# OLS+

Readings

Matrix OLS review

Inference

Interpretation

Nonlinear models

# Readings

# Readings

This week: misc topics in linear models least squares estimation.

- Appendix E
  - Section 4 minus "Wald Statistics. . . "
- Chapter 1: Bedtime reading
- Chapter 2: Study 2.4
- Chapter 3: Study . . .
  - in 3.1, p. 69-71
  - in 3.2, "Interpreting . . . " and "The meaning of . . . "
  - in 3.3, "Omitted variable bias: the simple case"
  - 3.6
- Chapter 4: Skip
- Chapter 5: Study everything until "Other large sample tests. . . "
- Chapter 6: Study 6.2
- Chapter 8: Study 8.1 and 8.2 (skip for test 1)

# Matrix OLS review

# Model equation

The population regression model can now be written in matrix form, for all the observations in our random sample jointly:

$$y = X\beta + u$$

# Estimator

- The **ordinary least squares** estimator for $\beta$ is

$$\hat{\beta} = (X'X)^{-1}X'y$$

- **Projection matrix** $P_X = X(X'X)^{-1}X'$:
  - $P_X y = \hat{y}$
  - $P_X P_X = P_X$
- **Residual maker** $M_X = I_{k+1} - P_X$
  - $M_X y = y - \hat{y} = \hat{u}$
  - $M_X M_X = M_X$

# Assumptions

**Assumption E.**

1. The population regression equation can be written as $y = X\beta + u$
2. The matrix $X$ has full column rank, $rank(X) = k + 1$
3. Zero conditional mean: $E(u|X) = O_{n \times 1}$
4. Conditional variance of $u$ given $X$ is

$$Var(u|X) = \sigma^2 I_n$$

# Estimator properties

- Unbiasedness:
$$E(\hat{\beta}) = \beta$$
- Consistency:
$$\hat{\beta} \xrightarrow{p} \beta$$
- Variance:
$$Var\left(\hat{\beta}\middle| X\right) = \sigma^2(X'X)^{-1}$$
- Efficiency
- Asymptotic normality?

# Inference

# Asymptotic distribution

Related readings:

- E.4, Ch.5, Ch. 8

Asymptotic normality derivation: *whiteboard*

# Unknown $\sigma$

Our asympotic normality result says that, under E.1-E.4, then

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 A^{-1}\right)$$

where $A = \text{plim} \frac{1}{n} \sum_i x_i' x_i$.

To use this result, we need an estimator for $\sigma^2$.

# Estimating $\sigma$

Since

$$\sigma^2 = E[u_i^2|X_i] \\ = E[u_i^2]$$

a reasonable estimator is

$$\widehat{\sigma^2} = \frac{1}{n}\sum_i (y_i - X_i\hat{\beta})^2$$

# Intermezzo: unbiased

In previous course, you performed a degree-of-freedom correction. Why?

$$
\begin{aligned}
E[s^2] &= E\left[\frac{1}{n-k-1}\sum_i (y_i - X_i\hat{\beta})^2\right] \\
&= E\left[\frac{1}{n-k-1}(M_X y)'(M_X y)\right] \\
&= \frac{1}{n-k-1}E[(M_X u)'(M_X u)] \\
&= \frac{1}{n-k-1}E[u' M_X u] \\
&= \frac{1}{n-k-1}E[tr(M_X)u'u)] \\
&= \frac{tr(M_X)}{n-k-1}E[u'u)] \\
&= \sigma^2.
\end{aligned}
$$

# No homoskedasticity

However, assumption E.4 imposes strong assumptions.

```
library(Ecdat, quietly=TRUE, warn.conflicts=FALSE, verbose=
```

```
##
## Attaching package: 'Ecfun'
```

```
## The following object is masked from 'package:base':
##
##     sign
```

```
data("Bwages")
```

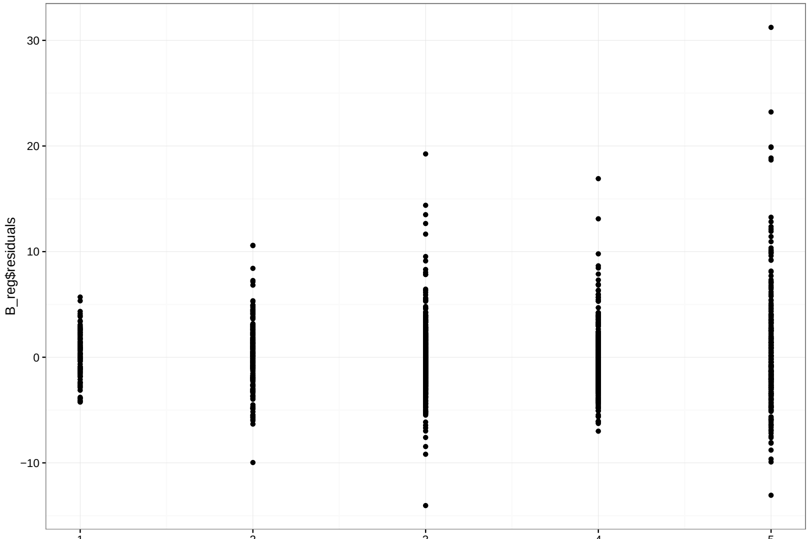
(...)

```
B_reg <- lm(wage ~ educ + exper, data=Bwages)
summary(B_reg)
```

```
##
## Call:
## lm(formula = wage ~ educ + exper, data = Bwages)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.0436  -2.0808  -0.4068   1.5915  31.2307
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.07374    0.37269   2.881  0.00402 **
## educ         1.93037    0.08154  23.674  < 2e-16 ***
## exper        0.20069    0.00966  20.774  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

# Residual plot

```
library(ggplot2)
qplot(Bwages$educ,B_reg$residuals) + theme_bw()
```



# Belgian wages

Do you believe that E.4 holds? In other words, do you believe that the errors are *homoskedastic*?

# Limit distribution without E.4

$$\sqrt{n}\left(\hat{\beta}-\beta\right) = \sqrt{n}(X'X)^{-1}X'u$$
$$= \left(\frac{1}{n}\sum_i x_i' x_i\right)^{-1}\left(\frac{\sqrt{n}}{n}\sum_i x_i' u_i\right)$$

and

- first term converges to $A^{-1}$ in probability
- second term converges to

$$\mathcal{N}\left(0, E\left(u_i^2 x_i' x_i\right)\right)$$

# Asymptotic variance

Consequentially,

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, A^{-1}BA^{-1}\right)$$

where

$$A = E(x_i'x_i)$$
$$B = E(u_i^2 x_i'x_i)$$

and we call

$$Avar(\hat{\beta}) = A^{-1}BA^{-1}$$

the **asymptotic variance** of the OLS estimator.

# Specialization under E.4

Under E.4, $B = \sigma^2 A$ so that

$$Avar(\hat{\beta}) = \sigma^2 A^{-1}$$

and we are back to the homoskedastic case.

# Without E.4

Estimate

$$\hat{A} = \frac{1}{n}\sum_i x_i' x_i$$

$$\hat{B} = \frac{1}{n}\sum_i \hat{u}_i^2 x_i' x_i$$

$$\widehat{Avar\left(\hat{\beta}\right)} = \hat{A}^{-1}\hat{B}\hat{A}^{-1}$$

# CLT without E.4

For each component $j$ of $\hat{\beta}_j$, let

$$SE(\hat{\beta}_j) = \sqrt{\widehat{Avar\left(\hat{\beta}_j\right)}}$$

then

$$\frac{\sqrt{n}(\hat{\beta}_j - \beta_j)}{SE(\hat{\beta}_j)} \xrightarrow{d} \mathcal{N}(0, 1)$$

as $n \to \infty$

# Results

```r
library(sandwich)
non_robust <- summary(B_reg)
robust_SE <- sqrt(diag(vcovHC(B_reg)))
full_results <- cbind(non_robust$coefficients[,1:2],robust_
print(full_results,digits=3)
```

```
##             Estimate Std. Error robust_SE
## (Intercept)    1.074    0.37269    0.4154
## educ           1.930    0.08154    0.0968
## exper          0.201    0.00966    0.0114
```

# Using the CLT

- Implicitly used above
- A 95% confidence interval for $\beta_j$ is given by

$$\left[\hat{\beta}_j \pm 1.96 SE(\hat{\beta}_j)\right]$$

- Self-study

# Interpretation

# Self-study

How to interpreting OLS coefficients is discussed in the required readings.

The Frisch-Waugh-Lovell theorem that follows gives a **mechanical** interpretation

# FWL: intro

Split your regressor matrix $X$ into two:

- $X_1$ contains the first $k_1$ regressors
- $X_2$ contains the second $k_2$ regressors

Split $\beta$ similarly. Then write:

$$y = X_1\beta_1 + X_2\beta_2 + u$$

# FWL: FOC

Solve for $\hat{\beta}_1$ and $\hat{\beta}_2$ using FOCs:

$$0 = X_1'(y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2)$$
$$0 = X_2'(y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2)$$

Or:

$$X_1'X_1\hat{\beta}_1 = X_1'y - X_1'X_2\hat{\beta}_2$$
$$X_2'X_2\hat{\beta}_2 = X_2'y - X_2'X_1\hat{\beta}_1$$

# FWL: Step 1

Solve for $\hat{\beta}_1$:

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1(y - X_2\hat{\beta}_2)$$

- Looks like a regression of $y$ free of $X_2$ onto $X_1$.
- Substitute that in the other FOC.

# FWL: Step 1.5

We will replace, in

$$X_2'X_2\hat{\beta}_2 = X_2'y - X_2'X_1\hat{\beta}_1$$

the term

$$\begin{aligned} X_2'X_1\hat{\beta}_1 &= X_2'X_1(X_1'X_1)^{-1}X_1(y - X_2\hat{\beta}_2) \\ &= X_2'P_{X_1}y - X_2'P_{X_1}X_2\hat{\beta}_2 \end{aligned}$$

# FWL: Step 2

Leads to

$$X_2'X_2\hat{\beta}_2 - X_2'P_{X_1}X_2\hat{\beta}_2 = X_2'y - X_2'P_{X_1}y$$

which can we rewritten as

$$(X_2'M_{X_1}X_2)\hat{\beta}_2 = X_2M_{X_1}y$$

- Recall what $P_{X_1}$ and $M_{X_1}$ do?
- Recall that they are both idempotent, i.e. $X_2'M_{X_1}y = X_2'M_{X_1}'M_{X_1}y$
- Is $X_2'P_{X_1}X_2$ invertible?

# FWL: Step 3

Conclusion:

$$\hat{\beta}_2 = (X_2' M_{X_1} X_2)^{-1} X_2 M_{X_1} y$$

# FWL: interpretation

Consider two ways of obtaining an estimator for $\beta_2$:

1. Regress $y$ on $X_1$ and $X_2$ and look at the coefficients on $X_2$
2. Three steps
    2.1 Regress $y$ on $X_1$ and get residuals ($M_{X_1} y$)
    2.2 Regress $X_2$ on $X_1$ and get residuals ($M_{X_1} X_2$)
    2.3 Regress a. on b.

FWL says: 1. and 2. are identical!

# FWL: demo

```
summary(B_reg)
```

```
##
## Call:
## lm(formula = wage ~ educ + exper, data = Bwages)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.0436  -2.0808  -0.4068   1.5915  31.2307
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.07374    0.37269   2.881  0.00402 **
## educ         1.93037    0.08154  23.674  < 2e-16 ***
## exper        0.20069    0.00966  20.774  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
```

# FWL: demo (2)

```
step1 <- lm(wage~educ,data=Bwages)
Bwages$y_resid <- step1$residuals
step2 <- lm(exper~educ,data=Bwages)
Bwages$X_resid <- step2$residuals
step3 <- lm(y_resid~X_resid,data=Bwages)
```

# FWL demo (3)

```r
summary(step3)
```

```
##
## Call:
## lm(formula = y_resid ~ X_resid, data = Bwages)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.0436  -2.0808  -0.4068   1.5915  31.2307
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.338e-16  9.395e-02    0.00        1
## X_resid      2.007e-01  9.657e-03   20.78   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## Residual standard error: 3.604 on 1470 degrees of freedo
```

# FWL: conclusion

*The regression coefficient estimate can be interpreted as the effect of $X$ on $y$ after controlling for all the other variables.*

# Nonlinear models

# Nonlinear models

Some nonlinear models can be estimated using OLS. We look at

- Quadratic
- Logarithmic (self-study)
  - log-log
  - log-linear
  - linear-log

# Quadratic model

Consider the population regression function

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + u_i$$

- What is in the error term?
- What is the interpretation of $\beta_1$?

# Quadratic model

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + u_i$$

- This model looks nonlinear.
- It is **linear in the parameters.**
- Can be estimated using OLS by letting $X_{i3} = X_{i2}^2$

# QM (2)

```
BM_quad <- lm(wage~educ + exper + I(exper^2),data=Bwages)
summary(BM_quad)
```

```
##
## Call:
## lm(formula = wage ~ educ + exper + I(exper^2), data = Bw
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.2549  -2.0108  -0.3923   1.5720  31.1177
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.057261   0.422808  -0.135    0.892
## educ         1.932966   0.080745  23.939  < 2e-16 ***
## exper        0.368841   0.032129  11.480  < 2e-16 ***
## I(exper^2)  -0.004435   0.000809  -5.482 4.94e-08 ***
## ---
```

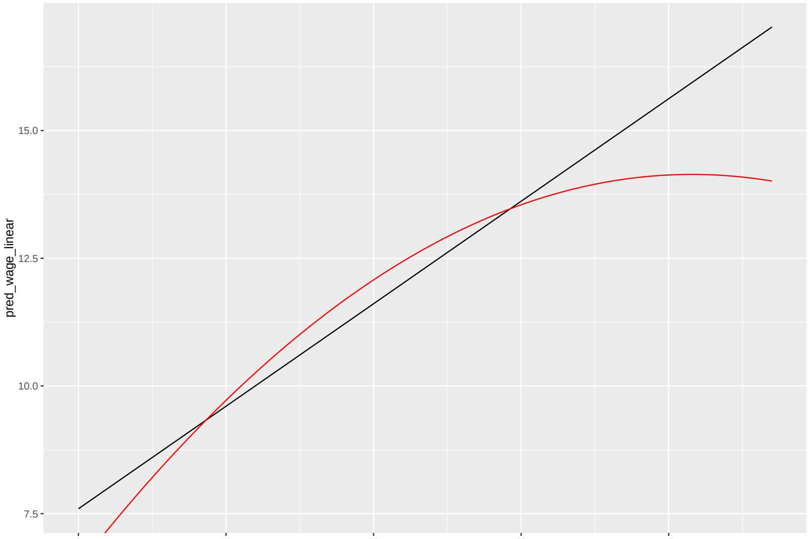
# QM (3)

Graphing for comparison:

```
av_educ <- mean(Bwages$educ)
min_exper <- min(Bwages$exper)
max_exper <- max(Bwages$exper)
exper_vals <- seq(from=min_exper,to=max_exper,length=100)
pred_wage_linear <- B_reg$coefficients[1] +
  B_reg$coefficients["educ"]*av_educ +
  B_reg$coefficients["exper"]*seq(from=min_exper,to=max_exp
pred_wage_quadratic <- BM_quad$coefficients[1] +
  BM_quad$coefficients["educ"]*av_educ +
  BM_quad$coefficients["exper"]*exper_vals +
  BM_quad$coefficients["I(exper^2)"]*exper_vals^2
```

# Plot

```
qplot(x=exper_vals,y=pred_wage_linear,geom="line") +
  geom_line(y=pred_wage_quadratic,colour="red")
```



# Turning point

The quadratic relationship $y = b_0 + b_1 x + b_2 x^2$ has a turning point if $b_1$ and $b_2$ have different signs, with turning point

$$x^* = \frac{-b_1}{2b_2}.$$

In our case

```
-BM_quad$coefficients["exper"]/(2*BM_quad$coefficients["I(e
```

```
##    exper
## 41.58001
```

whereas

```
max_exper
```

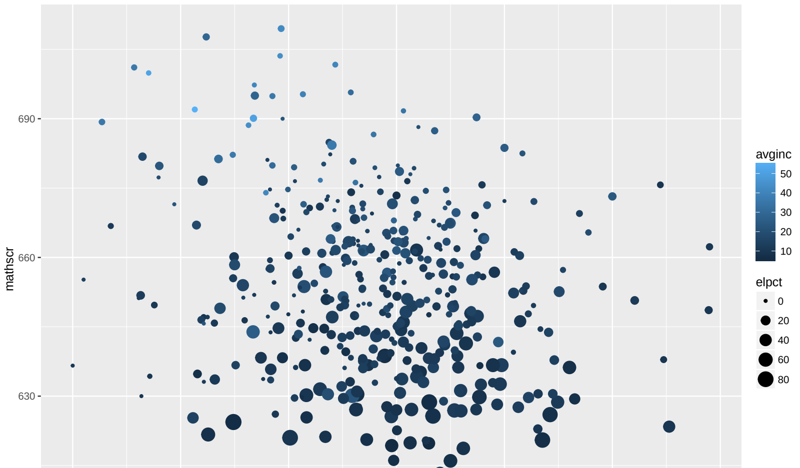
```
## [1] 47
```

# Interpretation

Except for the turning point, it is hard to pick **one number** to represent **the** effect of experience on wages, due to the nonlinear relationship between experience and wages.

# Example 2: Class-size

```
data("Caschool")
ggplot(data=Caschool,aes(x=str,y=mathscr,
                         size=elpct,color=avginc)) +
      geom_point()
```



# Class-size (2)

```
str_quad <- lm(mathscr ~ str + I(str^2) + mealpct + avginc
summary(str_quad)
```

```
##
## Call:
## lm(formula = mathscr ~ str + I(str^2) + mealpct + avginc
##     elpct, data = Caschool)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.405  -6.673   0.063   5.883  31.610
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.926e+02  3.595e+01  19.269  < 2e-16 ***
## str         -2.927e+00  3.419e+00  -0.856  0.39250
## I(str^2)     6.622e-02  8.579e-02   0.772  0.44061
## mealpct     -3.856e-01  3.323e-02 -11.605  < 2e-16 ***
```

# Class size (3)

```
-str_quad$coefficients["str"]/(2*str_quad$coefficients["I(s
```

```
##     str
## 22.0993
```