Bootstrap methods

A. Colin Cameron
U.C.-Davis

July, 2022

1. Introduction

- The bootstrap is a method for obtaining properties of statistics through resampling.
- There are many ways to construct bootstrap resamples.
- There are many uses of the bootstrap.
- The most common use of the bootstrap in econometrics is
  - to obtain standard errors of estimates.
- Occasionally use a more advanced bootstrap to potentially enable better finite sample inference
  - confidence intervals with better coverage
  - tests with true size closer to nominal size.

Summary

1. Introduction
2. Bootstrap (without asymptotic refinement)
3. Bootstrap in General
4. Bootstrap with asymptotic refinement
5. Wild Bootstrap
6. Stata commands

2. Bootstrap pairs estimate of standard error

- The most common bootstrap is the pairs bootstrap
  - views the sample as the $\{(y_1, \mathbf{x}_1), ..., (y_N, \mathbf{x}_N)\}$ as the population
  - assumes that $(y_i, \mathbf{x}_i)$ are i.i.d.
  - obtains B random samples from this population by resampling with replacement
    - ★ e.g. in bootstrap resample 1 observation may appear once, observation 2 not at all, observation 2 times, ...
- This yields B estimates $\widehat{\boldsymbol{\theta}}_1, ..., \widehat{\boldsymbol{\theta}}_B$.
  - so estimate $\mathrm{Var}[\widehat{\boldsymbol{\theta}}]$ using the usual variance of the B estimates.
- For scalar θ we have

$$\widehat{\mathrm{V}}[\widehat{\theta}] = \tfrac{1}{B-1} \sum\nolimits_{b=1}^{B} (\widehat{\theta}_b - \overline{\widehat{\theta}})^2, \quad \text{where } \overline{\widehat{\theta}} = \tfrac{1}{B} \sum\nolimits_{b=1}^{B} \widehat{\theta}_b.$$

  - Square root of this is called a bootstrap standard error.

Regression application

- Data: Doctor visits (count) and chronic conditions. $N = 50$.

```
. * Summmarize and Poisson with robust se's
. summarize

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      docvis |         50        4.12     7.82106          0         43
         age |         50       4.162    1.160382        2.6        6.2
     chronic |         50         .28    .4535574          0          1

. poisson docvis chronic, nolog vce(robust)

Poisson regression                              Number of obs     =         50
                                                Wald chi2(1)      =       3.64
                                                Prob > chi2       =     0.0565
Log pseudolikelihood = -238.75384               Pseudo R2         =     0.0917

------------------------------------------------------------------------------
             |               Robust
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     chronic |   .9833014   .5154894     1.91   0.056    -.0270391    1.993642
       _cons |   1.031602   .3446734     2.99   0.003     .3560541    1.707149
------------------------------------------------------------------------------
```

Bootstrap standard errors after Poisson regression

- Use option vce(boot)
  - Set the seed!
  - Set the number of bootstrap repetitions!

```
. * Compute bootstrap standard errors using option vce(bootstrap) to
. poisson docvis chronic, vce(boot, reps(400) seed(10101) nodots)

Poisson regression                              Number of obs     =         50
                                                Replications      =        400
                                                Wald chi2(1)      =       3.33
                                                Prob > chi2       =     0.0679
Log likelihood = -238.75384                     Pseudo R2         =     0.0917

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     chronic |   .9833014   .5386575     1.83   0.068    -.0724478    2.039051
       _cons |   1.031602   .3536507     2.92   0.004      .338459    1.724744
------------------------------------------------------------------------------
```

- Bootstrap se = 0.539 versus White robust se = 0.515.

Results vary with seed and number of reps

```
. * Bootstrap standard errors for different reps and seeds
. quietly poisson docvis chronic, vce(boot, reps(50) seed(10101))

. estimates store boot50

. quietly poisson docvis chronic, vce(boot, reps(50) seed(20202))

. estimates store boot50diff

. quietly poisson docvis chronic, vce(boot, reps(2000) seed(10101))

. estimates store boot2000

. quietly poisson docvis chronic, vce(robust)

. estimates store robust

. estimates table boot50 boot50diff boot2000 robust, b(%8.5f) se(%8.5f)
```

Variable	boot50	boot50~f	boot2000	robust
chronic	0.98330	0.98330	0.98330	0.98330
	0.45444	0.59923	0.54178	0.51549
_cons	1.03160	1.03160	1.03160	1.03160
	0.37131	0.37533	0.36414	0.34467

legend: b/se

Leading uses of bootstrap standard errors

- Sequential two-step m-estimator
  - First step gives $\widehat{\alpha}$ used to create a regressor $z(\widehat{\alpha})$
  - Second step regresses y on x and $z(\widehat{\alpha})$
  - Do a paired bootstrap resampling (x, y, z)
  - e.g. Heckman two-step estimator.
- Hausman test where under $H_0 : \widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}} \xrightarrow{p} 0$
  - use bootstrap to compute $\mathrm{V}[\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}] = \mathrm{Var}[\widehat{\boldsymbol{\theta}}] + \mathrm{V}[\widetilde{\boldsymbol{\theta}}] - 2\mathrm{Cov}[\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}]$
  - e.g. to test OLS versus IV
    - the simpler form of Hausman test assumes i.i.d. errors.
- Functions of other estimates e.g. $\widehat{\theta} = \widehat{\alpha} \times \widehat{\beta}$
  - replaces delta method
- Clustered data with many small clusters, such as short panels.
  -
    - Then resample the clusters.
    - But be careful if model includes cluster-specific fixed effects.
- For these in Stata need to use prefix command `bootstrap:`

3. The Bootstrap in General: Bootstrap algorithm

- A general bootstrap algorithm is as follows:
  - **1.** Given data $\mathbf{w}_1, ..., \mathbf{w}_N$
    - ★ draw a **bootstrap sample** of size N (see below for different ways)
    - ★ denote this new sample $\mathbf{w}_1^*, ..., \mathbf{w}_N^*$.
  - **2.** Calculate an **appropriate statistic** using the bootstrap sample. Examples include:
    - ★ (a) estimate $\widehat{\theta}^*$ of θ;
    - ★ (b) standard error $s_{\widehat{\theta}^*}$ of estimate $\widehat{\theta}^*$
    - ★ (c) $t-$statistic $t^* = (\widehat{\theta}^* - \widehat{\theta})/s_{\widehat{\theta}^*}$ centered at $\widehat{\theta}$.
  - **3.** Repeat steps 1-2 B independent times.
    - ★ Gives B bootstrap replications of $\widehat{\theta}_1^*, ..., \widehat{\theta}_B^*$ or $t_1^*, \ldots, t_B^*$ or .....
  - **4.** Use these B bootstrap replications to obtain a bootstrapped version of the statistic (see below).

Implementation

- Number of bootstraps: B high is best but increases computer time.
  - CT use 400 for se's and 999 for tests and confidence intervals.
  - Defaults are often too low. And set the seed!
- Various resampling methods
  - 1. Paired (or nonparametric or empirical dist. func.) is most common
    - ★ $\mathbf{w}_1^*, ..., \mathbf{w}_N^*$ obtained by sampling with replacement from $\mathbf{w}_1, ..., \mathbf{w}_N$.
  - 2. Parametric bootstrap for fully parametric models.
    - ★ Suppose $y|\mathbf{x} \sim F(\mathbf{x}, \boldsymbol{\theta}_0)$ and generate y_i^* by draws from $F(\mathbf{x}_i, \widehat{\boldsymbol{\theta}})$
  - 3. Residual bootstrap for regression with additive errors
    - ★ Resample fitted residuals $\widehat{u}_1, ..., \widehat{u}_N$ to get $(\widehat{u}_1^*, ..., \widehat{u}_N^*)$ and form new $(y_1^*, \mathbf{x}_1), ..., (y_N^*, \mathbf{x}_N)$.
  - 4. Moving blocks bootstrap
    - ★ for autocorrelated time series data
  - 5. Wild bootstrap
    - ★ for asymptotic refinement with heteroskedastic or clustered data.

- Need the underlying resampling to be i.i.d.
  - resample over clusters if data are clustered
    - But be careful if model includes cluster-specific fixed effects.
  - resample over moving blocks if data are serially correlated.

Bootstrap failure

- The bootstrap always provides estimates even when it makes no sense
  - e.g. can always get bootstrap standard errors for the mean of a Cauchy sample, even though the mean of the Cauchy does not exist.
- The following are cases where standard bootstraps fail
  - so need to adjust standard bootstraps.
- GMM (and empirical likelihood) in over-identified models
  - For overidentified models need to recenter or use empirical likelihood.
- Nonparametric Regression:
  - Nonparametric density and regression estimators converge at rate less than root-N and are asymptotically biased.
  - This complicates inference such as confidence intervals.
- Non-smooth estimators.

Jackknife

- The jackknife uses a leave-one-out resampling scheme.
- The jackknife estimate of the variance of an estimator $\widehat{\theta}$ is

$$\widehat{\mathrm{V}}[\widehat{\theta}] = \tfrac{N-1}{N}\sum_{i=1}^{N}(\widehat{\theta}_{(-i)} - \overline{\widehat{\theta}})^2, \quad \text{where } N^{-1}\sum_i \widehat{\boldsymbol{\theta}}_{(-i)}.$$

  - where $\widehat{\theta}_{(-i)}$ is $\widehat{\theta}$ obtained from the sample with observation i omitted.
- The jackknife is a "rough and ready" method for bias reduction in many situations, but not the ideal method in any.
  - it can be viewed as a linear approximation of the bootstrap (Efron and Tibsharani (1993, p.146)).
  - it requires less computation than the bootstrap in small samples, as then $N < B$ is likely
  - but it is outperformed by the bootstrap as $B \rightarrow \infty$.
- E.g. `poisson docvis chronic, vce(jackknife)`

4. Bootstrap Confidence Intervals (no refinement)

- **"Normal-based"** confidence intervals
  - 95% CI is $\widehat{\theta} \pm 1.96 \times se_{boot}(\widehat{\theta})$
  - asymptotically equivalent to $\widehat{\theta} \pm 1.96 \times se(\widehat{\theta})$
  - what Stata `vce(boot)` gives.
- **Percentile bootstrap** confidence intervals
  - 95% CI is $(\widehat{\theta}^*_{0.025}, \widehat{\theta}^*_{0.975})$
  - from the 2.5 to 97.5 percentiles of the bootstrap $\widehat{\theta}_b$, $b = 1, ..., B$
  - asymptotically equivalent to $\widehat{\theta} \pm 1.96 \times se(\widehat{\theta})$.
- Validity of bootstrap confidence intervals and tests requires convergence of the bootstrap distribution.
- Validity of bootstrap standard errors requires stronger uniform integrability conditions, because convergence in distribution does not imply convergence in moments.
- So percentile method requires weaker assumptions than the "normal-based" method.

Bootstrap Wald Test (no refinement)

- Consider test of $H_0 : \theta = \theta_0$ against $H_0 : \theta \neq \theta_0$ at level α.
- **"Normal-based"** Wald test
  - $t = (\widehat{\theta} - \theta_0)/se(\widehat{\theta})$
  - $p = \Pr[|t| > z_{1-\alpha/2}] = 2 \times (1 - \Phi(|t|)$
  - what Stata `vce(boot)` gives.
- **Percentile bootstrap** symmetric two-sided Wald test
  - $p = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\{|\widehat{\theta}_b^* - \widehat{\theta}| > |\widehat{\theta} - \theta|\}$
  - reject at level α if $p < \alpha$.

5. Bootstrap with asymptotic refinement

- The simplest bootstraps are no better than usual asymptotic theory
  - advantage is easy to implement, e.g. standard errors.
- More complicated bootstraps provide asymptotic refinement
  - this may provide a better finite-sample approximation.
- Several methods have for asymptotic refinement have been proposed
  - econometricians use the percentile-t method,

Asymptotic refinement (continued)

- Let T denote the Wald test t-ratio.
- Most conventional asymptotic tests
  - α = nominal size for a test, e.g. $\alpha = 0.05$.
  - actual size= $\alpha + O(N^{-1/2})$ for T or $\alpha + O(N^{-1})$ for $|T|$
  - e.g. see Hansen (2022), *Probability and Statistics*, p.186.
- Most tests with asymptotic refinement
  - actual size= $\alpha + O(N^{-1})$ for T or $\alpha + O(N^{-1/2})$ for $|T|$.
- Asymptotic bias of size $O(N^{-1}) < O(N^{-1/2})$ is smaller asymptotically.
  - but need simulation studies to confirm finite sample gains.
    - ★ e.g. if $N = 100$ then $100/N = O(N^{-1}) > 5/\sqrt{N} = O(N^{-1/2})$.

Asymptotically pivotal statistic and studentized t-statistic

- Econometricians rarely use asymptotic refinement.
- Asymptotic refinement bootstraps an asymptotically pivotal statistic
  - this means limit distribution does not depend on unknown parameters.
- An estimator $\widehat{\theta} - \theta_0 \overset{a}{\sim} \mathcal{N}[0, \sigma^2_{\widehat{\theta}}]$ is not asymptotically pivotal
  - since $\sigma^2_{\widehat{\theta}}$ is an unknown parameter.
- But the studentized $t-$statistic is asymptotically pivotal
  - since $t = (\widehat{\theta} - \theta_0)/s_{\widehat{\theta}} \overset{a}{\sim} \mathcal{N}[0, 1]$ has no unknown parameters.
- So bootstrap Wald test statistic to get tests and confidence intervals with asymptotic refinement.
- Formally this is an empirical way of implementing an Edgeworth expansion
  - a higher order expansion than the usual one used for asymptotic theory
  - analogous to going out extra terms in a Taylor series expansion.

Edgeworth Expansion

- Consider $Z_N = \sum_i X_i / \sqrt{N}$ where X_i are i.i.d. $[0, 1]$.
- The usual CLT leads to $Z_N \xrightarrow{d} N(0, 1)$. More precisely Z_N has c.d.f.

$$G_N(z) = \Pr[Z_N \leq z] = \Phi(z) + O(N^{-1/2}),$$

  - where $\Phi(\cdot)$ is the standard normal c.d.f.
- The CLT uses an approximation of $E[e^{isZ_N}]$, the **characteristic function** of Z_N, , where $i = \sqrt{-1}$.
- A better approximation expands $E[e^{isZ_N}]$ in powers of $N^{-1/2}$.
- The usual **Edgeworth Expansion** adds two additional terms, so

$$G_N(z) = \Pr[Z_N \leq z] = \Phi(z) + \frac{g_1(z)}{\sqrt{N}} + \frac{g_2(z)}{N} + O(N^{-3/2}),$$

  - where $g_1(z) = -(z^2 - 1)\phi(z)\kappa_3/6$
  - $\phi(\cdot)$ is the standard normal density
  - κ_3 is the third cumulant of Z_N
    - the 3^{rd} coefficient in the expansion $\ln(\mathrm{E}[e^{isZ_N}]) = \sum_{r=0}^{\infty} \kappa_r (is)^r / r!$
  - $g_2(\cdot)$ is given in Rothenberg (1984, p.895) or Amemiya (1985, p.93).

Bootstrap and Edgeworth Expansion

- We have

$$G_N(z) = \Pr[Z_N \leq z] = \Phi(z) + \frac{g_1(z)}{\sqrt{N}} + \frac{g_2(z)}{N} + O(N^{-3/2}),$$

- Using this directly is problematic as $g_1(z)$ depends on κ_3.
- Instead, the bootstrap for an asymptotically pivotal statistic can be shown to eliminate the term $g_1(z)/\sqrt{N}$
  - see P. Hall (1982), *The Bootstrap and Edgeworth Expansions*, Springer-Verlag.
  - or Cameron and Trivedi (2005), *Microeconometrics Methods and Applications*, pp.371-372.
  - or Hansen (2022), *Econometrics*, p.285.
- This leads to actual size$= \alpha + O(N^{-1})$.

Percentile-t pairs bootstrap

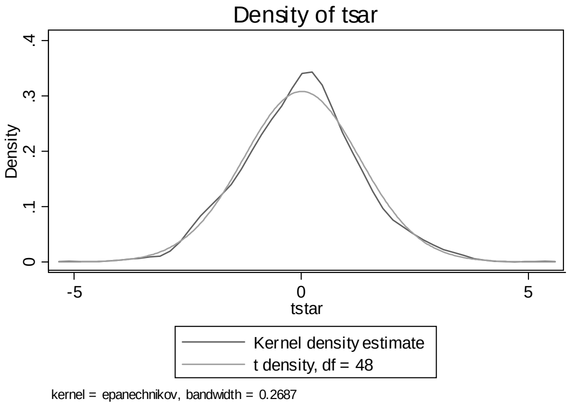
- Bootstrap $t = (\widehat{\theta} - \theta_0)/s_{\widehat{\theta}} \overset{a}{\sim} \mathcal{N}[0,1]$
  - by recomputing $t_b^* = (\widehat{\theta}_b - \widehat{\theta})/s_{\widehat{\theta}_b}$ where $\widehat{\theta}$ = original sample estimate
    - the original sample is now the population and the population $\theta = \widehat{\theta}$.

```
. * Percentile-t for a single coefficient: Bootstrap the t statistic
. use bootdata.dta, clear

. quietly poisson docvis chronic, vce(robust)

. local theta = _b[chronic]

. local setheta = _se[chronic]

. bootstrap tstar=((_b[chronic]-`theta')/_se[chronic]), seed(10101)        ///
>   reps(999) nodots saving(percentilet, replace): poisson docvis chronic, ///
>   vce(robust)
(note: file percentilet.dta not found)

Bootstrap results                               Number of obs      =        50
                                                Replications       =       999

      command:  poisson docvis chronic, vce(robust)
        tstar:  (_b[chronic]-.9833014421442415)/_se[chronic]
```

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
tstar	0	1.288018	0.00	1.000	-2.52447	2.52447

Percentile-t pairs bootstrap (continued)

- The 999 values of tstar $(= t_b^* = (\widehat{\theta}_b - \widehat{\theta})/s_{\widehat{\theta}_b})$ trace the bootstrap estimated density of the t-statistic .
- The plot is of the kernel density estimate and $T(48)$



Percentile-t Wald test

- Let t be the original full sample test statistic.
- For an **equal-tailed test**
  - p-value $=$ the proportion of times that $t < t_b^*$ or $t > t_b^*$, $b = 1, ..., B$.
- For a **symmetric two-sided test**
  - p-value $=$ the proportion of times that $|t_b^*| > |t|$, $t = 1, ..., B$,
  - where t is the original full sample test statistic.

Percentile-t Confidence Interval

- Let $\widehat{\theta}$ and $se(\widehat{\theta})$ be from the original full sample.
- Let $t^*_{0.025}$ and $t^*_{0.975}$ denote the 2.5 and 97.5 percentiles of t^*_b, $b = 1, ..., B$.
- For a **symmetric 95% confidence interval** we use
  - $\widehat{\theta} \pm |t^*|_{0.95} \times se(\widehat{\theta})$
- For an **equal-tailed 95% confidence interval** we use
  - $[\widehat{\theta} - t^*_{0.975} \times se(\widehat{\theta}), \widehat{\theta} - t^*_{0.025} \times se(\widehat{\theta})]$.
  - For explanation see Efron and Tibsharani (1993, 173-174) or Hansen (2022, 283-284).

Percentile-t Wald test

- Let t be $se(\widehat{\theta})$ be from the original full sample.
- For an equal-tailed 95% confidence interval test at 5% the critical t-values are the 2.5 and 97.5 percentiles of t^*.
- For a symmetric two-sided test the p-value is the proportion of times that $|t^*| > |t|$

BC and BCa confidence interval

- (N) is observed coefficient $\pm$ 1.96 $\times$ bootstrap s.e.
- (P) is 2.5 to 97.5 percentile of the bootstrap estimates $\widehat{\beta}_1^*, ..., \widehat{\beta}_B^*$.
- (BC bias-corrected) and (BCa) also have asymptotic refinement
  - not used in practice - instead use percentile-t method.

```
. * Bootstrap confidence intervals: normal-based, percentile, BC, and BCa
. quietly poisson docvis chronic, vce(boot, reps(999) seed(10101) bca)

. estat bootstrap, all

Poisson regression                              Number of obs      =        50
                                                Replications       =       999

------------------------------------------------------------------------------
             |    Observed               Bootstrap
      docvis |       Coef.       Bias    Std. Err.  [95% Conf. Interval]
-------------+----------------------------------------------------------------
     chronic |   .98330144   .0132307   .54137854    -.077781   2.044384   (N)
             |                                     -.0139438   2.061742   (P)
             |                                      -.079079   2.019438  (BC)
             |                                      .0295944    2.08349 (BCa)
       _cons |   1.0316016  -.0769223   .35685342    .3321817   1.731021   (N)
             |                                       .186586   1.582409   (P)
             |                                       .268264   1.641356  (BC)
             |                                       .386773   1.771351 (BCa)
------------------------------------------------------------------------------
(N)    normal confidence interval
(P)    percentile confidence interval
(BC)   bias-corrected confidence interval
(BCa)  bias-corrected and accelerated confidence interval
```

6. Percentile-t confidence intervals with Wild bootstrap

- The wild bootstrap was proposed for some non-i.i.d. models
  - C.F.J. Wu (1986), "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," Ann. Statist. 14(4), 1261-1295.
  - Regina Y. Liu (1988), "Bootstrap Procedures under some Non-I.I.D. Models," Ann. Statist., 16(4), 1696 - 1708.
- Initially used in econometrics for OLS regression with heteroskedastic errors
  - but few applications as usually N is reasonably large.
- Then used for clustered data
  - A. Colin Cameron, Douglas Miller and Jonah Gelbach (2008), "Bootstrap-Based Improvements for Inference with Clustered Errors," R.E.Stat, 90, 414-427
  - great need as asymptotics work poorly with few clusters
  - and does better than percentile-t with cluster pairs bootstrap.

Wild bootstrap for OLS and independent observations

- The wild bootstrap for linear regression
  - conditions on the sample value of the $\mathbf{x}'s$.
  - only y is resampled. $\mathbf{x}$ is not resampled.
- In original sample do restricted OLS of y_i on $\mathbf{x}_i$ where impose $H_0 : \beta = 0$ on the coefficient of interest
  - get residual $\widehat{u}_i = y_i - \mathbf{x}_i' \widehat{\boldsymbol{\beta}}$.
- In the b^{th} resample
  - set $y_{i,b} = \mathbf{x}_i' \widehat{\boldsymbol{\beta}} + \widehat{u}_i^*$
    where $\widehat{u}_i^* = a_g \widehat{u}_i$ and $\begin{cases} a_g = 1 & \text{with probability } 0.5 \\ a_g = -1 & \text{with probability } 0.5 \end{cases}$
  - do OLS regression using sample $(y_{1,b}, \mathbf{x}_1)$, ..., $(y_{N,b}, \mathbf{x}_N)$ gives $t_b^* = (\widehat{\beta}_b - \widehat{\beta}) / s_{\widehat{\beta}_b}$.
  - seems "wild" as $y_{i,b}$ can only take one of two values
  - but with N observations possibly as many as 2^N distinct samples.
- Gives an asymptotic refinement for OLS with heteroskedastic errors.

Wild bootstrap (continued)

- Not used much in practice for independent observations.
  - usually if N is low then estimates are statistically insignificant.
- But with clustered data and the number of clusters G is small
  - estimates may be highly statistically significant if many observations per cluster
  - yet tests have poor size
- In theory one could instead use a simpler pairs cluster bootstrap
  - where resample clusters $(\mathbf{y}_g, \mathbf{X}_g)$ with replacement
  - but this worked poorly in Monte Carlos.
- Instead do a Wild bootstrap where resample $\widehat{\mathbf{u}}_g$ over clusters.
  - i.e. $\mathbf{y}_{g,b} = \mathbf{X}_g\widehat{\boldsymbol{\beta}} + \widehat{\mathbf{u}}_g$ or $\mathbf{y}_{g,b} = \mathbf{X}_g\widehat{\boldsymbol{\beta}} - \widehat{\mathbf{u}}_g$ in cluster g and use the percentile-t method as before
  - important $\widehat{\mathbf{u}}_g$ imposes H_0.
  - Cameron, Gelbach and Miller (2008) proposed this
  - Webb (2017) proposed six-point resampling when $G < 10$.

Wild Score bootstraps

- Instead do a Wild bootstrap where resample over clusters.
  - Wild cluster bootstrap with weights a_g (e.g. $a_g = -1$ or 1)
  - $\widehat{\beta}^* = \widehat{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \sum_g \mathbf{X}_g (a_g \widehat{\mathbf{u}}_g)$ resamples **residuals** $\widehat{\mathbf{u}}_g$
  - $\quad = \widehat{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \sum_g (a_g \mathbf{X}_g \widehat{\mathbf{u}}_g)$ resamples **score** $\mathbf{X}_g \widehat{\mathbf{u}}_g$
- The latter generalizes to score of other estimators such as ML
  - Score bootstrap of Kline and Santos (2012).
- The Stata `boottest` command due to Roodman, MacKinnon, Nielsen and Webb (2018) implements
  - regular Wild and score Wild bootstraps
  - for independent and clustered data
  - for OLS, IV and nonlinear regression.

Wild score bootstrap example

- For the current Poisson example with independent observations
  - Default standard errors are too small giving $z = 3.6714$
  - But the method adjusts for this and yields $p = 0.0951$.

```
. * Wild score bootstrap for Poisson
. quietly poisson docvis chronic

. boottest chronic, seed(10101)

Re-running regression with null imposed.


Iteration 0:   log likelihood = -276.04852
Iteration 1:   log likelihood = -263.56766
Iteration 2:   log likelihood =  -262.8458
Iteration 3:   log likelihood = -262.84466
Iteration 4:   log likelihood = -262.84466

Poisson regression                              Number of obs     =         50
                                                Wald chi2(0)      =          .
Log likelihood = -262.84466                     Prob > chi2       =          .

 ( 1)  [docvis]chronic = 0
------------------------------------------------------------------------------
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     chronic |          0  (omitted)
       _cons |   1.415853   .0696733    20.32   0.000     1.279296     1.55241
------------------------------------------------------------------------------

Score bootstrap, null imposed, 999 replications, Wald test, Rademacher weights:
  chronic

                                    z =     3.6714
                             Prob>|z| =     0.0951
```

Wild score bootstrap example (continued)

- Repeat previous but use heteroskedastic-robust standard errors
  - heteroskedastic-robust standard errors yield smaller $z = 1.5280$
  - But the same $p = 0.0951$.

```
. * Note that with robust se's gives same p-value though different t-stat
. quietly poisson docvis chronic, vce(robust)

. boottest chronic, seed(10101)

Re-running regression with null imposed.


Iteration 0:   log likelihood = -276.04852
Iteration 1:   log likelihood = -263.56766
Iteration 2:   log likelihood =  -262.8458
Iteration 3:   log likelihood = -262.84466
Iteration 4:   log likelihood = -262.84466

Poisson regression                              Number of obs     =         50
                                                Wald chi2(0)      =          .
Log likelihood = -262.84466                     Prob > chi2       =          .

 ( 1)  [docvis]chronic = 0
------------------------------------------------------------------------------
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     chronic |          0  (omitted)
       _cons |   1.415853   .0696733    20.32   0.000     1.279296     1.55241
------------------------------------------------------------------------------

Score bootstrap, null imposed, 999 replications, Wald test, Rademacher weights:
  chronic

                             z =     1.5280
                      Prob>|z| =     0.0951
```

Wild score confidence intervals

- A better way that imposes a range of $H_0' s$
  - similar to procedure for AR confidence intervals with weak instruments.
- A confidence interval, or more generally a confidence region, can be obtained by inverting a test.
- Specifically, to obtain a 95% confidence set for a parameter θ we perform a two-sided test of $\theta = \theta_0$ for a range of values of θ_0.
- The confidence set is then those values of θ_0 for which the test has p > 0.05, since the 95% confidence interval includes those values that we do not reject at level 0.05.

7. Stata commands

- Most commands have option `vce(bootstrap)` and `vce(jackknife)`
- For more complicated bootstraps write a program and use `bootstrap:`
- For replicability set the seed!!
- For published work the more bootstraps the better as the seed becomes less important
- For small clusters use user-written `boottest` command.

8. References

- A. Colin Cameron and Pravin K. Trivedi (2005), Microeconometrics: Methods and Applications (MMA), chapter 11, Cambridge Univ. Press.
- A. Colin Cameron and Pravin K. Trivedi (2022), Microeconometrics using Stata (MUS), Second edition, Volume 1, chapter 12, Stata Press.
- Bruce Hansen (2022), Econometrics, chapter 10, Princeton University Press.
- A. Colin Cameron, Jonah Gelbach and Douglas L. Miller (2008), "Bootstrap-Based Improvements for Inference with Clustered Errors", Review of Economics and Statistics, August 2008, Vol. 90, 414-427.
- Patrick Kline and Andres Santos (2012), "A Score Based Approach to Wild Bootstrap Inference", Journal of Econometric Methods, 23-41.
- Davis Roodman, James MacKinnon, Morten Nielsen and Matthew Webb (2019),"Fast and Wild Bootstrap Inference in Stata using boottest," The Stata Journal [https://ideas.repec.org/p/qed/wpaper/1406.html#download]