Monte Carlo, Bootstrap, and Jackknife Estimation

Assume that your true model is

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

where u is i.i.d. and $E(u|X) = 0$. Then the ordinary least squares (OLS) estimator of β is $\hat{\beta} = (X'X)^{-1}X'y$, which has desirable properties. You will want to estimate the variance of $\hat{\beta}$. An estimator of the $var(\hat{\beta}) = \sigma^2(X'X)^{-1}$ is

$$\hat{\sigma}^2(X'X)^{-1}, \tag{2}$$

where $\hat{\sigma}^2 = \hat{u}'\hat{u}/(T-K), \hat{u} = y - X\hat{\beta}, T =$ the number of observations, and $K =$ the number of regressors. To understand what is meant by the $var(\hat{\beta})$ and its estimator, consider the following Monte Carlo procedure. Keep in mind that you would never want to apply this procedure to the classical linear model in (1) for actual data, because you can easily evaluate (2).

1. MONTE CARLO ESTIMATION OF STANDARD ERRORS (= positive square root of the variances) OF $\hat{\beta}$:

   a. Assume a value for β, which is otherwise unobservable. You also select a matrix of values for X , which you hold constant over repeated trials.
   b. Draw u randomly with replacement from some distribution you assume to be correct, using a random number generator. This use of a random number generator yields the term "Monte Carlo," famed for its roulette wheels and games of chance.
   c. Compute y from equation (1).
   d. Estimate $\hat{\beta}$ by regressing your generated y on X obtaining

   $$\hat{\beta} = (X'X)^{-1}X'y. \tag{3}$$

   e. Repeat steps (a)-(d) many times (say 10,000). Note that you have generated 10,000 drawings of the random variable u and y through equation (1), from which we could compute 10,000 estimates of $\hat{\beta}$ using (3). Our estimate of the sample variance of $\hat{\beta}$ over 10,000 such outcomes is our sample measure of the population variance.
   g. The main use of the Monte Carlo method is to compute the bias and mean-square error of your estimator when it is difficult to do so analytically. However,

it is also useful to demonstrate omitted variable bias and other related models to econometrics students. Keep in mind that the assumptions made in the Monte Carlo method may make your results specific to your exact model.

2. BOOTSTRAP ESTIMATION OF THE STANDARD ERRORS (= positive square root of the variances) OF $\hat{\beta}$. Here you are confined to one actual data set and wish to resample from the empirical distribution of the residuals, rather than assume some distribution of the true error as with Monte Carlo analysis. Given (1) as your true model, you could easily evaluate (2) and not do any bootstrapping. With more complex models containing non-normal error terms and non-linearities in β, or in two-stage models where you need to correct the estimated standard errors of the second-stage estimators, the derivation of analytical formulas for the variance of $\hat{\beta}$ is complex. Also, the analytical formulas typically underestimate the true covariance with small samples. An alternative is to bootstrap these estimates (as in to pull oneself up by one's bootstraps). Theory and Monte Carlo evidence indicate that the bootstrap estimates are more accurate in small samples than the asymptotic formula. A nice summary is found in "Bootstrap Inference in Econometrics" by James MacKinnon, Dept. of Economics Working Paper, Queens Univ., June, 2002.

There are three basic bootstrap methods we will focus on: the naive bootstrap, pairs bootstrapping, and the wild bootstrap. These methods are most easily explained for the simple model (1). As with the Monte Carlo method, you assume that the model generating your data is the same as in (1). However, now you do not assume knowledge of β:

1. Naive Bootstrap

   a. Estimate $\hat{\beta} = (X'X)^{-1}X'y$.
   b. Compute $\hat{u} = y - X\hat{\beta}$. You work with $\hat{u}$ instead of assuming the distribution of u as in Monte Carlo estimation.
   c. Draw with replacement a sample of size T using a discrete uniform random number generator U[1,T], where T is your sample size. Let these random numbers be represented by $z_1, \ldots, z_T$. Generate element u_i^* as element z_i of $\hat{u}$, $i = 1, \ldots, T$. What this means is that each element of $\hat{u}$ has probability $1/T$ of being drawn. You can accomplish this more easily by using the TSP command RANDOM(DRAW=UHAT) USTAR, where UHAT is the residual, $\hat{u}_i$, and USTAR is u_i^*.
   d. Treating $y^* = X\hat{\beta} + u^*$ as your true model, compute y^*.

e. Compute $\beta^* = (X'X)^{-1}X'y^*$.
f. Repeat (c)-(e) many times (say 10,000). See MacKinnon for details.
g. Compute the square root of the sample variance of these β^* estimates. This is the estimate of the standard error of $\hat{\beta}$.

2. Pairs Bootstrap

a. Follow step a above.
b. Then draw pairs randomly with replacement, where the probability of any pair being drawn is equal to $1/T$, from $\{X, y\}$ to obtain $\{X^*, y^*\}$ using thse TSP commands:

MMAKE MATDAT Y1 X1 X2 X3 ;
RANDOM (DRAW=MATDAT) Y1STAR X1STAR X2STAR X3STAR

c. Then evaluate β^* with $\{X^*, y^*\}$ data and proceed as with the naive bootstrap.

Note that the pairs bootstrap produces the White robust covariance matrix.

3. Wild Bootstrap The wild bootsrap also produces this covariance matrix; see MacKinnon (2002) for details.

Also, note that for all three methods, $\hat{u} = Mu$, so that the residuals from OLS are biased downward. This means that you must correct the estimated variance by $T/(T-K)$. See Atkinson and Wilson, "The bias of bootstrapped versus conventional standard errors in the general linear and sur models" *Econometric Theory*, 8: 258-275 (1990).

3. BIAS ADJUSTMENT USING THE BOOTSTRAP OR JACKKNIFE.

In small samples many sandwich estimators may be biased. Weak instruments may also cause bias. We can correct for these biases using the bootstrap or the jackknife using the following methods, where B is the number of bootstrap replications:

1. Since the Bootstrap estimator of $\hat{\beta}$ is $1/B\sum_{i=1}^{B}\beta_i^*$, we can compute the bias correction for $\hat{\beta}$ as $\hat{\beta} - (1/B\sum_{i=1}^{B}\beta_i^* - \hat{\beta}) = 2\hat{\beta} - 1/B\sum_{i=1}^{B}\beta_i^*$. The intuition is that since we do not know β, we treat $\hat{\beta}$ as the "true" value and determine the bias of the bootstrap estimator relative to this value. We then adjust $\hat{\beta}$ by this computed bias, assuming that the bias of the bootstrap estimator relative to $\hat{\beta}$ is the same as the bias of $\hat{\beta}$ relative to β.

2. We can compute the jackknife estimator of the standard deviation of $\hat{\beta}$ for a sample of size $T, i = 1, \ldots, T$, by computing T jackknife estimates of β obtained by successively dropping observation i and recomputing $\beta_{J,i}$, where J stands for Jackknife. Then compute the standard deviation of the T estimates and multiply by $T-1$ to get the estimated standard deviation of $\hat{\beta}$. We can employ the jackknife two-stage-least-squares (JK2SLS) estimator of Hahn, J., and J. Hausman (2003), "Weak Instruments: Diagnosis and Cures in Empirical Econometrics," *American Economics Review Papers and Proceedings* 93: 118–125, to correct for the bias caused by weak instruments. The formula for the jackknife bias correction is given in Shao and Tu (1995). To compute the jackknife bias correction for the estimated coefficients, let $\hat{\beta}$ be the estimator of β for a sample of size T. First compute T jackknife estimates of $\hat{\beta}$ obtained by successively dropping one observation and recomputing $\hat{\beta}$. Call each of these T estimates $\beta_{J,i}$, $i = 1 \ldots, T$, and their average $\bar{\beta}_J = \sum_{i=1}^{T} \beta_{J,i}$. Define the jackknife bias estimator as

$$\text{BIAS}_J = (T-1)(\bar{\beta}_J - \hat{\beta}). \quad (1)$$

Then the jackknife bias-adjusted (BA) estimator of β is

$$\hat{\beta}_{BA} = \hat{\beta} - \text{BIAS}_J = T\hat{\beta} - (T-1)(\bar{\beta}_J). \quad (2)$$

Again, the intuition is that since we do not know β, we treat $\hat{\beta}$ as the "true" value and determine the bias of the jackknife estimator relative to this value. We then adjust $\hat{\beta}$ by this computed bias, assuming that the bias of the jacknife estimator relative to $\hat{\beta}$ is the same as the bias of $\hat{\beta}$ relative to β.