 999$. Define $LO = [J \cdot Z_{(25)}]$ and $UP = [J \cdot Z_{(975)}]$ where $[\cdot]$ is the greatest integer function, e.g., $[26.3] = 26$. The double bootstrap confidence interval for the pivot is $[R_{(LO)}^*, R_{(UP)}^*]$. A confidence interval for μ can be obtained by unraveling the pivot:

$$\text{(5)} \quad [\bar{x} - \hat{R}_{(UP)}^*\hat{\sigma}, \bar{x} - \hat{R}_{(LO)}^*\hat{\sigma}].$$

[2] For an exact test, it is important the $\alpha(J + 1)$ be an integer. The reason is that we are not ordering just the J bootstrap estimates $\bar{x}_j^*$, but the J bootstrap estimates and the initial statistic, $\bar{x}$. Thus the rank of $\bar{x}$ can have $J + 1$ possible outcomes, not just J. So if $\alpha = 0.05$ and $J = 100$ yields $0.05(101) = 5.05$, we would reject if the rank of $\bar{x}$ was less than 5, which would have a probability of $5/101 = 0.0495$ that does not exactly equal α. If $J = 99$, there would be 100 possible ranks, and we would reject when the rank of $\bar{x}$ is less than five, which has probability 5/100, which is exact. See Davidson and MacKinnon (1997) for an extensive discussion of this point.

[3] Some readers might be concerned that $R_{(975)}$, the upper critical value for R, is used to calculate the lower limit of the confidence interval for μ and might instinctively think that the confidence interval should be constructed as $[\bar{x} + \hat{R}_{(25)}^*\hat{\sigma}, \bar{x} + \hat{R}_{(975)}^*]$ (recall that $\hat{R}_{(25)}^* < 0$; hence, the plus sign for the lower limit). However, Hall (1986) shows that such a construction is incorrect.

Booth and Hall provide needed guidance concerning the choice of the number of J first-stage iterations and K second-stage iterations. J and K make separate contributions to the overall accuracy of the bootstrap approximation, the former contributing to variance and the latter to (squared) bias. For fixed $L = JK$, it is possible to choose J and K to minimize the square error of the bootstrap approximation. In particular, for fixed J there is an optimal K, both above and below which performance degrades. J and K cannot be chosen arbitrarily. Their values are determined, in part, by the nominal coverage of the confidence interval, $1 - \alpha$, the sample size, n, and whether the interval in question is one-sided or two-sided.

Booth and Hall recommend that L should be at least size n^3, preferably an order of magnitude larger. Given the sample sizes that predominate in economics, this is not an onerous requirement. Due to the discrete nature of the bootstrap distribution, it is best to take $K = (J + 1)\alpha$, $(J + 1)/K$, and $K/2$ to be integers. Because J is large, this typically is not difficult to do. Choose $J = \gamma L^{2/3}$ and $K = L^{1/3}/\gamma^{-1}$. For a one-sided interval, $\gamma = \{\alpha(1 - 2\alpha)(1/2 - \alpha)^{-2}\}^{1/3}$ and for a two-sided interval choose $\gamma = \{0.5(1 - \alpha)^{-2}\, \alpha(5/4 - \alpha)\}^{1/3}$. For the example presented in the subsequent section, $\alpha = 0.10$ is set for a two-sided interval. A natural choice for J is 999. However, there is no nearby K that meets the integer-divide requirements. Another natural choice is $J = 1{,}999$, for which optimal K is 245.76. Therefore, $J = 1{,}999$ and $K = 246$ are used, whose product more than suffices for the order of magnitude requirement.

Shi's Double Bootstrap

The usual double bootstrap cannot be executed if a pivot is not available, in which case Shi proves that the correct unpivoted confidence interval is given by

$$(6) \quad P[P(\theta^{**} \leq \theta|\hat{F}_n^*) \leq \hat{\delta}|\hat{F}_n] = \alpha$$

where $\hat{F}_n^*$ is the empirical distribution of F based on a sample of size n and $\hat{\delta}$ is an estimate of the correction factor by which the critical value must be adjusted so that it has size α. The inner probability $Q^* = P(\hat{\theta}^{**} \leq \hat{\theta}|\hat{F}_n^*)$ in equation (6) shows precisely how $\hat{\delta}$ is determined. Shi proves that his one-sided bootstrap confidence limit has coverage error at least $O(1/n)$, which is the same as the one-sided pivoted single bootstrap interval. The two-sided interval has even faster convergence. The coverage error of Shi's double bootstrap is better than the percentile method and no larger than the more computationally complex single bootstrap methods such as (accelerated) bias-correction. Moreover, Shi's double bootstrap does not rely on normality assumptions to effect transformations, nor does it require higher moment calculations that can be extremely imprecise for small samples. Thus, when a pivot is not available, Shi's double bootstrap is preferred over the more sophisticated single bootstraps. Shi's Monte Carlo study demonstrates that the double bootstrap produces more reliable confidence intervals than single bootstrap methods with no pivot.

The double bootstrap confidence interval with no pivot is constructed as follows. Instead of forming Z_j, form Q_j as follows:

$$(7) \quad Q_l = \#(\bar{x}_{jk}^{**} \leq \bar{x})/K, \quad k = 1, 2, \ldots, K$$

noting that $Q_j \in [0,1]$ and is asymptotically symmetrically distributed if the model is correctly specified. Let the ordered values be $Q_{(1)}, Q_{(2)}, \ldots, Q_{(J)}$. Define $l = [J \cdot Q_{(25)}]$ and $u = [J \cdot Q_{(975)}]$. Shi's 95% double bootstrap confidence interval for μ is $[\bar{x}_{(l)}^*, \bar{x}_{(u)}^*]$.

In more general situations, replace the initial, single and double bootstrap estimators $\bar{x}$, $\bar{x}_j^*$, and $\bar{x}_{jk}^{**}$ of μ, the population parameter, with estimators $\hat{\theta}$, $\hat{\theta}_j^*$, and $\hat{\theta}_{jk}^{**}$ of the parameter of interest, θ, and proceed in the same fashion. In the case of estimating substitution elasticities from a translog production function, a stable pivot is not available, so the usual double bootstrap is not applicable. However, Shi's double bootstrap is applicable. On the initial regression obtain an estimate of the desired elasticity and denote it $\hat{\theta}$. On first- and second-stage iterations obtain $\hat{\theta}_j^*$, $\hat{\theta}_{jk}^{**}$, and Q_j, and proceed as above to determine the double bootstrap confidence interval. McCullough and Vinod (1998) apply Shi's procedure to translog estimation, and also double bootstrap a nonlinear estimation problem using Davidson and MacKinnon's (1998) artificial regressions method for decreasing computational time when bootstrapping nonlinear regressions. The next section applies Shi's double bootstrap to obtain confidence intervals for flexibility estimates.

Application to Waugh Data

For comparison with previous work (DKS; Miller, Capps, and Wells) this method is applied to Waugh's demand study. Waugh uses ordinary least squares (OLS) to estimate price and income flexibilities for potatoes, sweet potatoes, tomatoes, grapefruit, apples, and beef, with demand equations of the form

$$(8) \quad P_t = \beta_0 + \beta_1 Q_t + \beta_2 M_t + e_t$$

where P, Q, and M are price, quantity, and income. The subscript t denotes the annual time series observations from 1948 to 1962 except in the case of tomatoes, 1950 to 1962. With price as the dependent variable, the estimated price flexibility is $\beta_1(\bar{Q}/\bar{P})$ where $\bar{Q}$ and $\bar{P}$ are the estimated means.

To derive confidence intervals for these flexibilities, Miller, Capps, and Wells use simulations based on a normal error. DKS use Fieller's interval, the delta method, and three versions of the single bootstrap: percentile, bias-corrected percentile, and accelerated. DKS find that the three bootstraps all have approximately the same coverage and that the difference between nominal and actual coverage is greater for the bootstrap than for the other two methods. We use Shi's double bootstrap.

Only small differences between the single and double bootstrap intervals are anticipated for two reasons. First, DKS note that the degree of bias in Waugh's estimates is not appreciable (p. 1011). Because the double bootstrap makes this bias correction more readily than the single bootstrap, the difference between the single bootstrap and the double bootstrap is not likely to be appreciable. Second, the double bootstrap yields better correction for non-normality of the statistic. However, in this case, the various statistics are all approximately normally distributed. The single bootstrap distributions of the statistics are analyzed, and not one rejected the hypothesis of normality by the Jarque–Bera test. In summary, if the estimator is non-normal and its bias is non-negligible, the double bootstrap is expected to offer appreciably different intervals compared to the single bootstrap. Thus, the Waugh data provide a conservative test for the improvement of Shi's double bootstrap over the single bootstrap.

The model is

$$(9) \quad y = \beta X + \epsilon$$

where $E[\epsilon\epsilon'] = \sigma^2 \boldsymbol{I}$ with β, ϵ, and σ^2, respectively. The matrix of regressors $\boldsymbol{X}$ is fixed and the dependent variable comprises observations on a random variable Y with mean μ and finite variance. y is Waugh's P series, and X corresponds to the matrix formed by Waugh's constant, Q and M series. Thus, the true flexibility of Y with respect to the ith independent variable X_i is $\eta_{\mu i} = \beta_i\bar{X}_i/\mu$, which is estimated by $\hat{\eta}_{\mu,i} = b_i\bar{X}_i/\bar{Y}$ where $\bar{Y}$ is the mean of the vector $\boldsymbol{y}$ and $\bar{X}$ is the mean of the ith column of $\boldsymbol{X}$. For notational convenience, the μ,i notation is suppressed, and we consider the case of an individual flexibility.

The regression in equation (9) yields least squares estimates $\boldsymbol{b}$ and $\boldsymbol{e}$ of β and ϵ, respectively, and the flexibility estimate $\hat{\eta}$. Proceeding to the first-stage bootstrap, the residuals are scaled by $\sqrt{n(n-k)}$ where n is the number of observations and k is the number of columns of $\boldsymbol{X}$. On each first-stage iteration j a bootstrap replicate of the dependent variable is formed by

$$(10) \quad \boldsymbol{y}_j^* = \boldsymbol{Xb} + \boldsymbol{e}_j^*$$

where $\boldsymbol{e}_j^*$ is an n-vector of bootstrap residuals drawn from rescaled $\boldsymbol{e}$ by random uniform selection with replacement. The regression of $\boldsymbol{y}_j^*$ on $\boldsymbol{X}$ yields first-stage bootstrap estimates of the coefficients, $\boldsymbol{b}_j^*$, and the flexibility, $\hat{\eta}_j^*, j = 1, , 2, \ldots, J$.

On each first-stage iteration, K second-stage iterations are performed as follows. The second-stage bootstrap residuals $\boldsymbol{e}_{jk}^{**}$ are formed by random uniform selection with replacement from $\boldsymbol{e}_j^*$. The second-stage replicate of the dependent variable is $\boldsymbol{y}_{jk}^{**} = \boldsymbol{b}_j^*\boldsymbol{X} + \boldsymbol{e}_{jk}^{**}$. Regressing y_{jk}^{**} on $\boldsymbol{X}$ yields second-stage estimates of the coefficients, $\boldsymbol{b}_{jk}^{**}$, and the flexibility, $\hat{\eta}_{jk}^{**}$. On each first-stage iteration j one value of Q_j^* is computed as follows: $Q_j^* = \#(\hat{\eta}_{jk}^{**} \leq \hat{\eta})/K$, $k = 1, 2, \ldots, K$. Completion of all bootstrapping operations yields two j-vectors, bootstrap estimates $\hat{\boldsymbol{\eta}}^*$ and inner probability estimates $\boldsymbol{Q}^*$. The program is written in RATS v4.2 and takes approximately forty minutes on a 90-MHz Pentium.

Double bootstrap confidence intervals for Shi's method are obtained using the estimates $\hat{\boldsymbol{\eta}}^*$ and $\boldsymbol{Q}^*$. Sort $\hat{\boldsymbol{\eta}}^*$ in ascending order and let the ordered values be $\hat{\eta}_{(1)}, \hat{\eta}_{(2)}, \ldots, \hat{\eta}_{(J)}$. In the case of a 90% income flexibility for grapefruit, the percentile single lower limit is given by $\hat{\eta}^*_{[(J+1)0.05]} = \hat{\eta}^*_{(100)} = 0.2658$ and the upper limit is given by $\hat{\eta}^*_{[(J+1)0.95]} = \hat{\eta}^*_{(1,900)} = 0.78592$. $\boldsymbol{Q}^*$ can be used to obtain a more reliable con-

Table 1. Ninety Percent Bootstrap Confidence Intervals for Flexibilities

	Price		Income	
Commodity	Single	Double	Single	Double
Potatoes	[−3.94, −1.23]	[−3.98, −1.17]	[−0.11, 0.55]	[−0.13, 0.56]
Sweet potatoes	[−1.20, −0.36]	[−1.21, −0.33]	[−1.26, 0.02]	[−1.33, 0.07]
Grapefruit	[−1.19, −0.47]	[−1.18, −0.42]	[0.27, 0.79]	[0.29, 0.82]
Apples	[−1.09, −0.54]	[−1.10, −0.51]	[0.11, 0.54]	[0.10, 0.55]
Tomatoes	[−1.62, −0.10]	[−1.67, −0.04]	[0.16, 0.56]	[0.14, 0.58]
Beef	[−1.63, −1.26]	[−1.65, −1.26]	[1.11, 1.47]	[1.10, 1.48]

fidence interval from $\hat{\eta}$. Sort Q^* in ascending order: $Q^*_{(1)}, Q^*_{(2)}, \ldots, Q^*_{(J)}$. Because $J = 1{,}999$, $0.05(J + 1) = 100$ and $0.95(J + 1) = 1{,}900$, so $Q^*_{(100)} = 0.060$ and $Q^*_{(1,900)} = 0.972$. Thus, the lower limit will be $\hat{\eta}^*_{[(J+1)0.060]} = \hat{\eta}^*_{(120)} = 0.28505$ and the upper limit is $\hat{\eta}^*_{[(J+1)0.972]} = \hat{\eta}^*_{(1,944)} = 0.82158$. Similar calculations provide single and double bootstrap 90% confidence intervals for price and income flexibilities for the commodities studied by DKS, which are presented in table 1. The discrepancy between the single bootstrap intervals in table 1 and those of DKS can be attributed solely to the difference between their $J = 500$ and our $J = 1{,}999$. That the single bootstrap and double bootstraps are close confirms the DKS simulation result that the extent of the bias is negligible and our conclusion of the approximate normality of the statistics.

The effect of the second stage can vary. The single bootstrap interval can be left-shifted (potatoes/price) or right-shifted (grapefruit/income). In the case of sweet potato income flexibility, the interval is expanded by almost 10% of its length. The double bootstrap can produce different answers than the single bootstrap, and there is theoretical support that the double bootstrap is better. As a practical matter, a Monte Carlo study is employed to determine whether the double bootstrap really does produce more accurate confidence intervals.

Monte Carlo

Shi's double bootstrap modestly adjusts the single bootstrapped intervals reported by DKS. However, does this modest adjustment matter in a practical sense? Because DKS's Monte Carlo study is designed to answer the question, Does the interval achieve nominal coverage? their Monte Carlo design is used: 500 replications per experiment; two commodities, beef and tomatoes; and five error distributions, the standard normal, and beta with parameters (1.0, 1.0), (1.5, 1.5), (5.0, 5.0), and (1.5, 5.0) renormalized to have zero mean and unit variance. DKS chose these two commodities because tomatoes (beef) is most (least) likely to satisfy the Hayya, Armstrong, and Gressis conditions for the estimated flexibility (which is a ratio of random variables) to be distributed normally. This is a judicious choice, as demonstrated by the fact that while DKS find all three bootstrap varieties fail to achieve nominal coverage, the Fieller and Taylor intervals fail only for beef but achieved nominal coverage for tomatoes.

For each replication, nominal 90% single and double bootstrap intervals are computed, and for each experiment the number of times the calculated interval covers the true parameter is counted. With the double bootstrap program used in the previous section taking forty minutes to run, at 500 replications, each of the twenty experiments would take fourteen days because the regression command was invoked for each regression, thus inverting the $X'X$ matrix JK times. For the Monte Carlo analysis, a useful artifice decreases computational time when standard errors of the coefficients are not needed, as with Shi's double bootstrap. Instead of using the regression command JK times, one time we calculate $\boldsymbol{P} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'}$, form $\boldsymbol{y}^*_j$ and $\boldsymbol{y}^{**}_{jk}$ in the usual fashion, and form $\boldsymbol{b}^*_j$ and $\boldsymbol{b}^{**}_{jk}$ as $\boldsymbol{Py}^*_j$ and $\boldsymbol{Py}^{**}_{jk}$, respectively. This decreases the computational time from forty minutes to less than eight minutes, allowing the entire Monte Carlo study to be completed on a single computer in about eight weeks rather than forty. Because the time required to invert $\boldsymbol{X'X}$ increases proportionally to the square of the number of observations, the time savings will be even more dramatic for larger sample sizes.

DKS compare percentile, bias-corrected, and bias-corrected accelerated bootstraps. We rely on the DKS result that the three single

Table 2. Monte Carlo Results—% Coverage for Nominal 90% Intervals, 500 Replications

Error Distribution	Price		Income	
	Single	Double	Single	Double
		Tomatoes		
$N(0,1)$	85.2 (−3.58)*	88.0 (−1.49)	88.2 (−1.34)	89.2 (−0.60)
β(1.0,1.0)	87.2 (−2.08)*	89.4 (−0.45)	87.4 (−1.94)*	89.8 (−0.15)
β(1.5,1.5)	86.8 (−2.38)*	89.4 (−0.45)	87.8 (−1.65)*	89.0 (−0.75)
β(5.0,5.0)	87.2 (−2.08)*	89.2 (−0.50)	87.8 (−1.65)*	89.4 (−0.45)
β(1.5,5.0)	83.8 (−4.62)*	87.0 (−2.24)*	86.2 (−2.87)*	88.6 (−1.04)
		Beef		
$N(0,1)$	87.6 (−1.79)*	89.8 (−0.19)	86.8 (−2.39)*	91.0 (0.75)
β(1.0,1.0)	86.4 (−2.68)*	87.6 (−1.79)*	88.8 (−0.88)	90.8 (−0.60)
β(1.5,1.5)	87.0 (−2.24)*	90.6 (0.45)	87.8 (−2.39)*	90.4 (0.30)
β(5.0,5.0)	87.8 (−2.39)*	90.4 (0.30)	86.8 (−2.38)*	89.8 (−0.15)
β(1.5,5.0)	88.0 (−1.49)	90.8 (0.60)	87.8 (−1.65)*	90.4 (0.30)

Note: In parentheses, t-statistic for H_0: actual coverage = 0.90. Asterisk denotes rejection at 10%.

bootstraps all achieved approximately the same level of coverage. Thus, only the percentile bootstrap is compared to Shi's double bootstrap. Results are presented in table 2.

As in DKS, for each experiment the null-hypothesis that the true coverage equals the nominal coverage of 90% at the 10% significance level is tested with the t-statistics in parentheses, with significance denoted by an asterisk. The failure of the single bootstrap to achieve nominal coverage (H_0 rejected nine of ten times for tomatoes and eight of ten for beef) comports with the DKS study. For purposes of comparison, recall that DKS find that the Fieller and Taylor series achieved nominal coverage (each covering nine of ten times) for tomatoes, the case most favorable to the Hayya, Armstrong, and Gressis conditions, while each failed to cover in six of ten times for beef. Shi's double bootstrap, by contrast, covers nine of ten times for both tomatoes and beef.

Conclusions

Elasticity and flexibility estimation is crucial to empirical economic analysis. However, generating confidence intervals for elasticities and flexibilities can be problematic. A double bootstrap procedure is proposed as an improvement to interval generation. Shi's double bootstrapping method has superior coverage and consistency properties, as compared to single bootstrapping methods. Shi's method is demonstrated with an application to Waugh's classic study of demand, for which the bias is small and the sampling distribution of the parameters is approximately normal. Because the double bootstrap yields an improvement over the single bootstrap that is proportional to the degree of bias and non-normality, the use of the Waugh data implies a conservative assessment of the double bootstrap's performance. Shi's method modestly adjusts the single bootstrap intervals. A Monte Carlo study shows that Shi's double bootstrap achieves nominal coverage while the single bootstrap does not. Because pivoting transformations are frequently not available, Shi's method can both improve upon and extend confidence interval generation not only for elasticity and flexibility estimates but for any case when a pivot is not available and the usual double bootstrap cannot be applied.

[Received July 1996; accepted October 1997.]

References

Beran, R. "Prediction in Random Coefficient Regression." *J. Statistical Planning and Inference* 43(January 1995):205–13.

Beran, R. "Prepivoting to Reduce Level Error of Confidence Sets." *Biometrika* 74(September 1987):457–68.

Booth, J.G., and P. Hall. "Monte Carlo Approximation and the Iterated Bootstrap." *Biometrika* 81(June 1994):331–40.

Cooper, J.C. "A Comparison of Approaches to Calculating Confidence Intervals for Benefit Measures from Dichotomous Choice Contingent Valuation Surveys." *Land Econ.* 70(February 1994):111–22.

Davidson, R., and J.G. MacKinnon. "Bootstrap Tests: How Many Bootstraps?" Discussion Paper No. 951, Queen's University Institute for Economic Research, Ontario, 1997.

——. "Bootstrap Testing in Nonlinear Models." *Int. Econ. Rev.,* in press, 1999.

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

Dorfman, J., C. Kling, and R.J. Sexton. "Confidence Intervals for Elasticities and Flexibilities: Reevaluating the Ratio of Normals Case." *Amer. J. Agr. Econ.* 72(November 1990):1006–17.

Eakin, B., D. McMillen, and M. Buono. "Constructing Confidence Intervals Using the Bootstrap: An Application to a Multi-product Cost Function." *Rev. Econ. and Statist.* 72(1990):339–44.

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

Hall, P. "On the Bootstrap and Confidence Intervals." *Ann. Statist.* 14(August 1986): 1431–52.

——. *The Bootstrap and Edgeworth Expansion.* New York: Springer-Verlag, 1992.

Hayya, J., D. Armstrong, and N. Gressis. "A Note on the Ratio of Two Normally Distributed Variables." *Manage. Sci.* 21(1975):1338–41.

Kling, C. "Estimating the Precision of Welfare Measures." *J. Environ. Econ. and Manage.* 21(November 1991):244–51.

McCullough, B.D. "Bootstrapping Autoregressive Processes: An Application to AR(p) Models." *J. Forecasting* 13(January 1994): 51–66.

——. "Consistent Forecast Intervals when the Forecast-Period Exogenous Variables are Stochastic." *J. Forecasting* 15(July 1996): 293–304.

McCullough, B.D., and H.D. Vinod. "Implementing the Single Bootstrap: Some Computational Considerations." *Computational Econ.* 6(February 1993):1–15.

——. "Implementing the Double Bootstrap." *Computational Econ.* in press, 1998.

McCullough, B.D., and T. Waldon. "The Substitutability of Network and National Spot Television Advertising." *Quart. J. Econ. and Bus.* in press, 1998.

Miller, S.E., O. Capps, and G. Wells. "Confidence Intervals for Elasticities and Flexibilities from Linear Equations." *Amer. J. Agr. Econ.* 66(August 1984):392–96.

Prescott, D.M., and T. Stengos. "Bootstrapping Confidence Intervals: Application to Forecasting the Supply of Pork." *Amer. J. Agr. Econ.* 69(May 1987):266–73.

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

Shi, S. "Accurate and Efficient Double Bootstrap Confidence Limit Method." *Computational Statist. and Data Analysis* 13(1992):21–32.

Thombs, L., and W. Schucany. "Bootstrap Prediction Intervals for Autoregression." *J. Amer. Statist. Assoc.* 85(1990):486–92.

Veall, M.R. "Bootstrapping the Probability Distribution of Peak Electricity Demand." *Int. Econ. Rev.* 28(1987):302–12.

Vinod, H.D. "Bootstrap, Jackknife, and Resampling Methods: Applications in Econometrics." *Handbook of Statistics: Econometrics,* Vol. 11. G.S. Maddala, C.R. Rao, and H.D. Vinod, eds., pp. 629–661. New York: North-Holland, 1993.

——. "Double Bootstrap for Shrinkage Estimators." *J. Econometrics* 68(July 1995):287–302.

——. "Concave Consumption, Euler Equations and Inference with Estimating Functions." Proceedings of the 1997 ASA Meeting, Business and Economics Statistics Section, in press, 1998.

Vinod, H.D., and B.D. McCullough. "Estimating Cointegrating Parameters: An Application of the Double Bootstrap." *J. Statist. Planning and Inference* 43(January 1995a):147–56.

——. "Bootstrapping Supply and Demand Elasticities." *J. Asian Econ.* 5(October 1995b): 367–79.

Waugh, F.V. *Demand and Price Analysis: Some Estimates from Agriculture.* U.S. Department of Agriculture/Economic Research Service, Technical Bulletin 1316, 1964.

LINKED CITATIONS

Better Confidence Intervals: The Double Bootstrap with No Pivot
David Letson; B. D. McCullough
American Journal of Agricultural Economics, Vol. 80, No. 3. (Aug., 1998), pp. 552-559.
Stable URL:
http://links.jstor.org/sici?sici=0002-9092%28199808%2980%3A3%3C552%3ABCITDB%3E2.0.CO%3B2-5

[Footnotes]

[3] **On the Bootstrap and Confidence Intervals**
Peter Hall
The Annals of Statistics, Vol. 14, No. 4. (Dec., 1986), pp. 1431-1452.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28198612%2914%3A4%3C1431%3AOTBACI%3E2.0.CO%3B2-V

References

Prepivoting to Reduce Level Error of Confidence Sets
Rudolf Beran
Biometrika, Vol. 74, No. 3. (Sep., 1987), pp. 457-468.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28198709%2974%3A3%3C457%3APTRLEO%3E2.0.CO%3B2-0

Monte Carlo Approximation and the Iterated Boostrap
James G. Booth; Peter Hall
Biometrika, Vol. 81, No. 2. (Jun., 1994), pp. 331-340.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28199406%2981%3A2%3C331%3AMCAATI%3E2.0.CO%3B2-R

LINKED CITATIONS

A Comparison of Approaches to Calculating Confidence Intervals for Benefit Measures from Dichotomous Choice Contingent Valuation Surveys
Joseph C. Cooper
Land Economics, Vol. 70, No. 1. (Feb., 1994), pp. 111-122.
Stable URL:
http://links.jstor.org/sici?sici=0023-7639%28199402%2970%3A1%3C111%3AACOATC%3E2.0.CO%3B2-M

Confidence Intervals for Elasticities and Flexibilities: Reevaluating the Ratios of Normals Case
Jeffrey H. Dorfman; Catherine L. Kling; Richard J. Sexton
American Journal of Agricultural Economics, Vol. 72, No. 4. (Nov., 1990), pp. 1006-1017.
Stable URL:
http://links.jstor.org/sici?sici=0002-9092%28199011%2972%3A4%3C1006%3ACIFEAF%3E2.0.CO%3B2-5

Constructing Confidence Intervals Using the Bootstrap: An Application to a Multi-Product Cost Function
B. Kelly Eakin; Daniel P. McMillen; Mark J. Buono
The Review of Economics and Statistics, Vol. 72, No. 2. (May, 1990), pp. 339-344.
Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28199005%2972%3A2%3C339%3ACCIUTB%3E2.0.CO%3B2-V

On the Bootstrap and Confidence Intervals
Peter Hall
The Annals of Statistics, Vol. 14, No. 4. (Dec., 1986), pp. 1431-1452.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28198612%2914%3A4%3C1431%3AOTBACI%3E2.0.CO%3B2-V

A Note on the Ratio of Two Normally Distributed Variables
Jack Hayya; Donald Armstrong; Nicolas Gressis
Management Science, Vol. 21, No. 11, Theory Series. (Jul., 1975), pp. 1338-1341.
Stable URL:
http://links.jstor.org/sici?sici=0025-1909%28197507%2921%3A11%3C1338%3AANOTRO%3E2.0.CO%3B2-2

Confidence Intervals for Elasticities and Flexibilities from Linear Equations
Stephen E. Miller; Oral Capps, Jr.; Gary J. Wells
American Journal of Agricultural Economics, Vol. 66, No. 3. (Aug., 1984), pp. 392-396.
Stable URL:
http://links.jstor.org/sici?sici=0002-9092%28198408%2966%3A3%3C392%3ACIFEAF%3E2.0.CO%3B2-S

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

Bootstrapping Confidence Intervals: An Application to Forecasting the Supply of Pork
David M. Prescott; Thanasis Stengos
American Journal of Agricultural Economics, Vol. 69, No. 2. (May, 1987), pp. 266-273.
Stable URL:
http://links.jstor.org/sici?sici=0002-9092%28198705%2969%3A2%3C266%3ABCIAAT%3E2.0.CO%3B2-I

NOTE: *The reference numbering from the original has been maintained in this citation list.*