Monday Part 2
Counts: Cross-section Inference

A. Colin Cameron
Univ. of Calif. - Davis

. .

Nonlinear Cross-section and Panel Regression Models for Count Data

. .

Queens University, Canada
Department of Economics

May 13-16, 2019

1. Introduction

- Count data models are for dependent variable $y = 0, 1, 2, \ldots$
- These slides focus on inference for Poisson quasi-MLE
  - heteroskedastic-robust standard errors
  - cluster-robust standard errors
  - bootstrap.

Outline

2. Standard Errors for OLS

- First consider **OLS in linear model**: $\widehat{\boldsymbol{\beta}} = \left(\sum_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \times \sum_i \mathbf{x}_i y_i$.
- Substitute in $y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i$ and simplify gives

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\sum_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \times \sum_i \mathbf{x}_i u_i.$$

- For simplicity assume that the $x_i's$ are fixed
  - if $\mathrm{E}[u_i] = 0$ then $\mathrm{E}[\widehat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$.
  - it follows that then

$$\begin{aligned} \mathrm{V}[\widehat{\boldsymbol{\beta}}] &= \mathrm{E}[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2] \\ &= \left(\sum_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \times \mathrm{V}\left[\sum_i \mathbf{x}_i u_i\right] \times \left(\sum_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1}. \end{aligned}$$

OLS Review (continued)

- If observations are independent over i then $\mathrm{V}[\sum_i \mathbf{x}_i u_i] = \sum_i \mathrm{V}[\mathbf{x}_i u_i]$ so

$$\mathrm{V}[\widehat{\boldsymbol{\beta}}] = \left(\sum_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \times \left[\sum_i \mathrm{E}[\mathbf{x}_i \mathbf{x}_i' u_i^2]\right] \times \left(\sum_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1}.$$

- Original approach: Assume a model for $\mathrm{E}[u_i^2|\mathbf{x}_i]$ and fit this model
  - e.g. $\mathrm{E}[u_i^2] = \exp(\mathbf{x}_i'\boldsymbol{\alpha})$ and use $\widehat{\mathrm{E}}[\mathbf{x}_i \mathbf{x}_i' u_i^2] = \mathbf{x}_i \mathbf{x}_i' \exp(\mathbf{x}_i'\widehat{\boldsymbol{\alpha}})$
- 1980's on: There is no need for such a model!
  - simply estimate $\sum_i \mathrm{E}[\mathbf{x}_i \mathbf{x}_i' u_i^2]$ by $\sum_i \mathbf{x}_i \mathbf{x}_i' \widehat{u}_i^2$ where $\widehat{u}_i = y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}$.
  - even though $\widehat{u}_i^2$ is not a good estimate of $\mathrm{E}[u_i^2]$
  - heteroskedastic-robust due to White (1982) and Huber (1967)

$$\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}] = \left(\sum_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \times \left[\sum_i \mathbf{x}_i \mathbf{x}_i' \widehat{u}_i^2\right] \times \left(\sum_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1}$$

- Can extend to
  - serially correlated errors (HAC robust)
  - clustered errors (cluster-robust).

3. Standard Errors for Poisson quasi-MLE

- Recall Poisson MLE solves $\sum_i (y_i - \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}))\mathbf{x}_i = \mathbf{0}$.
- Take a first-order Taylor series expansion of left-hand side about $\boldsymbol{\beta}$
  - so $\mathbf{g}(\widehat{\boldsymbol{\beta}}) \simeq \mathbf{g}(\boldsymbol{\beta}) + \mathbf{g}'(\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$

$$\sum_i (y_i - \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}))\mathbf{x}_i \simeq \sum_i (y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{x}_i - \sum_i \exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i\mathbf{x}_i'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

- Set r.h.s. to zero (note: we can ignore the remainder asymptotically)

$$\sum_i (y_i - \mu_i)\mathbf{x}_i + \left(\sum_i -\mu_i\mathbf{x}_i\mathbf{x}_i'\right)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \simeq \mathbf{0}.$$

- Solve for $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \simeq \left(\sum_i \mu_i\mathbf{x}_i\mathbf{x}_i'\right)^{-1} \times \sum_i (y_i - \mu_i)\mathbf{x}_i.$$

- This is like earlier OLS $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (\sum_i \mathbf{x}_i\mathbf{x}_i')^{-1} \times \sum_i \mathbf{x}_i u_i$
  - so proceed in the same way.

Poisson heteroskedastic-robust standard errors

- We have

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \simeq \left(\sum_i \mu_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \times \sum_i (y_i - \mu_i)\mathbf{x}_i.$$

- Then

$$\mathsf{V}[\widehat{\boldsymbol{\beta}}] \simeq \left(\sum_i \mu_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \times \mathsf{V}\left[\sum_i (y_i - \mu_i)\mathbf{x}_i\right] \times \left(\sum_i \mu_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1}.$$

- Given independence over i, $\mathsf{V}[\sum_i (y_i - \mu_i)\mathbf{x}_i] = \sum_i \mathsf{V}[(y_i - \mu_i)\mathbf{x}_i]$ so

$$\mathsf{V}[\widehat{\boldsymbol{\beta}}] \simeq \left(\sum_i \mu_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \times \left[\sum_i \mathsf{E}[(y_i - \mu_i)^2 \mathbf{x}_i \mathbf{x}_i']\right] \times \left(\sum_i \mu_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1}.$$

- And we use

$$\widehat{\mathsf{V}}_{\mathsf{HET}}[\widehat{\boldsymbol{\beta}}] = \left(\sum_i \mu_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \times \left[\sum_i (y_i - \widehat{\mu}_i)^2 \mathbf{x}_i \mathbf{x}_i'\right] \times \left(\sum_i \mu_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1}.$$

- Whereas MLE variance uses $\mathsf{E}[(y_i - \mu_i)^2 \mathbf{x}_i \mathbf{x}_i'] = \mu_i \mathbf{x}_i \mathbf{x}_i'$
and GLM variance uses $\mathsf{E}[(y_i - \mu_i)^2 \mathbf{x}_i \mathbf{x}_i'] = \alpha \times \mu_i \mathbf{x}_i \mathbf{x}_i'$.

Poisson cluster-robust standard errors

- Sometimes errors are correlated within cluster or group and independent across clusters
  - e.g. correlated if in same village and independent if in different villages.
- Let there be G such clusters, $g = 1, ..., G$. Then

$$\begin{aligned} \mathrm{V}\left[\sum_i (y_i - \mu_i)\mathbf{x}_i\right] &= \sum_{g=1}^G \mathrm{V}\left[\sum_{i \in g} (y_i - \mu_i)\mathbf{x}_i\right] \text{ as indep. over } g \\ &= \sum_{g=1}^G \left(\sum_{i \in g} \sum_{j \in g} \mathrm{E}[(y_i - \mu_i)\mathbf{x}_i (y_j - \mu_j)\mathbf{x}_j'] \right) \end{aligned}$$

- Then we use (provided $G \rightarrow \infty$)

$$\begin{aligned} \widehat{\mathrm{V}}_{\mathrm{CLU}}[\widehat{\beta}] &= \left(\sum_i \mu_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \\ &\quad \times \sum_{g=1}^G \left(\sum_{i \in g} \sum_{j \in g} (y_i - \widehat{\mu}_i)(y_j - \widehat{\mu}_j)\mathbf{x}_i \mathbf{x}_j' \right) \\ &\quad \times \left(\sum_i \mu_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1}. \end{aligned}$$

- Here little difference but in general cluster-robust standard errors can be much larger than heteroskedastic-robust.

```
. * Poisson with cluster robust standard errors - illustration
. * Here cluster on age for illustration
. * In practice the grouping variable would be, for example, village
. poisson docvis $xlist, vce(cluster age) nolog  // Poisson robust SEs

Poisson regression                              Number of obs     =      3,677
                                                Wald chi2(7)      =     917.07
                                                Prob > chi2       =     0.0000
Log pseudolikelihood =  -15019.64               Pseudo R2         =     0.1297

                                  (Std. Err. adjusted for 26 clusters in age)
```

docvis	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
private	.1422324	.0357205	3.98	0.000	.0722215	.2122434
medicaid	.0970005	.0653316	1.48	0.138	-.0310471	.2250481
age	.2936722	.0471694	6.23	0.000	.2012219	.3861226
age2	-.0019311	.0003162	-6.11	0.000	-.0025508	-.0013114
educyr	.0295562	.0054728	5.40	0.000	.0188296	.0402828
actlim	.1864213	.0374476	4.98	0.000	.1130254	.2598172
totchr	.2483898	.011554	21.50	0.000	.2257444	.2710353
_cons	-10.18221	1.749064	-5.82	0.000	-13.61031	-6.754105

4. Bootstrap estimate of standard error

- Basic idea is view $\{(y_1, \mathbf{x}_1), ..., (y_N, \mathbf{x}_N)\}$ as the population.
- Then obtain B random samples from this population
  - Get B estimates $\widehat{\theta}_1, ..., \widehat{\theta}_B$.
  - Then estimate Var$[\widehat{\theta}]$ as the usual standard deviation of B estimates

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

  - Square root of this is called a bootstrap standard error.
- Nonparametric bootstrap gets B different samples of size N we resample with replacement from $\{(y_1, \mathbf{x}_1), ..., (y_N, \mathbf{x}_N)\}$
  - In each bootstrap sample some original data points appear more than once while others not appear at all.
- **IMPORTANT: Stata 14 changed to a different random number generator (mt64) than earlier versions (kiss32).**
  - **These slides use the old Stata 13 generator.**
  - **To get my slide results using Stata 14 or 15:** `set rng kiss32`

Poisson regression application

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

```
Contains data from musbootdata.dta
  obs:            50
 vars:             3                          16 Apr 2010 10:32
 size:           750 (99.9% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
docvis          int    %8.0g                  number of doctor visits
age             float  %9.0g                  Age in years / 10
chronic         byte   %8.0g                  = 1 if a chronic condition
-------------------------------------------------------------------------------
Sorted by:

. 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
```

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)      =       3.50
                                                Prob > chi2       =     0.0612
Log likelihood = -238.75384                     Pseudo R2         =     0.0917

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     chronic |   .9833014   .5253149     1.87   0.061    -.0462968      2.0129
       _cons |   1.031602   .3497212     2.95   0.003     .3461607    1.717042
------------------------------------------------------------------------------
```

- Bootstrap se $= 0.525$ versus White heteroskedastic-robust se $= 0.515$.
- Note that if $B \rightarrow \infty$ the bootstrap se is asymptotically equivalent to White heteroskedastic-robust se!

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
	0.47010	0.50673	0.53479	0.51549
_cons	1.03160	1.03160	1.03160	1.03160
	0.39545	0.32575	0.34885	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.
- 2SLS estimator with heteroskedastic errors (if no White option)
  - Paired bootstrap gives heteroskedastic robust standard 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:`

The bootstrap: general 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)
    - ★ 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 $(y_1^*, \mathbf{x}_1), ..., (y_N^*, \mathbf{x}_N)$.
- Need to resample over i.i.d. observations
  - 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 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: e.g. LAD.

Jackknife

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

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

  - 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 is outperformed by the bootstrap as $B \rightarrow \infty$.

Clustered data

- The bootstrap relies on independence over the quantity being bootstrapped.
- So for clustered data we resample over clusters rather than observations
  - Let the g^{th} cluster be $\mathbf{y}_g = (y_{g1}, ..., y_{Ng})$ and similarly define $\mathbf{X}_g$.
  - Then view the G clusters $\{(\mathbf{y}_1, \mathbf{X}_1), ..., (\mathbf{y}_G, \mathbf{X}_G)\}$ as the population
  - And pick with replacement G clusters, etcetera.
  - e.g. `poisson y x, vce(boot, cluster(id_cluster))`
- Similarly for the jackknife use a delete-one-cluster jackknife

$$\widehat{\mathrm{V}}[\widehat{\theta}] = \frac{G-1}{G} \sum_{g=1}^{G} (\widehat{\theta}_{(-g)} - \overline{\widehat{\theta}})^2, \quad \text{where } \overline{\widehat{\theta}} = G^{-1} \sum_g \widehat{\boldsymbol{\theta}}_{(-g)}.$$

Cluster bootstrap: cluster on age for illustrative purposes

```
. poisson docvis chronic, vce(boot, cluster(age) reps(400) seed(10101) nodots)

Poisson regression                              Number of obs     =         50
                                                Replications      =        400
                                                Wald chi2(1)      =       4.12
                                                Prob > chi2       =     0.0423
Log likelihood = -238.75384                     Pseudo R2         =     0.0917

                                   (Replications based on 26 clusters in age)
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     chronic |   .9833014    .484145     2.03   0.042     .0343947    1.932208
       _cons |   1.031602    .303356     3.40   0.001     .4370348    1.626168
------------------------------------------------------------------------------
```

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.
- Conventional asymptotic tests (such as Wald test).
  - α = nominal size for a test, e.g. $\alpha = 0.05$.
  - Actual size= $\alpha + O(N^{-1/2})$.
- Tests with asymptotic refinement
  - Actual size= $\alpha + O(N^{-1})$.
  - 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

- 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 asymptotically refinement.
- For confidence intervals can also use BC (bias-corrected) and BCa methods.
- Econometricians rarely use asymptotic refinement.

```
. * 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  -.0244473   .54040762    -.075878   2.042481   (N)
             |                                       -.1316499   2.076792   (P)
             |                                       -.0820317   2.100361  (BC)
             |                                       -.0215526   2.181476 (BCa)
       _cons |   1.0316016  -.0503223   .35257252    .3405721   1.722631   (N)
             |                                        .2177235   1.598568   (P)
             |                                        .2578293   1.649789  (BC)
             |                                        .3794897   1.781907 (BCa)
------------------------------------------------------------------------------
(N)    normal confidence interval
(P)    percentile confidence interval
(BC)   bias-corrected confidence interval
(BCa)  bias-corrected and accelerated 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) and (BCa) have asymptotic refinement.

Percentile-t bootstrap

```
.
. * 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)

Bootstrap results                               Number of obs     =         50
                                                Replications      =        999

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

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       tstar |          0      1.3004     0.00   1.000    -2.548736    2.548736
------------------------------------------------------------------------------
```

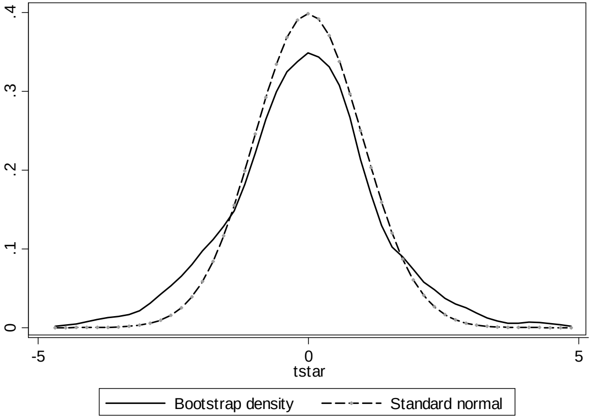
Bootstrap $t = (\widehat{\theta} - \theta_0)/s_{\widehat{\theta}} \overset{a}{\sim} \mathcal{N}[0,1]$

The 999 tstar values are the bootstrap estimated density of the t-statistic

Percentile-t bootstrap (continued)

Plot of the bootstrap estimated density of the t-statistic

tstar is $t_b = (\widehat{\theta}_b - \widehat{\theta}) / s_{\widehat{\theta}_b}$



Percentile-t bootstrap (continued)

Critical t-values are 2.5 and 97.5 percentiles

```
. use percentilet, clear
(bootstrap: poisson)

. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       tstar |       999   -.0874198      1.3004  -4.435354   4.611352

. centile tstar, c(2.5,50,97.5)

                                                       -- Binom. Interp. --
    Variable |     Obs  Percentile      Centile        [95% Conf. Interval]
-------------+-------------------------------------------------------------
       tstar |     999        2.5    -2.756196        -3.030972   -2.567785
             |                 50    -.0569957        -.1464812    .0312477
             |               97.5     2.568691           2.3092    2.917802

. kdensity tstar, generate(evalpoint densityest) xtitle("tstar from the bootstra
```

Wild bootstrap

- In practice bootstraps with asymptotic refinement are not often used in econometrics.
- Datasets are usually large and if small estimation is imprecise.
- But with clustered errors and few clusters it can be worthwhile to cluster bootstrap with asymptotic refinement
  - A. Colin Cameron and Douglas L. Miller, "A Practitioner's Guide to Cluster-Robust Inference", Journal of Human Resources, Spring 2015, Vol.50, No. 2, pp.317-373.
- The best bootstrap appears to be a Wild cluster bootstrap.

Wild bootstrap for OLS

- Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{u}_i$ be the original sample estimates.
- With independent data the Wild bootstrap resamples as follows
  - in b^{th} resample the i^{th} observation is $(y_i^*, \mathbf{x}_i)$ where
    - ★ $y_i^{*(b)} = \mathbf{x}_i'\widehat{\boldsymbol{\beta}} + \widehat{u}_i^{(b)}$ and
    - ★ $\widehat{u}_i^{(b)} = \widehat{u}_i$ with probability 0.5 and $\widehat{u}_i^{(b)} = -\widehat{u}_i$ with probability 0.5
    - ★ then $\widehat{\boldsymbol{\beta}}^{(b)}$ is OLS from regress $y_i^{*(b)}$ on $\mathbf{x}_i$.
- In cluster case in b^{th} resample for the g^{th} cluster is $(\mathbf{y}_i^*, \mathbf{X}_i)$ where
  - $\mathbf{y}_g^{*(b)} = \mathbf{X}_g\widehat{\boldsymbol{\beta}} + \widehat{\mathbf{u}}_g^{(b)}$ and
  - $\widehat{\mathbf{u}}_g^{(b)} = \widehat{\mathbf{u}}_g$ with probability 0.5 and $\widehat{\mathbf{u}}_g^{(b)} = -\widehat{\mathbf{u}}_g$ with probability 0.5.
- A variation that works better is to let $\widehat{\boldsymbol{\beta}}$ and $\widehat{u}_i$ (or $\widehat{\mathbf{u}}_g$) be original sample estimates that impose the null hypothesis restriction of interest
  - typically $H_0 : \beta_j = 0$
  - then $\widehat{\boldsymbol{\beta}}$ is from estimation dropping the j^{th} regressor.

Wild score bootstrap for Poisson

- For Poisson there is no residual
  - so instead resample the score (Kline and Santos)
  - recall $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \widehat{\mathbf{H}}^{-1}\widehat{\mathbf{g}}$.
- Then let
  - $\widehat{\mathbf{H}} = -\sum_{i=1}^{n} \mathbf{x}_i'\mathbf{x}_i$ be the Hessian from original estimation
  - $\widehat{\mathbf{g}} = \sum_{i=1}^{n} \widehat{\mathbf{g}}_i$ be the score or gradient vector
  - $\widehat{\mathbf{g}}_i = \sum_{i=1}^{n} \mathbf{x}_i(y_i - \exp(\widehat{\boldsymbol{\beta}}))$ be the contribution to the score
  - $\mathbf{g}^{(b)} = \widehat{\mathbf{g}}_i$ with probability 0.5 and $\mathbf{g}_i^{(b)} = -\widehat{\mathbf{g}}_i$ with probability 0.5
    - ★ more generally $\mathbf{g}_i^{(b)} = \widehat{\mathbf{g}}_i w_i$ where i.i.d. w_i have $E[w_i] = 0$ and $E[w_i^2] = 1$
  - $\widehat{\boldsymbol{\beta}}^{(b)} = \widehat{\boldsymbol{\beta}} + \widehat{\mathbf{H}}^{-1} \sum_{i=1}^{n} \mathbf{g}_i^{(b)}$.
- In cluster case repeat but at the cluster level.
- This can be implemented by the Stata add-on `boottest` command.