Lecture 6

Sampling errors for OLS regressions

1. review of "classic normal" case from Lecture 5

2. more general case - "robust" standard errors

3. even more general case - "clustered" standard errors

Recap: The population regression is:

$$y_i = x_i' \beta^* + u_i.$$

We assume:

1. independent sample of size N

2. no linear dependency in x_i

Classic normal case:

given the $x's$ each of the u_i are iid normals: $u_i \sim N(0,\sigma_u^2)$

Write:

$$
\begin{aligned}
\widehat{\beta} &= \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} x_i y_i \\
&= S_{xx}^{-1}\frac{1}{N}\sum_{i=1}^{N} x_i(x_i'\beta^* + u_i) \\
&= S_{xx}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)\beta^* + S_{xx}^{-1}\frac{1}{N}\sum_{i=1}^{N} x_i u_i \\
&= \beta^* + \sum_{i=1}^{N} a_i(X) u_i
\end{aligned}
$$

where $a_i(X) = \frac{1}{N}S_{xx}^{-1}x_i$. $\widehat{\beta} - \beta^*$ is a weighted sum of the $u_i's$ $\Rightarrow$ normal distribution, given X.

$$\widehat{\beta} - \beta^* = \sum_{i=1}^{N} a_i(X) u_i$$

$$E[\widehat{\beta} - \beta^* | X] = E[\sum_{i=1}^{N} a_i(X) u_i | X] = \sum_{i=1}^{N} a_i(X) E[u_i | X] = 0$$

so we know

$$E[\widehat{\beta} | X] = \beta^*$$

Now we want to get the *matrix of variances and covariances*, $V_\beta \equiv Var[\widehat{\beta} - \beta^*|X]$. This is a $K \times K$ matrix.

The (r,r) element is:

$$Var[\widehat{\beta}_r - \beta^*_r|X] = E[(\widehat{\beta}_r - \beta^*_r)^2|X]$$

The (r,s) element is:

$$Cov[\widehat{\beta}_r - \beta^*_r, \widehat{\beta}_s - \beta^*_s|X] = E[(\widehat{\beta}_r - \beta^*_r)(\widehat{\beta}_s - \beta^*_s)|X]$$

We know $\widehat{\beta} - \beta^* = \sum_{i=1}^{N} a_i u_i \Rightarrow \widehat{\beta}_r - \beta_r^* = \sum_{i=1}^{N} a_{ri} u_i$

AND all the $u_i' s$ are uncorrelated. So:

$$
\begin{aligned}
Var[(\widehat{\beta}_r - \beta_r^*)|X] &= \sum_{i=1}^{N} a_{ri}^2 \sigma_u^2 = \sigma_u^2 \sum_{i=1}^{N} a_{ri}^2 \\
Cov[\widehat{\beta}_r - \beta_r^*, \widehat{\beta}_s - \beta_s^*|X] &= \sum_{i=1}^{N} a_{ri} a_{si} \sigma_u^2 = \sigma_u^2 \sum_{i=1}^{N} a_{ri} a_{si} \\
\Rightarrow Var[\widehat{\beta} - \beta^*|X] &= \sigma_u^2 \sum_{i=1}^{N} a_i a_i'
\end{aligned}
$$

So we've shown

$$Var[\widehat{\beta}_r - \beta_r^*|X] = \sigma_u^2 \sum_{i=1}^{N} a_i a_i'$$

and using $a_i = \frac{1}{N} S_{xx}^{-1} x_i$, $a_i' = \frac{1}{N} x_i' S_{xx}^{-1}$ (since $S_{xx}$and $S_{xx}^{-1}$are symm.)

$$\begin{aligned} Var[\widehat{\beta}_r - \beta_r^*|X] &= \sigma_u^2 \sum_{i=1}^{N} \frac{1}{N^2} S_{xx}^{-1} x_i x_i' S_{xx}^{-1} \\ &= \frac{1}{N} \sigma_u^2 S_{xx}^{-1} \end{aligned}$$

In standard regression packages, the reported matrix of samp. errors is:

$$\widehat{V}_\beta = \frac{1}{N} \widehat{\sigma}_u^2 S_{xx}^{-1}$$

There are 2 limitations of the "classic normal" model.

1. The classic case says that $E[y_i|x_i] = x_i'\beta^*$. When that is not true, the error for the i^{th} observation is

$$\begin{aligned} u_i &= y_i - E[y_i|x_i] + E[y_i|x_i] - x_i'\beta^* \\ &= \varepsilon_i + v_i \end{aligned}$$

and $v_i \equiv E[y_i|x_i] - x_i'\beta^*$ (the specification error) depends on x_i.

2. In lots of cases $u_i \sim Normal$ is not appropriate (e.g., $y_i = 1[z_i > 0]$ discrete)

So in this lecture we "tech up" our sampling errors!

We need 4 results from statistics. Suppose we have an iid sample of size N; A_N is $K \times K$ matrix of sample statistics with property $plim\, A_N = A$

1. If A is invertible then $plim\,(A_N)^{-1} = (plim\, A_N)^{-1} = A^{-1}$

2. If b_N is $K \times 1$ vector with $plim\, b_N = 0$, then $plim\, A_N b_N = 0$.

3. (vector CLT). If z_i i.i.d $(K \times 1)$ vector with $E[z_i] = 0$, $Var[z_i] = V$ then:

$$\sqrt{N}\bar{z}_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} z_i \xrightarrow{a} N(0, V)$$

4. If $\sqrt{N} b_N \xrightarrow{a} N(0, V)$ then $\quad \sqrt{N} A_N b_N \xrightarrow{a} N(0, AVA')$

Using these results:

$$
\begin{aligned}
\widehat{\beta} &= \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} x_i y_i \\
&= S_{xx}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)\beta^* + S_{xx}^{-1}\frac{1}{N}\sum_{i=1}^{N} x_i u_i \\
&= \beta^* + S_{xx}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i u_i\right)
\end{aligned}
$$

$$
\Rightarrow \widehat{\beta} - \beta^* = S_{xx}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i u_i\right)
$$

$$\Rightarrow \widehat{\beta} - \beta^* = S_{xx}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i u_i\right)$$

Let $z_i = x_i u_i$. We know $E[z_i] = E[x_i u_i] = 0$ (FOC for β^*).

Let $E[(z_i - E[z_i])(z_i - E[z_i])'] = V$ (the variance of the r.v. z_i)

We know from *w.l.l.n* that $plim\,\frac{1}{N}\sum_{i=1}^{N} z_i = E[z_i] = 0$

We know from vector CLT that $\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} z_i\right) \xrightarrow{a} N(0, V)$

Assume: $plim\, S_{xx} = plim\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right) = S_{xx}^*$ exists and is invertible

and $Var[x_i u_i] = V$

Then

$$plim\, \widehat{\beta} - \beta^* = 0 \tag{1}$$

$$\sqrt{N}(\widehat{\beta} - \beta^*) \xrightarrow{a} N(0, [S_{xx}^*]^{-1}\, V\, [S_{xx}^*]^{-1}) \tag{2}$$

How to prove (1): $plim\,\widehat{\beta} - \beta^* = 0$?

$$
\begin{aligned}
plim\,\widehat{\beta} - \beta^* &= plim\, S_{xx}^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} x_i u_i \right) \\
&= [S_{xx}^*]^{-1}\, plim\, \frac{1}{N} \sum_{i=1}^{N} x_i u_i = 0
\end{aligned}
$$

How to prove (2): $\sqrt{N}(\widehat{\beta} - \beta^*) \xrightarrow{a} N(0, [S_{xx}^*]^{-1} \, V \, [S_{xx}^*]^{-1})$?

$$\sqrt{N}(\widehat{\beta} - \beta^*) \;\; = \;\; S_{xx}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i u_i$$

Now $plim\, S_{xx}^{-1} = [S_{xx}^*]^{-1}$. And by vecCLT, $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i u_i \xrightarrow{a} N(0, V)$. So using "result 4" we're done!

So in the “general case”, we have

$$(\widehat{\beta} - \beta^*) \approx N(0, \frac{1}{N}\,[S^*_{xx}]^{-1}\,V\,[S^*_{xx}]^{-1})$$

where $V = Var[x_i u_i]$

In the classic normal case, we have (conditional on X)

$$(\widehat{\beta} - \beta^*) \sim N(0, \frac{1}{N}\,[S_{xx}]^{-1}\,\sigma_u^2)$$

where $\sigma_u^2 = Var[u_i]$.

$$N(0, \frac{1}{N}\,[S_{xx}^*]^{-1}\,V\,[S_{xx}^*]^{-1}) \quad vs. \quad N(0, \frac{1}{N}\,[S_{xx}]^{-1}\,\sigma_u^2)$$

What's the difference?

-approx. vs exact distribution

-we aren't conditioning on X

-we don't assume $E[u_i|x_i] = 0$, only $E[x_i u_i] = 0$

$$N(0, \frac{1}{N}\,[S^*_{xx}]^{-1}\,V\,[S^*_{xx}]^{-1}) \quad vs. \quad N(0, \frac{1}{N}\,[S_{xx}]^{-1}\,\sigma_u^2)$$

How do we actually estimate the var-cov, $\widehat{V}_\beta$ in the general case?

a) Approximate $S^*_{xx} = S_{xx}$. so estimate of $[S^*_{xx}]^{-1}$ is $[S_{xx}]^{-1}$

b) estimate

$$\widehat{V} = \frac{1}{N}\sum_i (x_i\widehat{u}_i)(x_i\widehat{u}_i)' = \frac{1}{N}\sum \widehat{u}_i^2(x_ix_i')$$

c) form $\widehat{V}_\beta = \frac{1}{N}S_{xx}^{-1}\widehat{V}S_{xx}^{-1}$

We are estimating: $\widehat{V} = \frac{1}{N}\sum_i \widehat{u}_i^2 (x_i x_i')$. In the 3×3 case:

$$\widehat{V} = \frac{1}{N}\sum_i \begin{pmatrix} \widehat{u}_i^2 & \widehat{u}_i^2 x_{2i} & \widehat{u}_i^2 x_{3i} \\ \widehat{u}_i^2 x_{2i} & \widehat{u}_i^2 x_{2i}^2 & \widehat{u}_i^2 x_{3i} x_{2i} \\ \widehat{u}_i^2 x_{3i} & \widehat{u}_i^2 x_{3i} x_{2i} & \widehat{u}_i^2 x_{3i}^2 \end{pmatrix}$$

What happens if $\widehat{u}_i^2$ and $x_i x_i'$ are uncorrelated across obs?

USEFUL FACTOID:

If two random variables a_i and b_i are uncorrelated in a sample then

$$\frac{1}{N}\sum_i a_i b_i = \frac{1}{N}\sum_i a_i \frac{1}{N}\sum_i b_i = \bar{a}\bar{b}$$

Proof: write: $a_i = \bar{a} + a_i - \bar{a}$; $\quad b_i = \bar{b} + b_i - \bar{b}$. Finish as an exercise.

So if $\widehat{u}_i^2$ and $x_i x_i'$ uncorrelated in the sample,

$$\widehat{V} = \frac{1}{N}\sum_i \widehat{u}_i^2 (x_i x_i') = \frac{1}{N}\sum_i \widehat{u}_i^2 \frac{1}{N}\sum_i x_i x_i'$$

In the uncorrelated case

$$\widehat{V}_\beta = \frac{1}{N} S_{xx}^{-1} \widehat{V} S_{xx}^{-1} = \frac{1}{N}\left(\frac{1}{N}\sum_i \widehat{u}_i^2\right) S_{xx}^{-1}$$

$$\Rightarrow \widehat{V}_\beta = \frac{1}{N}\widetilde{\sigma}_u^2 S_{xx}^{-1}$$

which is the same as the classic normal except $\widetilde{\sigma}_u^2$ does not correct for d.f.!

Important extension - samples with group-level dependence

- many data sets have a "grouped" design

- for example: we might have observations from people who all live in the same state

- we are often concerned that their errors are correlated $\Rightarrow$non-i.i.d. sample!

- we will sketch a way to deal with this

Assume:

- data are from G groups ($g = 1, 2, ...G$)

- in each group there are m observations (so $N = Gm$)

- u_i and u_j may be correlated if i, j are in same group

- we can label each observation by 2 subscripts (so $i \to gj$):

$$y_{gj} = x_{gj}\beta^* + u_{gj}$$

$$\widehat{\beta} - \beta^* = S_{xx}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i u_i\right)$$

so lets look at the sum term:

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} x_i u_i &= \frac{1}{Gm}\sum_{g=1}^{G}\sum_{j} x_{gj}u_{gj} \\
&= \frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{m}\sum_{j} x_{gj}u_{gj}\right)
\end{aligned}$$

Let's call $z_g = \frac{1}{m}\sum_j x_{gj}u_{gj}$. Then we can write the sum term as:

$$\frac{1}{N}\sum_{i=1}^{N} x_i u_i = \frac{1}{G}\sum_{g=1}^{G} z_g$$

So we have that:

$$\widehat{\beta} - \beta^* = S_{xx}^{-1} \left(\frac{1}{G} \sum_{g=1}^{G} z_g \right)$$

Each of the z_g terms is independent (from separate groups) and $E[z_g] = 0$, so we apply a CLT to

$$\sqrt{G}(\widehat{\beta} - \beta^*) = S_{xx}^{-1} \left(\frac{1}{\sqrt{G}} \sum_{g=1}^{G} z_g \right)$$

And if we knew $V(z_g) = \Omega$ then we could say that

$$\sqrt{G}(\widehat{\beta} - \beta^*) \xrightarrow{a} N(0, [S_{xx}^*]^{-1}\, \Omega\, [S_{xx}^*]^{-1}).$$

The "trick" is how to get an estimate of $V(z_g) = \Omega$. Lets consider the case where $m = 2$

$$
\begin{aligned}
z_g &= \frac{1}{m}\sum_j x_{gi}u_{gi} \\
&= \frac{1}{2}(x_{g1}u_{g1} + x_{g2}u_{g2})
\end{aligned}
$$

The variance of this is a $k \times k$ matrix:

$$V(z_g) = \frac{1}{4}E[(x_{g1}x'_{g1})u^2_{g1} + (x_{g1}x'_{g2} + x_{g2}x'_{g1})u_{g1}u_{g2} + (x_{g2}x'_{g2})u^2_{g2}]$$

So we can estimate this using the estimated OLS residuals:

$$\widehat{\Omega} = \frac{1}{4G}\sum_g[(x_{g1}x'_{g1})\widehat{u}^2_{g1} + (x_{g1}x'_{g2} + x_{g2}x'_{g1})\widehat{u}_{g1}\widehat{u}_{g1} + (x_{g2}x'_{g2})\widehat{u}^2_{g2}]$$

Couple of points:

- need large number of groups or “clusters” – at least 30

- our “degrees of freedom” are the number of clusters

- classical standard errors: $\widehat{V}_\beta = \frac{1}{N}\widehat{\sigma}_u^2 S_{xx}^{-1}$

- robust standard errors: $\widehat{V}_\beta = \frac{1}{N} S_{xx}^{-1} \widehat{V} S_{xx}^{-1}$

- clustered standard errors: $\widehat{V}_\beta = \frac{1}{N} S_{xx}^{-1} \widehat{\Omega} S_{xx}^{-1}$

Clustered s.e.'s are good when we have "mis-specification" across groups of observations, as in the baby-weight example. The mis-specification means that the residuals for all observations with the same weight contain a shared error component

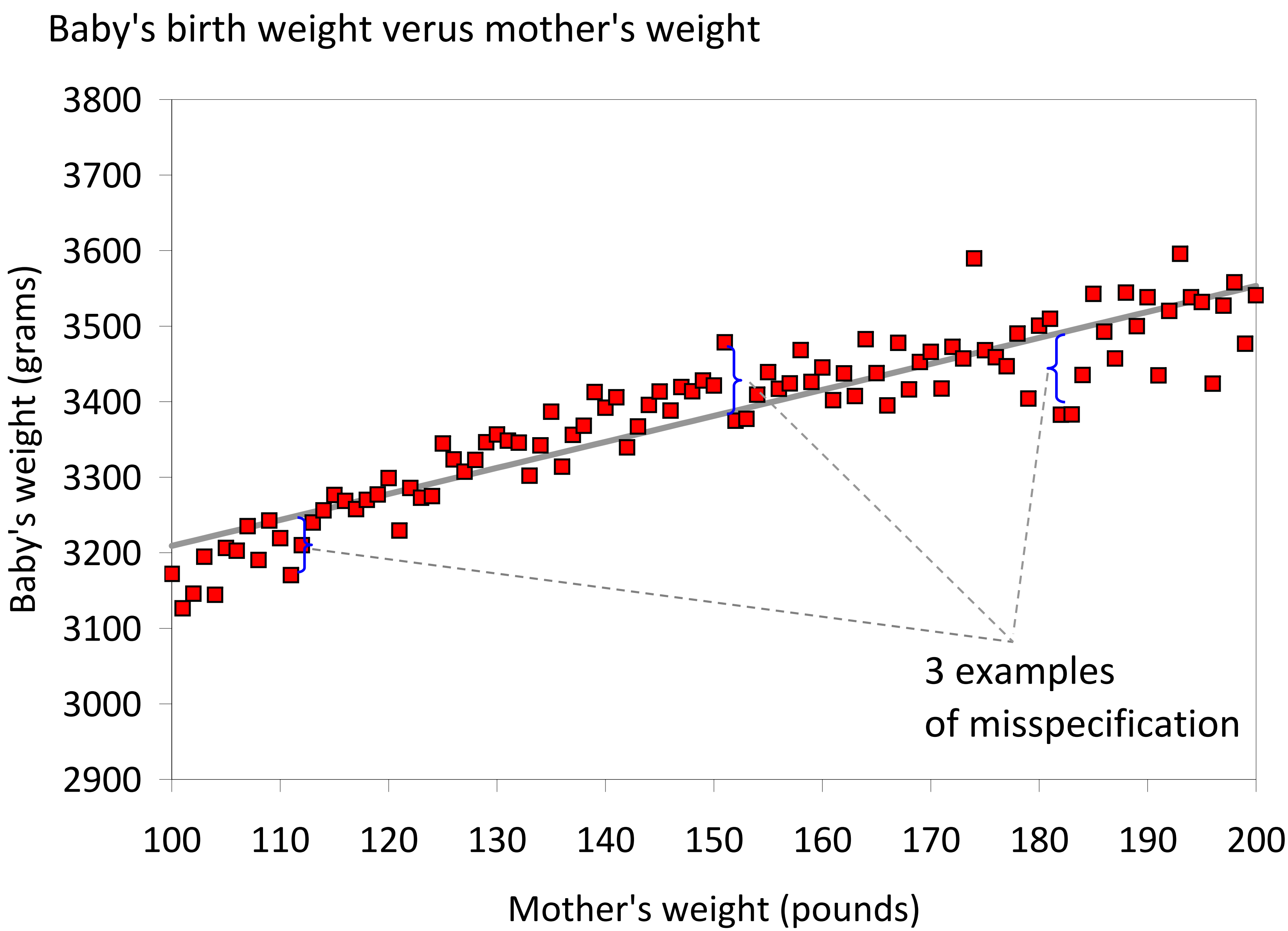
For the mom-baby linear regression (restricting to 200 $\leq momweight \leq$ 200) we have:

- $\widehat{\beta}_2$ =3.733

- classical standard errors: $\widehat{V}_\beta = \frac{1}{N}\widehat{\sigma}_u^2 S_{xx}^{-1} = .086056$

- robust standard errors: $\widehat{V}_\beta = \frac{1}{N} S_{xx}^{-1} \widehat{V} S_{xx}^{-1} = .087817$

- clustered standard errors: $\widehat{V}_\beta = \frac{1}{N} S_{xx}^{-1} \widehat{\Omega} S_{xx}^{-1} = .14617$



Baby's birth weight verus mother's weight
Baby's weight (grams)
3800
3700
3600
3500
3400
3300
3200
3100
3000
2900
100
110
120
130
140
150
160
170
180
190
200
Mother's weight (pounds)
3 examples
of misspecification