Short Guides to Microeconometrics
Fall 2020

Kurt Schmidheiny
Universität Basel

The Bootstrap

1 Introduction

The bootstrap is a method to derive properties (standard errors, confidence intervals and critical values) of the sampling distribution of estimators. It is very similar to Monte Carlo techniques (see the corresponding hand-out). However, instead of fully specifying the data generating process (DGP) we use information from the sample.

In short, the bootstrap takes the sample (the values of the independent and dependent variables) as the population and the estimates of the sample as true values. Instead of drawing from a specified distribution (such as the normal) by a random number generator, the bootstrap draws *with replacement* from the sample. It therefore takes the empirical distribution function (the step-function) as the true distribution function. In the example of a linear regression model, the sample provides the empirical distribution for the dependent variable, the independent variables and the error term as well as values for constant, slope and error variance. The great advantage compared to Monte Carlo methods is that we neither make assumption about the distributions nor about the true values of the parameters.

The bootstrap is typically used for consistent but biased estimators. In most cases we know the asymptotic properties of these estimators. So we could use asymptotic theory to derive the *approximate* sampling distribution. That is what we usually do when using, for example, maximum likelihood estimators. The bootstrap is an alternative way to produce *approximations* for the true small sample properties. So why (or when) would we use the bootstrap? There are two main reasons:

1a) *The asymptotic sampling distribution is very difficult to derive.*

1b) *The asymptotic sampling distribution is too difficult to derive for me.* This might apply to many multi-stage estimators. Example: the two stage estimator of the heckman sample selection model.

1c) *The asymptotic sampling distribution is too time-consuming and error-prone for me.* This might apply to forecasts or statistics that are (nonlinear) functions of the estimated model parameters. Example: elasticities calculated from slope coefficients.

2) *The bootstrap produces "better" approximations for some properties.* It can be shown that bootstrap approximations converge faster for certain statistics[1] than the approximations based on asymptotic theory. These bootstrap approximations are called asymptotic refinements. Example: the t-statistic of a mean or a slope coefficient.

Note that *both* asymptotic theory and the bootstrap only provide approximations for finite sample properties. The bootstrap produces consistent approximations for the sampling distribution for a variety of estimators such as the mean, median, the coefficients in OLS and most econometric models. However, there are estimators (e.g. the maximum) for which the bootstrap fails to produce consistent properties.

This handout covers the *nonparametric* bootstrap with *paired sampling*. This method is appropriate for randomly sampled cross-section data. Data from complex random samplings procedures (e.g. stratified sampling) require special attention. See the handout on "Clustering". Time-series data and panel data also require more sophisticated bootstrap techniques.

[1] These statistics are called *asymptotically pivotal*, i.e. there asymptotic distributions are independent of the data and of the true parameter values. This applies, for example, to all statistics with the standard normal or Chi-squared as limiting distribution.

2 The Method: Nonparametric Bootstrap

2.1 Bootstrap Samples

Consider a sample with $n = 1, ..., N$ independent observations of a dependent variable y and $K+1$ explanatory variables x. A *paired bootstrap sample* is obtained by independently drawing N pairs (x_i, y_i) from the observed sample *with replacement*. The bootstrap sample has the same number of observations, however some observations appear several times and others never. The bootstrap involves drawing a large number B of bootstrap samples. An individual bootstrap sample is denoted (x_b^*, y_b^*), where x_b^* is a $N \times (K+1)$ matrix and y_b^* an N-dimensional column vector of the data in the b-th bootstrap sample.

2.2 Bootstrap Standard Errors

The empirical standard deviation of a series of bootstrap replications of $\widehat{\theta}$ can be used to approximate the standard error $se(\widehat{\theta})$ of an estimator $\widehat{\theta}$.

1. Draw B independent bootstrap samples (x_b^*, y_b^*) of size N from (x, y). Usually $B = 100$ replications are sufficient.
2. Estimate the parameter θ of interest for each bootstrap sample:
$$\widehat{\theta}_b^* \quad \text{for} \quad b = 1, ..., B.$$
3. Estimate $se(\widehat{\theta})$ by
$$\widehat{se}(\widehat{\theta}) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\widehat{\theta}_b^* - \widehat{\theta}^*)^2}$$
where $\widehat{\theta}^* = \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}_b^*$.

The whole covariance matrix $V(\widehat{\theta})$ of a vector $\widehat{\theta}$ is estimated analogously.

In case the estimator $\widehat{\theta}$ is consistent and asymptotically normally distributed, bootstrap standard errors can be used to construct approximate confidence intervals and to perform asymptotic tests based on the normal distribution.

2.3 Confidence Intervals Based on Bootstrap Percentiles

We can construct a two-sided equal-tailed $(1-\alpha)$ confidence interval for an estimate $\widehat{\theta}$ from the empirical distribution function of a series of bootstrap replications. The $(\alpha/2)$ and the $(1-\alpha/2)$ empirical percentiles of the bootstrap replications are used as lower and upper confidence bounds. This procedure is called *percentile bootstrap*.

1. Draw B independent bootstrap samples (x_b^*, y_b^*) of size N from (x, y). It is recommended to use $B = 1000$ or more replications.
2. Estimate the parameter θ of interest for each bootstrap sample:
$$\widehat{\theta}_b^* \quad \text{for} \quad b = 1, ..., B.$$
3. Order the bootstrap replications of $\widehat{\theta}$ such that $\widehat{\theta}_1^* \leq ... \leq \widehat{\theta}_B^*$. The lower and upper confidence bounds are the $B \cdot \alpha/2$-th and $B \cdot (1-\alpha/2)$-th ordered elements, respectively. For $B = 1000$ and $\alpha = 5\%$ these are the 25th and 975th ordered elements. The estimated $(1-\alpha)$ confidence interval of $\widehat{\theta}$ is
$$[\widehat{\theta}^*_{B\cdot\alpha/2}, \widehat{\theta}^*_{B\cdot(1-\alpha/2)}].$$

Note that these confidence intervals are in general not symmetric.

2.4 Bootstrap Hypothesis Tests

The approximate confidence interval in section 2.3 can be used to perform an approximate two-sided test of a null hypothesis of the form $H_0 : \theta = \theta_0$. The null hypothesis is rejected on the significance level α if θ_0 lies outside the two-tailed $(1-\alpha)$ confidence interval.

2.5 The bootstrap-t

Assume that we have consistent estimates of $\widehat{\theta}$ and $\widehat{se}(\widehat{\theta})$ at hand and that the asymptotic distribution of the t-statistic is the standard normal

$$t = \frac{\widehat{\theta} - \theta_0}{\widehat{se}(\widehat{\theta})} \xrightarrow{d} N(0, 1).$$

Then we can calculate approximate critical values from percentiles of the empirical distribution of a series of bootstrap replications for the t-statistic.

1. Consistently estimate θ and $se(\widehat{\theta})$ using the observed sample:
$$\widehat{\theta},\ \widehat{se}(\widehat{\theta})$$
2. Draw B independent bootstrap samples (x_b^*, y_b^*) of size N from (x, y). It is recommended to use $B = 1000$ or more replications.
3. Estimate the t-value assuming $\theta_0 = \widehat{\theta}$ for each bootstrap sample:
$$t_b^* = \frac{\widehat{\theta}_b^* - \widehat{\theta}}{\widehat{se}_b^*(\widehat{\theta})} \quad \text{for} \quad b = 1, ..., B$$
where $\widehat{\theta}_b^*$ and $\widehat{se}_b^*(\widehat{\theta})$ are estimates of the parameter θ and its standard error using the bootstrap sample.
4. Order the bootstrap replications of t such that $t_1^* \leq ... \leq t_B^*$. The lower critical value and the upper critical values are then the $B \cdot \alpha/2$-th and $B \cdot (1 - \alpha/2)$-th elements, respectively. For $B = 1000$ and $\alpha = 5\%$ these are the 25th and 975th ordered elements.
$$t_{\alpha/2} = t^*_{B \cdot \alpha/2},\ t_{1-\alpha/2} = t^*_{B \cdot (1-\alpha/2)}$$

These critical values can now be used in otherwise usual t-tests for θ.

The above bootstrap lower $t_{B \cdot \alpha/2}$ and upper $t_{B \cdot (1-\alpha/2)}$ critical values generally differ in absolute values. Alternatively, we can estimate symmetric critical values by adapting step 4:

4. Order the bootstrap replications of t such that $|t_1^*| \leq ... \leq |t_B^*|$. The absolute critical value is then the the $B \cdot (1 - \alpha)$-th element. For $B = 1000$ and $\alpha = 5\%$ this is the 950th ordered element. The lower and upper critical values are, respectively:
$$t_{\alpha/2} = -|t^*_{B \cdot (1-\alpha)}|,\ t_{1-\alpha/2} = |t^*_{B \cdot (1-\alpha)}|$$

The symmetric bootstrap-t is the preferred method for bootstrap hypothesis testing as it makes use of the faster convergence of t-statistics relative to asymptotic approximations (i.e. critical values from the t- or standard normal tables).

The bootstrap-t procedure can also be used to create confidence intervals using bootstrap critical values instead of the ones from the standard normal tables:

$$[\widehat{\theta} + t_{\alpha/2} \cdot \widehat{se}(\widehat{\theta}), \widehat{\theta} + t_{1-\alpha/2} \cdot \widehat{se}(\widehat{\theta})]$$

The confidence interval from bootstrap-t is not necessarily better then the percentile method. However, it is consistent with bootstrap-t hypothesis testing.

3 Implementation in Stata 14.0

Stata has very conveniently implemented the bootstrap for cross-section data. Bootstrap sampling and summarizing the results is automatically done by Stata. The Stata commands are shown for the example of a univariate regression of a variable y on x.

Case 1: Bootstrap standard errors are implemented as option in the stata command

Many stata estimation commands such as `regress` have a built-in `vce` option to calculate bootstrap covariance estimates. For example

```
regress y x, vce(bootstrap, reps(100))
```

runs $B = 100$ bootstrap iterations of a linear regression and reports bootstrap standard errors along with confidence intervals and p-values based on the normal approximation and bootstrap standard errors. The postestimation command

```
regress y x, vce(bootstrap, reps(1000))
estat bootstrap, percentile
```

reports confidence bounds based on bootstrap percentiles rather than the normal approximation. Remember that it is recommended to use at least $B = 1000$ replications for bootstrap percentiles. The percentiles to be reported are defined with the confidence level option. For example, the 0.5% and 99.5% percentiles that create the 99% confidence interval are reported by

```
regress y x, vce(bootstrap, reps(1000)) level(99)
estat bootstrap, percentile
```

Case 2: The statistic of interest is returned by a single stata command

The command

```
bootstrap, reps(100): reg y x
```

runs $B = 100$ bootstrap iterations of a linear regression and reports bootstrap standard errors along with confidence intervals and p-values based on the normal approximation and bootstrap standard errors. The postestimation command `estat bootstrap` is used to report confidence intervals based on bootstrap percentiles from e.g. $B = 1000$ replications:

```
bootstrap, reps(1000): reg y x
estat bootstrap, percentile
```

We can select an specific statistic to be recorded in the bootstrap iterations. For example the slope coefficient only:

```
bootstrap _b[x], reps(100): reg y x
```


By default, Stata records the whole coefficient vector `_b`. Any value returned by a stata command (see `ereturn list`) can be selected.

We can also record functions of returned statistics. For example, the following commands create bootstrap critical values on the 5% significance level of the t-statistic for the slope coefficient:

```
reg y x
scalar b = _b[x]
bootstrap t=((_b[x]-b)/_se[x]), reps(1000): reg y x, level(95)
estat bootstrap, percentile
```

The respective *symmetric* critical values on the 5% significance level are calculated by

```
reg y x
scalar b = _b[x]
bootstrap t=abs((_b[x]-b)/_se[x]), reps(1000): reg y x, level(90)
estat bootstrap, percentile
```

We can save the bootstrap replications of the selected statistics in a normal stata `.dta` file to further investigate the bootstrap sampling distribution. For example,

```
bootstrap b=_b[x], reps(1000) saving(bs_b, replace): reg y x
use bs_b, replace
histogram b
```

shows the bootstrap histogram of the sampling distribution of the slope coefficient.

Note: it is important that all observations with *missing values* are dropped from the dataset before using the `bootstrap` command. Missing values will lead to different bootstrap sample sizes.

Case 3: The statistics of interest is calculated in a series of stata commands

The first task is to define a program that produces the statistic of interest for a single sample. This program might involve several estimation commands and intermediate results. For example, the following program calculates the t-statistic centered at $\widehat{\beta}$ in a regression of y on x

```
program tstat, rclass
   reg y x
   return scalar t = (_b[x]-b)/_se[x]
end
```

The last line of the program specifies the value that is investigated in the bootstrap: $(\widehat{\beta} - b)/\widehat{se}(\widehat{\beta})$ which will be returned under the name `t`. The definition of the program can be directly typed into the command window or is part of a do-file. The program should now be tested by typing

```
reg y x
scalar b = _b[x]
tstat
return list
```

The bootstrap is then performed by the Stata commands

```
reg y x
scalar b = _b[x]
bootstrap t=r(t), reps(1000): tstat
estat bootstrap, percentile
```

As in case 2, the bootstrap results can be saved and evaluated manually. For example,

```
reg y x
scalar b = _b[x]
bootstrap t=r(t), reps(1000) saving(bs_t): tstat
use bs_t, replace
centile t, centile(2.5, 97.5)
gen t_abs = abs(t)
centile t_abs, centile(95)
```

reports both asymmetric and symmetric critical values on the 5% significance level for t-tests on the slope coefficient.

4 See also ...

There is much more about the bootstrap than presented in this handout. Instead of paired resampling there is *residual resampling* which is often used in time-series context. There is also a *parametric bootstrap*. The bootstrap can also be used to reduce the small sample bias of an estimator by *bias corrections*. The m out of n bootstrap is used to overcome some bootstrap failures. A method very similar to the bootstrap is the *jackknife*.

References

Bradley Efron and Robert J. Tibshirani (1993), *An Introduction to the Bootstrap*, Boca Raton: Chapman & Hall. A fairly advanced but nicely and practically explained comprehensive text by the inventor of the bootstrap.

Brownstone, David and Robert Valetta (2001), The Bootstrap and Multiple Imputations: Harnessing Increased Computing Power for Improved Statistical Tests, *Journal of Economic Perspectives*, 15(4), 129-141. An intuitive pladoyer for the use of the bootstrap.

Cameron, A. C. and P. K. Trivedi (2005), Microeconometrics: Methods and Applications, Cambridge University Press. Sections 7.8 and chapter 11.

Horowitz, Joel L. (1999) The Bootstrap, In: Handbook of Econometrics, Vol. 5. This is a very advanced description of the (asymptotic) properties of the bootstrap.

Wooldridge, Jeffrey M. (2009), Introductory Econometrics: A Modern Approach, 4th ed. South-Western. Appendix 6A. A first glance at the bootstrap.