Lecture 8

Resampling

Lectures in SDPE: *Econometrics I* on February 18, 2024

Markus Jäntti
Swedish Institute For Social Research
Stockholm University

Introduction

- Statistical inference is traditionally based on *exact* or *asymptotic* approaches (and the latter overwhelmingly dominate modern econometrics, as exact inference requires exact distributional assumptions that are hard to defend).
- An alternative is to approximate unknown distributions using *resampling* out of the data, the idea being to let the sample represent the population from which it is drawn.
- There are advantages and disadvantages to their use:
  - widely applicable, powerful and can be more accurate than traditional methods;
  - computationally demanding and involve more complex theory.
- Two approaches to look at here:
  - the *jackknife* (leave-one-out; mostly for variance estimation);
  - the *bootstrap* (iid sampling with replacement from sample; variance estimation, confidence intervals and hypothesis testing).

- The most common use of the jackknife in econometrics is for estimating the variance of an estimator.
- Let θ be a parameter to be estimated and $\widehat{\theta}$ is an estimator of if. The "leave-one-out" estimator omitting i is $\widehat{\theta}_{(-i)}$ which has a mean

$$\overline{\theta} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\theta}_{(-i)}. \tag{1}$$

- The variance of the leave-one-out estimators can be directly used to estimate the variance of the estimator, but the suggested version is

$$\widehat{\boldsymbol{V}}_{\widehat{\theta}}^{\text{jack}} = \frac{n-1}{n}\sum_{i=1}^{n}\left(\widehat{\theta}_{(-i)} - \overline{\theta}\right)\left(\widehat{\theta}_{(-i)} - \overline{\theta}\right)'. \tag{2}$$

Jackknife estimation of variance

- The Tukey correction factor $(n-1)/n$ looks a little mysterious, so let's see why it is there. Consider the sample mean, $\bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i$. The leave-one-out estimator of the mean equals

$$\bar{Y}_{(-i)} = \frac{1}{n-1}\sum_{j\neq i} Y_j = \frac{n}{n-1}\bar{Y} - \frac{1}{n-1}Y_i \tag{3}$$

- The sample mean of the l-o-o estimators is

$$\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{(-i)} = \frac{n}{n-1}\bar{Y} - \frac{1}{n-1}\bar{Y} \tag{4}$$

so the deviation of a l-o-o estimator from the mean equals

$$\bar{Y}_{(-i)} - \bar{Y} = \frac{1}{n-1}\left(\bar{Y} - Y_i\right). \tag{5}$$

Jackknife estimation of variance

- A jackknife estimator of the variance is then

$$\begin{aligned} \widehat{\boldsymbol{V}}^{\text{jack}}_{\bar{Y}} &= \frac{n-1}{n}\sum_{i=1}^{n}\left(\frac{1}{n-1}\right)^2(\bar{Y}-Y_i)(\bar{Y}-Y_i)' \\ &= \frac{1}{n}\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(\bar{Y}-Y_i)(\bar{Y}-Y_i)' \end{aligned} \tag{6}$$

- The conventional (method-of-moments) estimator of the variance of the sample mean equals this times $(n-1)/n$.

Jackknife estimation of variance

- For least-squares regression, recall $\widetilde{e}_i = (1 - h_{ii})^{-1}\widehat{e}_i$ and $h_{ii} = \boldsymbol{X}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_i$.
- The leave-one-out estimator for i is found by noting that

$$\widehat{\beta}_{(-i)} = \widehat{\beta} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_i\widetilde{e}_i. \tag{7}$$

- Denoting $\widetilde{\mu} = n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i\widetilde{e}_i$, the sample mean of leave-one-out estimators is

$$\overline{\beta} = \widehat{\beta} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\widetilde{\mu} \tag{8}$$

so the deviation of an individual $\widehat{\beta}_{(-i)}$ from their mean is

$$\widehat{\beta}_{(-i)} - \overline{\beta} = -(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}_i\widetilde{e}_i - \widetilde{\mu}). \tag{9}$$

Jackknife estimation of variance

- Plugging this into eq 2 we get an expression that is closely related to the HC3 variance estimator that is based on prediction errors:

$$
\begin{aligned}
\widehat{\boldsymbol{V}}_{\widehat{\beta}}^{\text{jack}} &= \frac{n-1}{n}\sum_{i=1}^{n}\left(\widehat{\beta}_{(-i)}-\overline{\beta}\right)\left(\widehat{\beta}_{(-i)}-\overline{\beta}\right)' \\
&= \frac{n-1}{n}(\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\widetilde{e}_i^2-n\widetilde{\mu}\widetilde{\mu}'\right)(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \frac{n-1}{n}\widehat{\boldsymbol{V}}_{\widehat{\beta}}^{\text{HC3}}-(n-1)(\boldsymbol{X}'\boldsymbol{X})^{-1}\widetilde{\mu}\widetilde{\mu}'(\boldsymbol{X}'\boldsymbol{X})^{-1}.
\end{aligned} \tag{10}
$$

Jackknife estimation of variance

- The jackknife estimator can be used also for "smooth" functions of the estimator, $\theta = g(\beta)$. The leave-one-out estimator of $g()$, using a Taylor expansion, is

$$\widehat{\theta}_{(-i)} = g(\widehat{\beta}_{(-i)}) = g(\widehat{\beta} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_i\widetilde{e}_i) \simeq \widehat{\theta} - \widehat{\boldsymbol{G}}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_i\widetilde{e}_i. \quad (11)$$

- The estimator of the variance can be expressed as

$$\begin{aligned}\widehat{\boldsymbol{V}}^{\text{jack}}_{\widehat{\theta}} &= \frac{n-1}{n}\sum_{i=1}^{n}\left(\widehat{\theta}_{(-i)} - \overline{\theta}\right)\left(\widehat{\theta}_{(-i)} - \overline{\theta}\right)' \\ &= \frac{n-1}{n}\widehat{\boldsymbol{G}}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\widetilde{e}_i^2 - n\widetilde{\mu}\widetilde{\mu}'\right)(\boldsymbol{X}'\boldsymbol{X})^{-1}\widehat{\boldsymbol{G}} \quad (12) \\ &= \widehat{\boldsymbol{G}}'\widehat{\boldsymbol{V}}^{\text{jack}}_{\widehat{\beta}}\widehat{\boldsymbol{G}} \simeq \widehat{\boldsymbol{G}}'\widetilde{\boldsymbol{V}}_{\widehat{\beta}}\widehat{\boldsymbol{G}}.\end{aligned}$$

  The last formulation shows that the jackknife estimator approximates the estimator using asymptotic methods in section 4.16 (eq 4.34) of Hansen (2021).

- The restrictions used for testing hypotheses, which we denote as $\theta = r(\beta)$ are a special case of such "smooth" functions so replacing $g()$ with $r()$ and the (estimated) first derivative matrix $\widehat{\boldsymbol{G}}$ with $\widehat{\boldsymbol{R}}$ gives us the appropriate results.

The bootstrap algorithm and distribution

- The idea of the bootstrap is to let the empirical distribution F_n represent the unknown population distribution F.
- A bootstrap sample is randomly drawn from sample data (which has distribution F_n) *with replacement*; repeated a large number of times, this turn out to produce a good approximation to the distribution based on the true unknown distribution F.
- While variation is driven by the drawing with replacement, an individual observation has a reasonably strong probability of being included:

$$\mathbb{P}(\text{obs. included}) = 1 - \left(1 - \frac{1}{n}\right)^n \to 1 - e^{-1} =\simeq 0.632 \qquad (13)$$

  (The $n \to \infty$ works well even for small n.)
- The distribution of for *estimators* $\widehat{\theta}$ and test statistics is approximated by their distribution across the B bootstrap samples.
- Note that the bootstrap is an asymptotic method, with the asymptotics driven by *sample size n* (not number of bootstrap draws in the simulation, even if that matters also).

Definition of the bootstrap

- F is the unknown distribution of the k-dimensional random variable $(Y, \boldsymbol{X})$ for which we have a random sample $\{(Y_i, \boldsymbol{X}_i)\}_{i=1}^n$.
- The statistic of interest is a function of the sample values and of F

$$T_n = T_n((Y_1, \boldsymbol{X}_1), \ldots, (Y_n, \boldsymbol{X}_n), F). \tag{14}$$

  Examples are an estimator $\widehat{\theta}$ or a t-statistic, $(\widehat{\theta} - \theta)/s(\widehat{\theta})$.
- The true CDF of T_n, G, depends on F:

$$G_n(u, F) = \mathbb{P}(T_n \leq u | F) \tag{15}$$

- As F is unknown, we are in general unable to work out G_n. Asymptotic inference proceeds by approximating G_n by the limiting G when $n \longrightarrow \infty$.
- For this to work, $G(u, F) = G(u)$; i.e., G should not depend on F, in which case T_n is *asymptotically pivotal*.

- The bootstrap proceeds differently: in place of F, we use F_n *and* we simulate G by bootstrapping (repeatedly sampling with replacement from F_n) to obtain G_n^*.
- The *bootstrap distribution* on which *bootstrap inference* is based is then

$$G_n^*(u) \simeq G_n(u, F_n). \tag{16}$$

- The random variables with distribution F_n, $\{(Y_i^*, \boldsymbol{X}_i^*)\}_{i=1}^n$ are the *bootstrap data* and the *bootstrap statistic*

$$T_n^* = T_n((Y_1^*, \boldsymbol{X}_1^*), \ldots, (Y_n^*, \boldsymbol{X}_n^*), F_n) \tag{17}$$

has distribution G_n^*.

The empirical distribution function

- How should we estimate F_n?
- At every point $(y, \boldsymbol{x})$, the value of F is a population moment, and expectation of a function of the random variables:

$$F(y, \boldsymbol{x}) = \mathbb{P}(Y_i \leq y \cap \boldsymbol{X}_i \leq \boldsymbol{x}) = \mathrm{E}[1(Y_i \leq y)1(\boldsymbol{X}_i \leq \boldsymbol{x})] \qquad (18)$$

(1(.) is the indicator function, 1 when the condition in parentheses is true and 0 otherwise.)

- The "plugin estimator" of this is the estimator of the empirical distribution function (EDF), a step function with $n+1$ steps:

$$F_n(y, \boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} 1(Y_i \leq y)1(\boldsymbol{X}_i \leq \boldsymbol{x}). \qquad (19)$$

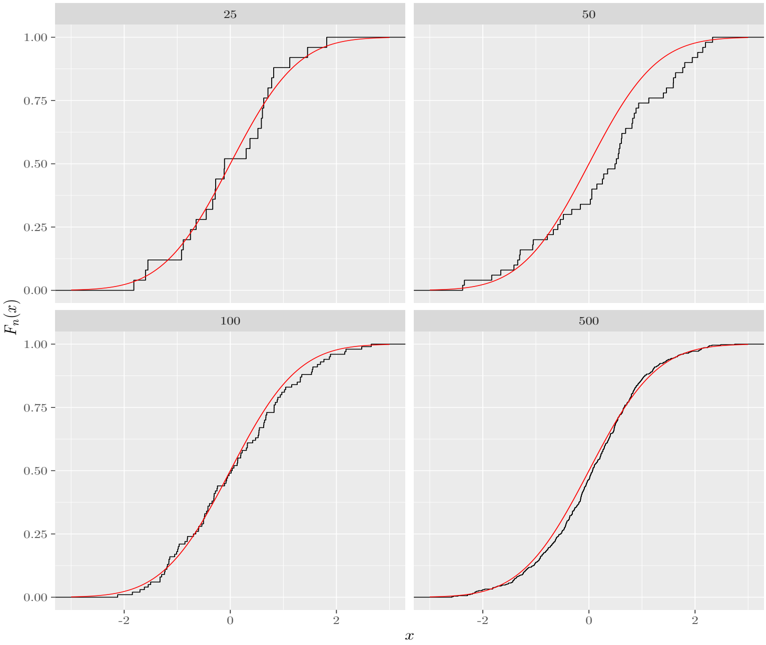
- F_n is consistent for F so by the weak law of large numbers

$$F_n(y, \boldsymbol{x}) \xrightarrow{p} F(y, \boldsymbol{x}) \qquad (20)$$

- By the central limit theorem,

$$\sqrt{n}(F_n(y, \boldsymbol{x}) - F(y, \boldsymbol{x})) \xrightarrow{d} N(0, F(y, \boldsymbol{x})(1 - F(y, \boldsymbol{x}))). \qquad (21)$$

The empirical distribution function (for different n)



The empirical distribution function

- The EDF is a discrete PD with probability $1/n$ at each $(Y_i, \boldsymbol{X}_i)$ pair.
- Denote each random pair $(Y_i^*, \boldsymbol{X}_i^*)$, with

$$\mathbb{P}(Y_i^* \leq y_i, \boldsymbol{X}_i^* \leq \boldsymbol{x}_i) = F_n(y_i, \boldsymbol{x}_i). \tag{22}$$

- We can treat this as a valid distribution, so the expectation of any function h equals the sample average:

$$\begin{aligned} \mathrm{E}[h(Y_i^*, \boldsymbol{X}_i^*)] &= \int_{\mathcal{Y},\mathcal{X}} h(y_i, \boldsymbol{x}_i) dF_n(y_i, \boldsymbol{x}_i) \\ &= \sum_{i=1}^{n} h(y_i, \boldsymbol{x}_i) \mathbb{P}(Y_i^* = y_i, \boldsymbol{X}_i^* = \boldsymbol{x}_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} h(y_i, \boldsymbol{x}_i). \end{aligned} \tag{23}$$

Nonparametric bootstrap

- A nonparametric bootstrap takes the EDF (eq. 19) in place of F in eq. 15.
- In principle, for fixed n, F_n is a multinomial random variable so G_n^* can be calculated directly, but there are $\binom{2n-1}{n}$ possible samples so it becomes computationally infeasible to do so.
- The alternative is to approximate the distribution using simulation;
  - each *bootstrap sample* should be of size n
  - the pairs $(Y_i^*, \boldsymbol{X}_i^*)$ are drawn randomly (with replacement) from the empirical distribution (equivalent for sampling $(Y_i, \boldsymbol{X}_i)$ from the original sample)
- The bootstrap statistic T_n^* (eq. 17) is calculated for each bootstrap sample, and this is repeated B times. B, the number of *bootstrap replications*, should be large (typically $B = 1,000$ or $B = 10,000$).
- T_n usually depends on F through some parameter – e.g., the t-ratio depends on θ, as $t = (\widehat{\theta} - \theta)/s(\widehat{\theta})$. Using θ_n replaces θ with the value associated with F_n (and is typically the point estimate, $\widehat{\theta}$).

Bootstrap asymptotics

- For bootstrap asymptotics, we have results that correspond to the conventional asymptotics, with the twist that they are *conditional* on the empirical distribution F_n.
- We have bootstrap convergence in probability, continuous mapping theorems, weak law of large numbers, central limit theorem and delta methods. (see Hansen 2021, sec 10.12)

Bootstrap estimation of variance

- $T_n = \widehat{\theta}$, so the variance and its bootstrap equivalent are

$$\mathrm{Var}[\widehat{\theta}] = V_\theta = \mathrm{E}[(T_n - \mathrm{E}[T_n])^2] \text{ with } V_\theta^* = \mathrm{E}[(T_n^* - \mathrm{E}[T_n^*])^2]. \quad (24)$$

- The bootstrapped estimate of this is

$$\widehat{V_\theta^*} = \frac{1}{B}\sum_{b=1}^{B}(\widehat{\theta}_b^* - \overline{\widehat{\theta}^*})^2, \quad (25)$$

  and its square root is the standard error.
- Often, the BS s.e. is taken to form the t-ratio of $\widehat{\theta}$. But the point of using the bootstrap is often that the asymptotic approximation to the normal is not good, so it is better to construct CI:s directly.
- Moreover, in more complex settings – e.g., estimating a ratio rather than, say, a mean – bootstrap variances can exhibit unwanted properties (i.e, not converge to the population moments they are supposed to be estimating) so trimming may be in order. Hansen (2021, p 276) gives as an example that the extreme 1 percent of each the bootstrap draws be discarded.

Bootstrap percentile intervals

- The (true) quantile function of $\widehat{\theta}$, $q(\alpha, F)$ is the solution to

$$G_n(q(\alpha, F), F) = \alpha. \tag{26}$$

  Its bootstrap counterpart is $q^*(\alpha, F_n)$, the quantile function of the bootstrap estimator $\widehat{\theta}^*$.
- In $(1-\alpha)$% of samples, $T_n = \widehat{\theta}$ lies in $[q_{\alpha/2}, q_{1-\alpha/2}]$ so the "plugin" confidence interval is based on the bootstrap quantile function

$$C^{\text{pc}} = [q^*_{\alpha/2}, q^*_{1-\alpha/2}]. \tag{27}$$

- Suppose we are interested in a monotonic increasing function m of θ. Then the confidence interval using this approach is $[m(q^*_{\alpha/2}), m(q^*_{1-\alpha/2})]$.

Coverage of BS percentile interval

- For the coverage of the BS percentile interval, we assume that

$$a_n(\widehat{\theta} - \theta) \xrightarrow{d} \xi \quad \text{and}$$

$$a_n(\widehat{\theta}^* - \widehat{\theta}) \xrightarrow{d} \xi.$$

- Denoting the quantiles of ξ by $\overline{q}_\alpha$, $a_n(q^*_\alpha - \widehat{\theta}) \xrightarrow{p} \overline{q}_\alpha$.
- Further, let the cdf of ξ be $H(x) = \mathbb{P}[\xi \le x]$. The coverage of the BS CI can be worked out as

$$\begin{aligned}
\mathbb{P}[\theta \in C^{\text{pc}}] =& \mathbb{P}[q^*_{\alpha/2} \le \theta \le q^*_{1-\alpha/2}] \\
=& \mathbb{P}[-a_n(q^*_{\alpha/2} - \widehat{\theta}) \ge a_n(\widehat{\theta} - \theta) \ge -a_n(q^*_{1-\alpha/2} - \widehat{\theta})] \\
\to& \mathbb{P}[-\overline{q}_{\alpha/2} \ge \xi \ge -\overline{q}_{1-\alpha/2}] \\
=& H(-\overline{q}_{\alpha/2}) - H(-\overline{q}_{1-\alpha/2}) \\
=& H(\overline{q}_{1-\alpha/2}) - H(\overline{q}_{\alpha/2}) \\
=& 1 - \alpha
\end{aligned} \tag{28}$$

This result hinges on *symmetry* of H, i.e., $H(-x) = 1 - H(x)$ (applied to moving lines 4 $\to$ 5).

Coverage of BS percentile interval

- The critical assumption above is that of symmetry of $H()$.
- For non-symmetric $H()$, the coverage is not $1-\alpha$. To set up one solution to that issue, consider an unknown but monotonic transformation $\psi()$ for which $\psi(\widehat{\theta}) - \psi(\theta)$ has a *pivotal*, symmetric (around zero) distribution $H(u)$ with quantiles $\overline{q}_\alpha$.
- As $H(u)$ is pivotal, $\psi(\widehat{\theta}^*) - \psi(\widehat{\theta})$ follows the same distribution.
- The coverage probability (closely following eq. 32) is

$$
\begin{aligned}
\mathbb{P}[\theta \in C^{\mathrm{pc}}] =& \mathbb{P}[q^*_{\alpha/2} \le \theta \le q^*_{1-\alpha/2}] \\
=& \mathbb{P}[\psi(q^*_{\alpha/2}) \le \psi(\theta) \le \psi(q^*_{1-\alpha/2})] \\
=& \mathbb{P}[\psi(\widehat{\theta}) - \psi(q^*_{\alpha/2}) \ge \psi(\widehat{\theta}) - \psi(\theta) \ge \psi(\widehat{\theta}) - \psi(q^*_{1-\alpha/2})] \\
\to& \mathbb{P}[-\overline{q}_{\alpha/2} \ge \psi(\widehat{\theta}) - \psi(\theta) \ge -\overline{q}_{1-\alpha/2}] \\
=& H(-\overline{q}_{\alpha/2}) - H(-\overline{q}_{1-\alpha/2}) \\
=& H(\overline{q}_{1-\alpha/2}) - H(\overline{q}_{\alpha/2}) \\
=& 1-\alpha
\end{aligned}
\tag{29}
$$

- The next to last row, again, hinges on $H()$ being symmetric around zero.
- The good news is that coverage is exact for any monotonic transformation; the bad news is the assumed symmetry.

Bias corrections to bootstrap quantiles

- There are several approaches to correcting the bias in BS percentile interval coverage.
- The bias-corrected (BC) is one, building on an unknown strictly increasing transformation $\psi(\theta)$ and an unknown constant z_0 for which

$$Z = \psi(\widehat{\theta}) - \psi(\theta) + z_0 \sim N(0, 1) \tag{30}$$

- The BC interval relies on the normal cdf $\Phi()$ and quantile function $\Phi^{-1}()$ as well as the bootstrap estimators $\widehat{\theta}^*_b$ and quantile function q^*_α, using

$$\begin{aligned} p^* =& \frac{1}{B}\sum_{b=1}^{B} \boldsymbol{I}[\widehat{\theta}^*_b < \widehat{\theta}] \\ z_0 =& \Phi^{-1}(p^*) \end{aligned} \tag{31}$$

  with p^* equal to median bias with z_0 to equal the quantile. With no bias, $p^* = .5$ and $z_0 = 0$.
- The adjustment is done by taking $x(\alpha) = \Phi(z_\alpha + 2z_0)$ and the BC interval as

$$C^{\mathrm{bc}} = [q^*_{x(\alpha/2)}, q^*_{x(1-\alpha/2)}]. \tag{32}$$

  i.e. the choice of quantiles is adjusted according to the bias.

Bias corrections to bootstrap quantiles

- To deduce the coverage probability of this (letting $\mathbb{P}^*$ denote "bootstrap probability"), use

$$\mathbb{P}[\psi(\widehat{\theta}) - \psi(\theta) + z_0 < x] = \Phi(x) \text{ and } \mathbb{P}^*[\psi(\widehat{\theta}^*) - \psi(\widehat{\theta}) + z_0 < x] = \Phi(x)$$
$$x(\alpha) = \mathbb{P}^*[\psi(\widehat{\theta}^*) - \psi(\widehat{\theta}) \leq z_\alpha + z_0] = \mathbb{P}^*[\widehat{\theta}^* \leq \psi^{-1}(\psi(\widehat{\theta}) + z_\alpha + z_0)]$$

- Then we can re-write C^{bc} as

$$C^{\mathrm{bc}} = [\psi^{-1}(\psi(\widehat{\theta}) + z_0 + z_{\alpha/2}), \psi^{-1}(\psi(\widehat{\theta}) + z_0 + z_{1-\alpha/2})] \tag{33}$$

- The coverage probability of this, in turn, is

$$\begin{aligned}
\mathbb{P}[\theta \in C^{\mathrm{bc}}] =& \mathbb{P}[\psi^{-1}(\psi(\widehat{\theta}) + z_0 + z_{\alpha/2}) \leq \theta \leq \psi^{-1}(\psi(\widehat{\theta}) + z_0 + z_{1-\alpha/2})] \\
=& \mathbb{P}[\psi(\widehat{\theta}) + z_0 + z_{\alpha/2} \leq \psi(\theta) \leq \psi(\widehat{\theta}) + z_0 + z_{1-\alpha/2}] \\
=& \mathbb{P}[-z_{\alpha/2} \geq \psi(\widehat{\theta}) - \psi(\theta) + z_0 \geq -z_{1-\alpha/2}] \\
=& \mathbb{P}[z_{1-\alpha/2} \geq Z \geq z_{\alpha/2}] \\
=& \Phi(z_{1-\alpha/2}) - \Phi(z_{\alpha/2}) = 1 - \alpha
\end{aligned} \tag{34}$$

- Further refinements exist, such as the "bootstrap accelerated" bias-corrected interval (BC_a). This involves a further correction based on skewness, which in turn needs to be estimated (e.g. by the jackknife).

Percentile-t interval and hypothesis test

- The t-ratio is

$$T = \frac{\widehat{\theta} - \theta}{s(\widehat{\theta})} \tag{35}$$

  and the bootstrap t-ratio is

$$T^* = \frac{\widehat{\theta}^* - \widehat{\theta}}{s(\widehat{\theta}^*)} \tag{36}$$

- The bootstrap percentile interval is based on the point estimate, standard error and the quantile function of T^*, q^*:

$$C^{\mathrm{pt}} = [\widehat{\theta} - s(\widehat{\theta})q^*_{1-\alpha/2}, \widehat{\theta} - s(\widehat{\theta})q^*_{\alpha/2})] \tag{37}$$

- A test of $\mathbb{H}_0 : \theta = \theta_0$ against $\mathbb{H}_1 : \theta \neq \theta_0$ is based on comparing $|T|$ to $q^*_{1-\alpha}$ (the quantile function of $|T^*|$). The bootstrap p-value $p^* = 1 - G^*_n(|T|)$ is estimated from

$$p^* = \frac{1}{B}\sum_{b=1}^{B} \boldsymbol{I}[|T^*_b| > |T|]. \tag{38}$$

Multivariate test statistics

- Wald based: With $\mathbb{H}_0 : \theta = \theta_0$ and two-sided alternative $\mathbb{H}_1 : \theta \neq \theta_0$, the asymptotically χ^2 Wald statistic is

$$W = n(\widehat{\theta} - \theta)' \widehat{\boldsymbol{V}_\theta}^{-1} (\widehat{\theta} - \theta), \tag{39}$$

the bootstrap replacement for which is

$$W^* = n(\widehat{\theta}^* - \widehat{\theta})' \widehat{\boldsymbol{V}_\theta^*}^{-1} (\widehat{\theta}^* - \widehat{\theta}) \tag{40}$$

- The bootstrap test rejects based on the test-statistic and bootstrap quantile function of W^* when $W \geq q^*(\alpha)$.
- Criterion based: With free and restricted criterion-based estimators $\widehat{\beta} = \arg\min_\beta J(\beta)$, we have $\widetilde{\beta} = \arg\min_{r(\beta)=\theta_0} J(\beta)$. A criterion-based test test for $\mathbb{H}_0 : r(\beta) = \theta_0$ is

$$J = J(\widetilde{\beta}) - J(\widehat{\beta}) \tag{41}$$

- The bootstrap test rejects if $J \geq q^*(\alpha)$, where q^* is the quantile function of J^* (where both $J(\widehat{\beta}^*)$ and $J(\widetilde{\beta}^*)$ are estimated in the bootstrap sample).
- Bootstrap p-values can be estimated based on the bootstrap distribution as

$$p^* = \frac{1}{B} \sum_{b=1}^{B} \boldsymbol{I}[W_b^* > W]. \tag{42}$$

- Suppose the conditional mean independence condition holds (so our estimates have a CEF interpretation) so

$$Y = \boldsymbol{X}'\beta + e,\ \mathrm{E}[e|\boldsymbol{X}] = 0. \tag{43}$$

- A non-parametric bootstrap might use the EDF to resample $(Y_i^*, \boldsymbol{X}_i^*)$ and impose

$$Y_i^* = \boldsymbol{X}_i^{*\prime}\beta + e_i^*,\ \mathrm{E}[e_i^* \boldsymbol{X}_i^*] = 0, \tag{44}$$

for which conditional mean independence does not, in general hold:

$$\mathrm{E}[e_i^*|\boldsymbol{X}_i^*] \neq 0. \tag{45}$$

- If the condition holds, estimation without imposing it is inefficient.

- One option is to *impose* independence in the bootstrapping, such as holding the $\boldsymbol{X}_i$ fixed and sampling e_i^* independently of them; then creating
$$Y_i^* = \boldsymbol{X}_i'\widehat{\beta} + e_i^* \tag{46}$$
and estimating the bootstrap $\widehat{\beta}^*$ using these new data.
- Imposing *independence* is overkill (and almost certainly not true); imposing *conditional mean independence* is less demanding, more accurate, but also more difficult than independence.

The *wild* bootstrap

- The wild bootstrap is one way to impose conditional mean independence rather than the stronger independence.
- The procedure creates a conditional distribution $e_i^*|\boldsymbol{X}_i$ such that

$$\begin{aligned} \mathrm{E}[e_i^*|\boldsymbol{X}_i] &= 0 \\ \mathrm{E}[e_i^{*2}|\boldsymbol{X}_i] &= \widehat{e_i}^2 \\ \mathrm{E}[e_i^{*3}|\boldsymbol{X}_i] &= \widehat{e_i}^3 \end{aligned} \tag{47}$$

- This can be achieved (somewhat surprisingly, perhaps) by sampling for each $\boldsymbol{X}_i$ from the two-point distribution

$$\mathbb{P}\left(e_i^* = \left(\frac{1 \pm \sqrt{5}}{2}\right)\widehat{e_i}\right) = \frac{\sqrt{5} \mp 1}{2\sqrt{5}} \tag{48}$$

Bootstrap methods for regression models – illustration

- use March 2015 CPS data to illustrate bootstrap sampling
- show BS distribution of both coefficient on female×education interaction, its t-value and and the error variance

Bootstrap methods for regression models – illustration

```
library(ggplot2)
load("cps2015.rda")
## table(cps$Gender <- factor(cps$female, labels=c("Man", "Woman")), cp
cps$oeducation <- cps$education
cps$education <- cps$eduyears
## focus on married non-white women and men
dim(td <- subset(cps, Married=="Married" & raceshort!="White" &
                  (experience>=11 & experience<=30)))
## set the number of bootstrap replications
B <- 1000
## get sample size (also the size of each BS sample)
n <- dim(td)[1]
## create, for simplicity, the dependent variable
td$y <- log(td$wage)
## the formula, to be reused in bootstrapping
fm <- formula(y~Gender*(education + experience + I(experience^2/100)))
## the original regression
lm.1 <- lm(fm, data=td)
## store fitted values in data.frame
td$x.beta.hat <- predict(lm.1)
## store the least squares residuals in data.frame
td$e.hat <- residuals(lm.1)
```

Bootstrap methods for regression models – illustration

```
## get the vector of point estimates
s2 <- summary(lm.1)$sigma^2
s2m <- var(subset(td, Gender=="Man")$e.hat)
s2w <- var(subset(td, Gender=="Woman")$e.hat)
b <- coef(lm.1)
theta.hat <- c(b, summary(lm.1)$coefficients[,2], summary(lm.1)$coeffic
names(theta.hat) <- c(names(b), paste("s", names(b), sep="."), paste("t
                      "s2", "s2m", "s2w", "ds2")
## setup the bootstrapping
## 1. Sample from pairs (y, x)
## 2. Sample e independently of x, generate y from fitted values plus e
## 3. Sample from the wild bootstrap
```

Bootstrap methods for regression models – illustration

```r
## 1. Sample from pairs (y, x)
bse1 <- function(x) {
    n <- dim(td)[1]
    btd <- td[sample(n, replace=TRUE),]
    lmb <- lm(fm, data=btd)
    btd$e.hat <- residuals(lmb)
    s2 <- var(btd$e.hat)
    s2m <- var(subset(btd, Gender=="Man")$e.hat)
    s2w <- var(subset(btd, Gender=="Woman")$e.hat)
    b <- coef(lmb)
    ret <- c(b, summary(lmb)$coefficients[,2], summary(lmb)$coefficient
    names(ret) <- c(names(b), paste("s", names(b), sep="."), paste("t",
                    "s2", "s2m", "s2w", "ds2")
    ret
    }
rmat1 <- t(sapply(1:B, bse1))
```

Bootstrap methods for regression models – illustration

```
## 2. Sample e independently of x, generate y from fitted values plus e
bse2 <- function(x) {
    n <- dim(td)[1]
    btd <- td
    e.hat <- btd[sample(n, replace=TRUE), "e.hat"]
    btd$y <- btd$x.beta.hat + e.hat
    lmb <- lm(fm, data=btd)
    btd$e.hat <- residuals(lmb)
    s2 <- var(btd$e.hat)
    s2m <- var(subset(btd, Gender=="Man")$e.hat)
    s2w <- var(subset(btd, Gender=="Woman")$e.hat)
    b <- coef(lmb)
    ret <- c(b, summary(lmb)$coefficients[,2], summary(lmb)$coefficient
    names(ret) <- c(names(b), paste("s", names(b), sep="."), paste("t",
                    "s2", "s2m", "s2w", "ds2")
    ret
    }
rmat2 <- t(sapply(1:B, bse2))
```

Bootstrap methods for regression models – illustration

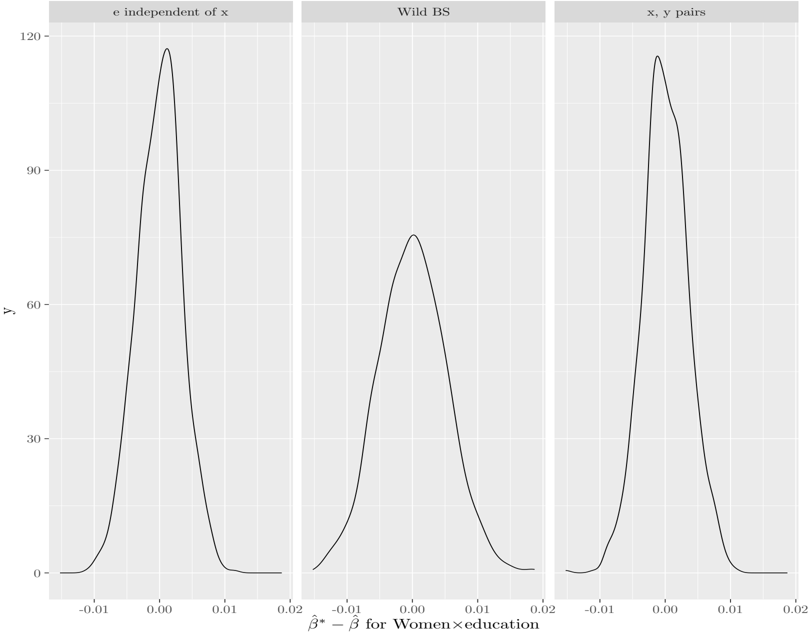
```
## 3. Sample from the wild bootstrap
bse3 <- function(x) {
    n <- dim(td)[1]
    btd <- td
    ## The LS residuals are in btd$e.hat
    ## create the two e.hats^* and sample
    ## from the two with known probabilities
    e.hat <- ifelse(runif(n)<((sqrt(5)-1)/(2*sqrt(5))),
                    ((1+sqrt(5))/2)*btd$e.hat,
                    ((1-sqrt(5))/2)*btd$e.hat)
    ## create the new y.hat
    btd$y <- btd$x.beta.hat + e.hat
    ## draw the BS sample
    btd <- btd[sample(n, replace=TRUE), ]
    lmb <- lm(fm, data=btd)
    btd$e.hat <- residuals(lmb)
    s2 <- var(btd$e.hat)
    s2m <- var(subset(btd, Gender=="Man")$e.hat)
    s2w <- var(subset(btd, Gender=="Woman")$e.hat)
    b <- coef(lmb)
    ret <- c(b, summary(lmb)$coefficients[,2], summary(lmb)$coefficient
    names(ret) <- c(names(b), paste("s", names(b), sep="."), paste("t",
```

Bootstrap methods for regression models – illustration

```r
## subtract the point estimates from each of the BS matrices
tm <- matrix(rep(theta.hat, B), nrow=B, byrow=TRUE)
drmat1 <- as.data.frame(rmat1-tm)
drmat2 <- as.data.frame(rmat2-tm)
drmat3 <- as.data.frame(rmat3-tm)
##
drmat1$method <- "x, y pairs"
drmat2$method <- "e independent of x"
drmat3$method <- "Wild BS"
btd <- rbind(drmat1, drmat2, drmat3)
## get rid of non-alphanumerics from the names
names(btd) <- gsub("\\W", "", names(btd))
```

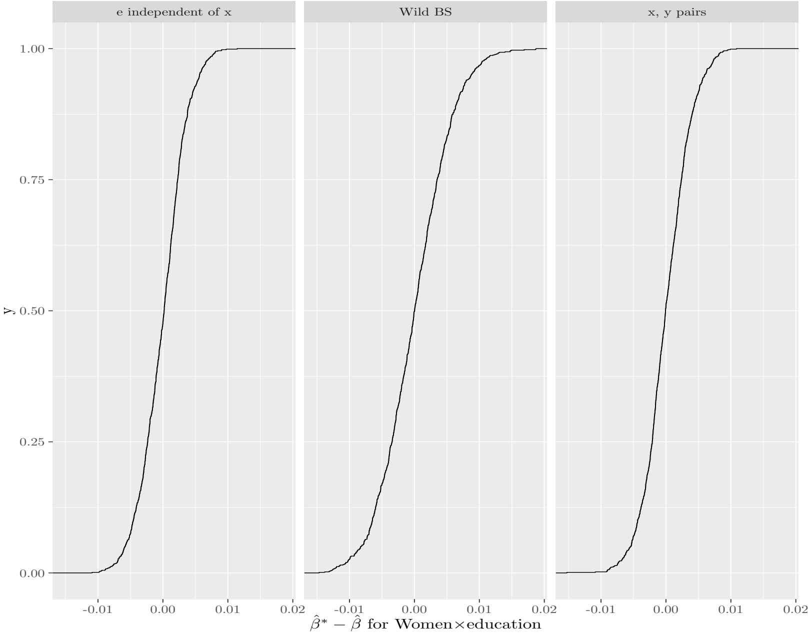
Bootstrap methods for regression models – illustration

Regression coefficient: Empirical density along with normal distribution



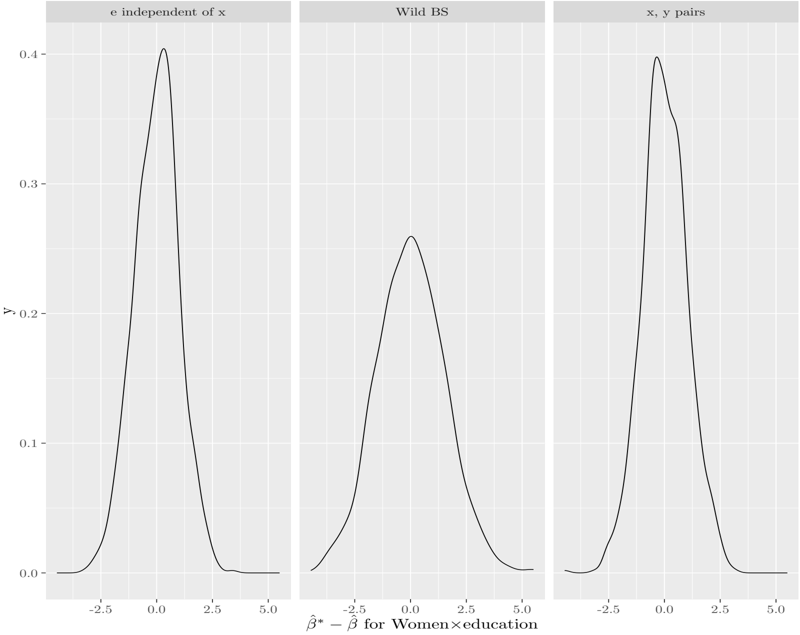
Bootstrap methods for regression models – illustration

Regression coefficient: Empirical CDF along with normal distribution



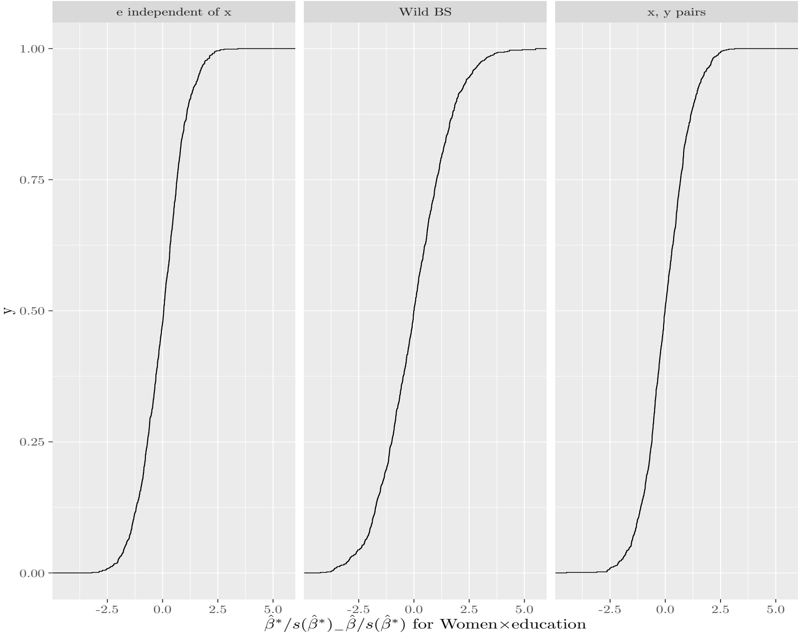
Bootstrap methods for regression models – illustration

t-statistic: Empirical density along with normal distribution



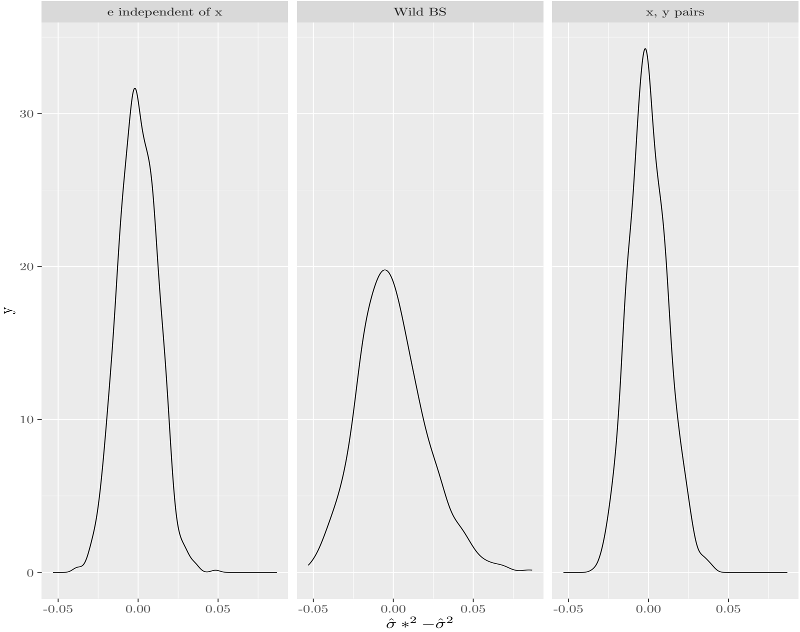
Bootstrap methods for regression models – illustration

t-statistic:Empirical CDF along with normal distribution



Bootstrap methods for regression models – illustration

Error variance: Empirical density along with normal distribution



Bootstrap methods for regression models – illustration

Confidence intervals for Woman×education interaction

```
##   -1*btd$GenderWomaneducation + 2*theta.hat["GenderWoman:education"]
## estimate C (theta.hat - q^*_n())
alpha <- .05
tx <-
 unlist(tapply(btd$GenderWomaneducation,
               list(btd$method),
               function(x) quantile(x, probs=c(1- alpha/2, alpha/2))))
ty <- rep(theta.hat["GenderWoman:education"], length(tx)) - tx
names(ty) <- names(tx)
ty

## e independent of x.97.5%  e independent of x.2.5%            Wild BS
##                   -0.0028                    0.0107                  -
##              Wild BS.2.5%         x, y pairs.97.5%          x, y pair
##                    0.0138                   -0.0034
```

Hansen, Bruce E (2021). *Econometrics*. Madison, WI: University of Wisconsin.