ECON207 Session 6

Hypothesis Testing

Anthony Tay

This Version: 31 Jul 2024

Session 6

- Recall Formulas and Results regarding OLS using Matrix Algebra
- Basic Hypothesis Tests in Linear Regression
- Specification Tests
  - Tests of Heteroskedasticity
  - RESET
  - Non-nested Alternatives
  - Testing for Non-normality in Noise Terms

Session 6.1

Session 6.1 Recap

- Quick recap of OLS estimation of MLR

OLS Estimation of MLR (Recap)

MLR: $Y = X\beta + \epsilon$ *X is $n \times (k+1)$, i.e., n obs, k regressors + intercept*

OLS Estimator: $\hat{\beta}^{ols} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$ *Requires $n > k$, full column rank*

Algebraic Properties:

- OLS Fitted values: $\hat{y}^{ols} = X\hat{\beta}^{ols} = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$
- OLS Residuals: $\hat{\epsilon}^{ols} = y - \hat{y}^{ols} = y - X\hat{\beta}^{ols} = (I_n - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}})y$
- Satisfies $X^{\mathrm{T}}\hat{\epsilon}^{ols} = 0_{(k+1)\times 1}$
- SST Decomposition: $y^{\mathrm{T}}M_0 y = \hat{y}_{ols}^{\mathrm{T}}M_0\hat{y}_{ols} + \hat{\epsilon}_{ols}^{\mathrm{T}}\hat{\epsilon}_{ols}$ or SST = SSE + SSR
- $R^2 = 1 - \dfrac{SSE}{SST}$, Adj.-$R^2 = 1 - \dfrac{SSR}{SST}\dfrac{n-1}{n-k-1}$

Properties of OLS estimators (Recap)

If $E(\epsilon \mid X) = 0$

- $\hat{\beta}^{ols}$ is unbiased
- het.-robust var: $\widehat{Var_{HC}(\hat{\beta}^{ols})} = \left(\sum_{i=1}^{n} X_{i*}^{\mathrm{T}} X_{i*}\right)^{-1} \left(\sum_{i=1}^{n} \hat{\epsilon}_{i,ols}^{2} X_{i*}^{\mathrm{T}} X_{i*}\right) \left(\sum_{i=1}^{n} X_{i*}^{\mathrm{T}} X_{i*}\right)^{-1}$

If $E(\epsilon \mid X) = 0$ and $E(\epsilon\epsilon^{\mathrm{T}} \mid X) = \sigma^2 I_n$

- $Var(\hat{\beta}^{ols} \mid X) = \sigma^2 (X^{\mathrm{T}} X)^{-1}$ estimated using $\widehat{Var(\hat{\beta}^{ols} \mid X)} = \widehat{\sigma^2} (X^{\mathrm{T}} X)^{-1}$
- Unbiased estimator for σ^2: $\widehat{\sigma^2} = \dfrac{\hat{\epsilon}_{ols}^{\mathrm{T}} \hat{\epsilon}_{ols}}{n-k-1}$
- $\hat{\beta}^{ols}$ is “best linear unbiased”

Properties of OLS estimators (Recap)

- Estimating $E(Y \mid X)$ (Predictive)
- Unbiased estimation assumes
  - no "sampling issues"
  - correct specification
- Causal interpretation for any particular regressor depends on
  - whether there are any simultaneity issues
  - whether you have appropriately controlled for all confounding factors

Session 6.2

Session 6.2 Hypothesis Testing in MLR

- Single hypothesis (t-test)
- Multiple hypotheses (F-test)

Hypothesis Testing in MLR

Suppose you wish to use MLR

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \epsilon_i$$

to test hypotheses, e.g., single hypothesis

- To test: $H_0 : \beta_1 = 0$ vs $H_A : \beta_1 \neq 0$
- To test: $H_0 : \beta_2 = 1$ vs $H_A : \beta_2 \neq 1$
- To test $\beta_1 + \beta_2 = 1$ vs $H_A : \beta_1 + \beta_2 \neq 1$

or joint hypotheses

- To test: $H_0 : \beta_1 = 0$ and $\beta_2 = 0$ vs $H_A : \beta_1 \neq 0$ or $\beta_1 \neq 0$ (or both)

Hypothesis Testing in MLR (Case 1: Normal Errors)

Assumption: $\epsilon \mid X \sim \text{Normal}_n(0, \sigma^2 I)$.

Digression:

If $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ is jointly Normal with mean μ and variance-covariance matrix Ω, where μ is $n \times 1$ and Ω is $n \times n$, then the pdf is

$$f(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) = (2\pi)^{-\frac{n}{2}} (\det \Omega)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\epsilon - \mu)^{\mathrm{T}} \Omega^{-1} (\epsilon - \mu)\right)$$

If $\mu = 0_{n \times 1}$, then this simplifies to

$$f(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) = (2\pi)^{-\frac{n}{2}} (\det \Omega)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\epsilon^{\mathrm{T}} \Omega^{-1} \epsilon\right)$$

Hypothesis Testing in MLR (Case 1: Normal Errors)

If furthermore $\Omega = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, so that the variables are uncorrelated, then

$$\begin{aligned} f(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) &= (2\pi)^{-\frac{n}{2}} \left(\prod_{i=1}^{n} \sigma_i^2 \right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \frac{\epsilon_i^2}{\sigma_i^2} \right) \\ &= \phi(\epsilon_1; 0, \sigma_1^2)\, \phi(\epsilon_2; 0, \sigma_2^2) \times \cdots \times \phi(\epsilon_n; 0, \sigma_n^2) \end{aligned}$$

where $\phi(x; \mu, \sigma)$ is the pdf of a univariate Normal with mean μ and variance σ_i^2. That is, uncorrelated multivariate Normal variables are independent

If $\sigma_i^2 = \sigma^2$ for all i, then

$$f(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \epsilon_i^2 \right)$$

Hypothesis Testing in MLR (Case 1: Normal Errors)

Impt: If $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ are distributed Multivariate Normal, then

$$a_1\epsilon_1 + a_2\epsilon_2 + \cdots + a_n\epsilon_n$$

remains Normally distributed

Therefore

$$\widehat{\beta}^{ols} = \beta + (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\epsilon$$

has distribution

$$\widehat{\beta}^{ols} \mid X \sim \text{Normal}_{k+1}(\beta, \sigma^2(X^{\mathrm{T}}X)^{-1})$$

This can be used to develop t and F-tests of linear hypotheses regarding elements of β

Hypothesis Testing in MLR (Case 1: Normal Errors)

General single linear hypothesis

$$H_0 : r^{\mathrm{T}}\beta = r_0 \quad \text{vs} \quad r^{\mathrm{T}}\beta \neq r_0.$$

E.g., in the regression

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$$

- To test: $\beta_1 = 0$, set $r^{\mathrm{T}} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$ and $r_0 = 0$
- To test: $\beta_2 = 1$, set $r^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$ and $r_0 = 1$
- To test $\beta_1 + \beta_2 = 1$, set $r^{\mathrm{T}} = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$ and $r_0 = 1$
- To test $\beta_1 = \beta_2$, set $r^{\mathrm{T}} = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix}$ and $r_0 = 0$

Hypothesis Testing in MLR (Case 1: Normal Errors)

From $\widehat{\beta}^{ols} \mid X \sim \text{Normal}_{k+1}(\beta, \sigma^2(X^{\mathrm{T}}X)^{-1})$, we have

$$r^{\mathrm{T}}\widehat{\beta}^{ols} \mid X \sim \text{Normal}(r^{\mathrm{T}}\beta, r^{\mathrm{T}}(\sigma^2(X^{\mathrm{T}}X)^{-1})r)$$

If the null hypothesis $r^{\mathrm{T}}\beta = r_0$ holds:

$$r^{\mathrm{T}}\widehat{\beta}^{ols} \mid X \sim \text{Normal}(r_0, r^{\mathrm{T}}(\sigma^2(X^{\mathrm{T}}X)^{-1})r)$$

$$\frac{r^{\mathrm{T}}\widehat{\beta}^{ols} - r_0}{\sqrt{r^{\mathrm{T}}(\sigma^2(X^{\mathrm{T}}X)^{-1})r}} \sim \text{Normal}(0, 1)$$

Hypothesis Testing in MLR (Case 1: Normal Errors)

If we replace σ^2 with $\widehat{\sigma^2}$, then it can be shown that

$$t = \frac{r^{\mathrm{T}}\widehat{\beta}^{ols} - r_0}{\sqrt{r^{\mathrm{T}}(\widehat{\sigma^2}(X^{\mathrm{T}}X)^{-1})r}} = \frac{r^{\mathrm{T}}\widehat{\beta}^{ols} - r_0}{\sqrt{r^{\mathrm{T}}\,\widehat{Var(\widehat{\beta}^{ols} \mid X)}r}} \sim t_{(n-k-1)}$$

We omit the proof of this last statement

Use this to test the hypothesis $H_0 : r^{\mathrm{T}}\beta = r_0$ in the usual way

Hypothesis Testing in MLR (Case 1: Normal Errors)

To test multiple hypotheses jointly, write the hypotheses as

$$H_0 : R\beta = r_0 \quad \text{vs} \quad H_A : R\beta \neq r_0$$

where R is a $J \times (k+1)$ matrix, and r_0 is a $J \times 1$ vector

E.g., In the regression

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$$

- To test the hypotheses

$$H_0 : \beta_1 + \beta_2 = 1 \text{ and } \beta_3 = 0 \quad \text{vs} \quad H_A : \beta_1 + \beta_2 \neq 1 \text{ or } \beta_3 \neq 0 \text{ (or both)}$$

$$\text{set:} \qquad R = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad r_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Hypothesis Testing in MLR (Case 1: Normal Errors)

Idea: Compare the sum of squared residuals from the restricted and unrestricted regressions

Restricted SSR will never be smaller then unrestricted SSR

It can be shown that if the null is true, then

$$F = \frac{(SSR_{res} - SSR_{unres})/J}{SSR_{unres}/n-k-1} = \frac{(\hat{\epsilon}_{res}^{\mathrm{T}}\hat{\epsilon}_{res} - \hat{\epsilon}_{ols}^{\mathrm{T}}\hat{\epsilon}_{ols})/J}{\hat{\epsilon}_{ols}^{\mathrm{T}}\hat{\epsilon}_{ols}/(n-k-1)} \sim F_{(J,n-k-1)}$$

Alt. Expression: $F = \dfrac{(R^2 - R_{res}^2)/J}{(1-R^2)/(n-k-1)}$

Hypothesis Testing in MLR (Case 1: Normal Errors)

E.g., $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$

$$H_0 : \beta_1 + \beta_2 = 1 \text{ and } \beta_3 = 0 \quad \text{vs} \quad H_A : \beta_1 + \beta_2 \neq 1 \text{ or } \beta_3 \neq 0 \text{ (or both)},$$

Restricted regression is

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + (1 - \beta_1) X_{i2} + 0 X_{i3} + \epsilon_i \\ &= \beta_0 + \beta_1 (X_{i1} - X_{i2}) + X_{i2} + \epsilon_i \end{aligned}$$

Hypothesis Testing in MLR (Case 1: Normal Errors)

- Regress equation without restrictions, get $SSR_{unres} = \hat{\epsilon}_{ols}^{\mathrm{T}}\hat{\epsilon}_{ols}$
- Regress $Y_i - X_{i2}$ on a constant and $X_{i1} - X_{i2}$
- Calculate restricted OLS residuals

$$\hat{\epsilon}_{i,res} = Y_i - \hat{\beta}_0^{res} - \hat{\beta}_1^{res}(X_{i1} - X_{i2}) - X_{i2}$$

- Calculate restricted OLS SSR

$$SSR_{res} = \sum_{i=1}^{n} \hat{\epsilon}_{i,res}^2 = \hat{\epsilon}_{res}^{\mathrm{T}}\hat{\epsilon}_{res}$$

- Calculate F-stat

Hypothesis Testing in MLR (Case 1: Normal Errors)

In practice you do not have to compute the restricted regression directly

It can be shown that

$$\hat{\epsilon}_{res}^{\mathrm{T}}\hat{\epsilon}_{res} - \hat{\epsilon}_{ols}^{\mathrm{T}}\hat{\epsilon}_{ols} = (R\hat{\beta}^{ols} - r_0)^{\mathrm{T}}[R(X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}]^{-1}(R\hat{\beta}^{ols} - r_0)$$

Furthermore, denominator of the F-statistic is $\widehat{\sigma^2}$, therefore

$$\begin{aligned} F &= (R\hat{\beta}^{ols} - r_0)^{\mathrm{T}}(R(\widehat{\sigma^2}(X^{\mathrm{T}}X)^{-1})R^{\mathrm{T}})^{-1}(R\hat{\beta}^{ols} - r_0)/J \\ &= (R\hat{\beta}^{ols} - r_0)^{\mathrm{T}}(R\,\widehat{Var(\hat{\beta}^{ols} \mid X)}\,R^{\mathrm{T}})^{-1}(R\hat{\beta}^{ols} - r_0)/J \\ &\sim F_{(J,n-k-1)} \end{aligned}$$

Empirical Example

- Recall Multiple Linear Regression using earnings.csv data

$$100\ln(earn) = \beta_0 + \beta_1 educ + \beta_2 educ^2 + \beta_3 height + \beta_4 male + \beta_5 \ln(wexp) + \beta_6 \ln(wexp)^2 + \beta_7 \ln(tenure) + \beta_8 \ln(tenure)^2 + \beta_9 age + \beta_{10} age^2 + \epsilon$$

```
dat <- read_csv("data\\earnings2019.csv", show_col_types=FALSE) %>%
        mutate(learn100 = 100*log(earn), educsq = educ^2, lwexp = log(wexp), lwexp2 = log(wexp)^2,
                                 ltenure=log(tenure), ltenure2=log(tenure)^2, agesq = age^2)
mdl1 <- lm(learn100 ~ educ + educsq + height + male +
            lwexp + lwexp2 + ltenure + ltenure2 + age + agesq, data=dat)
```

The following generates "standard" regression output

```
sum1 <- summary(mdl1);
```

We will use the following for "self-calculated" statistics

```
betahat <- as.matrix(coef(mdl1), ncol=1)
betavar <- as.matrix(vcov(mdl1), ncol=11)
df <- nrow(dat) - length(betahat)
```

Empirical Example

Estimation Results: lm.summary() output too long for slides

Recreating partial lm.summary() output below

```
sum1$coef %>% round(4)
```

```
            Estimate Std. Error  t value Pr(>|t|)
(Intercept)  81.1235    37.5369   2.1612   0.0307
educ        -15.2609     4.8159  -3.1689   0.0015
educsq        1.0065     0.1719   5.8565   0.0000
height        1.3433     0.2639   5.0905   0.0000
male         18.0716     2.1410   8.4405   0.0000
lwexp        -1.5820     2.7122  -0.5833   0.5597
lwexp2       -0.2263     0.7798  -0.2902   0.7717
ltenure       5.2235     2.6612   1.9628   0.0497
ltenure2      2.3144     0.7694   3.0078   0.0026
age           5.8237     0.4752  12.2563   0.0000
agesq        -0.0613     0.0052 -11.8434   0.0000
```

```
cat("\nResidual Std. Err.:", round(sum1$sigma,4),
    "on", sum1$df[2], "deg. of freedom")
cat("\nR2:", round(sum1$r.squared,4))
cat("\nAdjusted-R2:", round(sum1$adj.r.squared,4))
sum1f <- sum1$fstatistic
cat("\nF-stat:", sum1f[1], "on", sum1f[2], "and",
    sum1f[3], "DF,", "p-value:",
    1-pf(q=sum1f[1], df1=sum1f[2], df2=sum1$f[3]))
```

```
Residual Std. Err.: 54.514 on 4935 deg. of freedom
R2: 0.3039
Adjusted-R2: 0.3025
F-stat: 215.457 on 10 and 4935 DF, p-value: 0
```

Empirical Example

Standard regression output:

- t-statistics to test $H_0 : \beta_j = 0$ vs $H_A : \beta_j \neq 0$
- t-statistics are $t = \dfrac{\hat{\beta}_j^{ols}}{s.e.\left(\hat{\beta}_j^{ols}\right)}$, $j = 1, \ldots, k$, p-val based on $t(n-k-1)$
- F-statistic to test $H_0 : \beta_1 = \cdots = \beta_k = 0$ vs H_A : at least one $\beta_j \neq 0$, $j = 1, \ldots, k$
- F-statistic $F = \dfrac{R^2/k}{(1-R^2)/(n-k-1)}$, p-val based on $F(k, n-k-1)$

For other hypotheses:

- have to calculate own t- / F-statistics (we'll use both R packages and self-calc.)

Empirical Example

Our self-calculations will require estimating the var-cov matrix

Var-cov matrix assuming homoskedasticity

```
options(width=400)
round(betavar,3)     # VAR-COV ASSUMING HOMOSKEDASTICITY
```

	(Intercept)	educ	educsq	height	male	lwexp	lwexp2	ltenure	ltenure2	age	agesq
(Intercept)	1409.018	-154.411	5.479	-3.632	15.606	1.019	0.115	-3.887	1.708	-4.886	0.050
educ	-154.411	23.193	-0.825	-0.110	1.009	-0.360	0.076	0.215	-0.099	0.104	-0.001
educsq	5.479	-0.825	0.030	0.004	-0.033	0.014	-0.002	-0.010	0.004	-0.004	0.000
height	-3.632	-0.110	0.004	0.070	-0.385	-0.016	0.003	0.009	-0.002	-0.004	0.000
male	15.606	1.009	-0.033	-0.385	4.584	0.081	-0.020	-0.077	-0.002	0.038	0.000
lwexp	1.019	-0.360	0.014	-0.016	0.081	7.356	-1.988	-0.108	0.004	-0.136	0.002
lwexp2	0.115	0.076	-0.002	0.003	-0.020	-1.988	0.608	0.010	0.006	0.020	0.000
ltenure	-3.887	0.215	-0.010	0.009	-0.077	-0.108	0.010	7.082	-1.935	-0.120	0.002
ltenure2	1.708	-0.099	0.004	-0.002	-0.002	0.004	0.006	-1.935	0.592	0.012	0.000
age	-4.886	0.104	-0.004	-0.004	0.038	-0.136	0.020	-0.120	0.012	0.226	-0.002
agesq	0.050	-0.001	0.000	0.000	0.000	0.002	0.000	0.002	0.000	-0.002	0.000

Empirical Example

Heteroskedasticity-robust var-cov matrix

```
options(width=400)
varhat_HC0a = vcovHC(mdl1,type="HC0")
round(varhat_HC0a,3)  #HETEROSKEDASTICITY-ROBUST VAR-COV MATRIX
```

	(Intercept)	educ	educsq	height	male	lwexp	lwexp2	ltenure	ltenure2	age	agesq
(Intercept)	1345.536	-144.697	5.136	-3.296	14.791	-2.703	1.435	-10.416	3.477	-5.957	0.063
educ	-144.697	22.049	-0.785	-0.163	0.932	0.350	-0.173	0.829	-0.232	0.163	-0.002
educsq	5.136	-0.785	0.028	0.006	-0.028	-0.012	0.007	-0.032	0.009	-0.006	0.000
height	-3.296	-0.163	0.006	0.071	-0.382	-0.022	0.008	0.006	-0.001	-0.006	0.000
male	14.791	0.932	-0.028	-0.382	4.563	0.076	-0.048	0.098	-0.053	0.075	-0.001
lwexp	-2.703	0.350	-0.012	-0.022	0.076	8.468	-2.300	-0.121	0.041	-0.208	0.002
lwexp2	1.435	-0.173	0.007	0.008	-0.048	-2.300	0.705	0.029	-0.008	0.034	-0.001
ltenure	-10.416	0.829	-0.032	0.006	0.098	-0.121	0.029	7.554	-2.077	-0.030	0.001
ltenure2	3.477	-0.232	0.009	-0.001	-0.053	0.041	-0.008	-2.077	0.639	-0.022	0.000
age	-5.957	0.163	-0.006	-0.006	0.075	-0.208	0.034	-0.030	-0.022	0.266	-0.003
agesq	0.063	-0.002	0.000	0.000	-0.001	0.002	-0.001	0.001	0.000	-0.003	0.000

Empirical Example

We test $H_0 : \beta_7 = \beta_9$ vs $\beta_7 \neq \beta_9$, i.e., $H_0 : \beta_7 - \beta_9 = 0$ vs $\beta_7 - \beta_9 \neq 0$

Using formula developed in class:

```
r <- matrix(c(0,0,0,0,0,0,0,1,0,-1,0), ncol=1)
r0 <- 0
t <- t(r)%*%betahat / sqrt(t(r)%*%betavar%*%r)
tpval <- 2*(1-pt(abs(t),df))
cat("H0: ltenure = age. t-stat = ",
    round(t,6),", pval = ", round(tpval,6))
```

```
H0: ltenure = age. t-stat =  -0.21849 , pval =  0.827057
```

$$t = \frac{r^{\mathrm{T}} \hat{\beta}^{ols} - r_0}{\sqrt{r^{\mathrm{T}} \widehat{Var}(\hat{\beta}^{ols} \mid X) r}} \sim t_{(n-k-1)}$$

Do not reject the hypothesis at all conventional levels of significance

Empirical Example

$$H_0 : \beta_7 - \beta_9 = 0 \text{ vs } \beta_7 - \beta_9 \neq 0$$

```
linearHypothesis(mdl1, c('ltenure=age'))
```

- `linearHypothesis` from `lmtest` package produces an F-test
- The F-stat for one restriction is equal to the square of the t-stat (exercise)

```
Linear hypothesis test

Hypothesis:
ltenure - age = 0

Model 1: restricted model
Model 2: learn100 ~ educ + educsq + height + male + lwexp + lwexp2 + lten
    ltenure2 + age + agesq

  Res.Df      RSS Df Sum of Sq      F Pr(>F)
1   4936 14665880
2   4935 14665738  1    141.87 0.0477 0.8271
```

Do not reject the hypothesis at all conventional levels of significance

Empirical Example

Joint hypotheses $H_0 : \beta_5 = 0$ and $\beta_6 = 0$ vs $H_A : \beta_5 \neq 0$ or $\beta_6 \neq 0$

We use the F-test, with

$$R = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad r_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

```
R <- matrix(c(0,0,0,0,0, 1,0,0,0,0,0,
              0,0,0,0,0, 0,1,0,0,0,0), nrow=2, byrow=TRUE)
r0 <- matrix(c(0,0),nrow=2)
J = length(r0)
Rb <- R %*% betahat - r0
Fstat <- t(Rb) %*% solve(R%*%betavar%*%t(R)) %*% Rb / J
Fpval <- (1-pf(Fstat,J,df))
cat("H0: lwexp=0, lwexp2=0. F-stat ",round(Fstat,4),", pval", round(Fpval,6))
```

```
H0: lwexp=0, lwexp2=0. F-stat  3.1798 , pval 0.041679
```

Do not reject H_0 at 5% level of significance

Hypothesis Testing

```r
linearHypothesis(mdl1,c('lwexp = 0', 'lwexp2 = 0'))
```

```
Linear hypothesis test

Hypothesis:
lwexp = 0
lwexp2 = 0

Model 1: restricted model
Model 2: learn100 ~ educ + educsq + height + male + lwexp + lwexp2 + ltenure +
    ltenure2 + age + agesq

  Res.Df      RSS Df Sum of Sq      F  Pr(>F)
1   4937 14684637
2   4935 14665738  2     18899 3.1798 0.04168 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Non-Normality or Heteroskedasticity

Don't want to assume $\epsilon \mid X \sim \text{Normal}_n$?

Appeal to CLT:

- t-statistic approximately normal
- JF approximately $\chi^2(J)$

Heteroskedasticity

- We can use het. robust variances to construct het. robust t and F statistics
- replace $\widehat{Var(\hat{\beta}^{ols} \mid X)}$ with het. robust variance estimator in t and F statistic formula
    - "Heteroskedasticity-Robust" t- and F-statistics

Non-Normality or Heteroskedasticity

Self-calculated:

```
Fstat2 <- t(Rb) %*% solve(R%*%varhat_HC0a%*%t(R)) %*% Rb
Fpval2 <- (1-pchisq(Fstat2,J))
cat("H0: lwexp = 0, lwexp2 = 0. JF: ",round(Fstat2,4),", p-val", round(Fpval2,6))
```

```
H0: lwexp = 0, lwexp2 = 0. JF:  5.6432 , p-val 0.059512
```

Using packages (next slide):

Non-Normality or Heteroskedasticity

```
linearHypothesis(mdl1,c('lwexp=0', 'lwexp2=0'), vcov=varhat_HC0a, test="Chisq")
```

```
Linear hypothesis test

Hypothesis:
lwexp = 0
lwexp2 = 0

Model 1: restricted model
Model 2: learn100 ~ educ + educsq + height + male + lwexp + lwexp2 + ltenure +
    ltenure2 + age + agesq

Note: Coefficient covariance matrix supplied.

  Res.Df Df  Chisq Pr(>Chisq)
1   4937
2   4935  2 5.6432    0.05951 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Session 6.3

Session 6.3 Specification Tests

- Heteroskedasticity Testing
- Specification Tests
- Tests of Non-nested Alternatives
- Testing Normality of Error Terms

Will use

$$100ln(earn) = \beta_0 + \beta_1 wexp + \beta_2 tenure + \epsilon.$$

as illustration

Testing for Heteroskedasticity

Recall the Heteroskedasticity problem:

$Var(\epsilon_i) = \sigma_i^2$

- does not affect unbiasedness or consistency of OLS estimators (no problem)
- standard homoskedasticity-based estimator variance formulas incorrect
- OLS not efficient

Heteroskedasticity tests discussed here try to see if σ_i^2 depends on regressors

- Basic idea: use $\hat{\epsilon}_{i,ols}^2$ as proxy for $Var(\epsilon_i)$

Testing for Heteroskedasticity

Estimate main regression by OLS and obtain OLS residuals $\hat{\epsilon}_{i,ols}$

Then (Breusch-Pagan)

- Run the "test" regression

$$\hat{\epsilon}^2_{i,ols} = \alpha_0 + \alpha_1 X_{i1} + \ldots \alpha_k X_{ik} + u_i$$

  and test $H_0 : \alpha_1 = \cdots = \alpha_k = 0$ using an F-test.

- Use an "LM" test after running the regression above: under the null hypothesis, we have

$$nR^2_\epsilon \overset{a}{\sim} \chi^2(k)$$

  where R^2_ϵ is R^2 from the "test" regression, n is no. of obs.

Testing for Heteroskedasticity

- Can allow for non-linear forms by including powers and interaction terms in variance regression:

$$
\begin{aligned}
\hat{\epsilon}^2_{i,ols} = \alpha_0 &+ \alpha_1 X_{i1} + \cdots + \alpha_k X_{ik} \\
&+ \delta_1 X^2_{i1} + \cdots + \delta_k X^2_{ik} \\
&+ \gamma_{12} X_{i1} X_{i2} + \cdots + \gamma_{1k} X_{i1} X_{ik} \\
&+ \cdots + u_i
\end{aligned}
$$

  then test if slope coefficients are zero

- becomes infeasible pretty quickly!

Testing for Heteroskedasticity

- (White) Run the regression

$$\hat{\epsilon}^2_{i,ols} = \alpha_0 + \alpha_1 \hat{Y}_{i,ols} + \alpha_2 \hat{Y}^2_{i,ols} + u_i$$

  where $\hat{Y}_{i,ols}$ are OLS fitted values from main equation.
- Test $H_0 : \alpha_1 = \alpha_2 = 0$ using F-test or LM test.

Testing for Heteroskedasticity

Applying Breusch-Pagan test to

$$100ln(earn) = \beta_0 + \beta_1 wexp + \beta_2 tenure + \epsilon.$$

```
dat1 <- dat
mdl <- lm(learn100~wexp+tenure, data=dat1)              #--Main Equation
cat("Main Regression, Dependent Var.: ln(earnings)\n")  #--Main Regr Output Title
round(summary(mdl)$coefficients,4)                      #--Main Regr Output
```

```
Main Regression, Dependent Var.: ln(earnings)
            Estimate Std. Error  t value Pr(>|t|)
(Intercept) 302.4908     1.5582 194.1339        0
wexp         -0.4865     0.1113  -4.3724        0
tenure        1.8651     0.1086  17.1674        0
```

Testing for Heteroskedasticity

```
dat1$ehat <- residuals(mdl)                        #--Get OLS Residuals
heteq <- lm((ehat^2)~wexp+tenure, data=dat1)       #--BP-Test Regression
cat("Heteroskedasticity Test Regression\n")        #--Test Regr Output Title
round(summary(heteq)$coefficients,4)               #--Test Regr Output
```

```
Heteroskedasticity Test Regression
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 4015.3544   179.6718 22.3483   0.0000
wexp           5.3642    12.8288  0.4181   0.6759
tenure        -4.9394    12.5275 -0.3943   0.6934
```

Testing for Heteroskedasticity

```
## BP, F-version
f_het <- summary(heteq)$fstatistic  #--Retrieve F-Stat (stat, df1, df2)
cat("BP-F Stat: ", f_het[1], "  p-val: ", 1- pf(f_het[1], f_het[2], f_het[3]), "\n")
```

```
BP-F Stat:  0.1361259   p-val:  0.8727361
```

```
## BP, LM-version
lm_het <- nobs(heteq)*summary(heteq)$r.squared        # Calc. LM Stat.
lm_pval <- 1 - pchisq(lm_het, 2)  # 2 restrictions
cat("BP-LM Stat: ", lm_het, "     p-val: ", lm_pval, "\n", sep="")
```

```
BP-LM Stat: 0.272402     p-val: 0.8726672
```

Do not reject that homoskedastic noise terms

RESET Test for Functional Form Misspecification

Is Linear-in-Regressor specification correct?

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \epsilon_i \ ,$$

Check if adding

- powers (X_{i1}^2, X_{i2}^2,...) and
- interaction terms ($X_{i1}X_{i2}$, $X_{i1}X_{i3}$, etc.)

significantly improves the fit

RESET Test for Functional Form Misspecification

Test equation

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \alpha_2 \widehat{Y}_{i,ols}^2 + \cdots + \alpha_p \widehat{Y}_{i,ols}^p + \epsilon_i$$

Hypothesis of adequacy of functional form:

$$H_0 : \alpha_2 = \cdots = \alpha_p = 0$$

- Test equation cannot include $\widehat{Y}_{i,ols}$ (why?)
- Usually only the second or second and third powers are included

RESET Test for Functional Form Misspecification

Apply the RESET test to

$$\ln(earn) = \beta_0 + \beta_1 wexp + \beta_2 tenure + \epsilon$$

```r
mdl_base <- lm(learn100~wexp+tenure, data=dat1)
cat("Base Regression:\n")
round(summary(mdl_base)$coefficients, 4)
```

```
Base Regression:
            Estimate Std. Error  t value Pr(>|t|)
(Intercept) 302.4908     1.5582 194.1339        0
wexp         -0.4865     0.1113  -4.3724        0
tenure        1.8651     0.1086  17.1674        0
```

RESET Test for Functional Form Misspecification

```
dat1$yhat <- fitted(mdl_base)
mdl_test <- lm(learn100~wexp+tenure+I(yhat^2), data=dat1)
cat("\nTest Regression:\n")
round(summary(mdl_test)$coefficients, 4)
```

```
Test Regression:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1732.8527   239.5103  7.2350        0
wexp          -5.3101     0.8153 -6.5134        0
tenure        20.9247     3.1933  6.5528        0
I(yhat^2)     -0.0157     0.0026 -5.9722        0
```

The hypothesis of adequacy of the functional form in the base regression is rejected

Testing Nonnested Alternatives

Regression specifications such as

$$\begin{aligned} &[A] \quad Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i \\ \text{and} \quad &[B] \quad Y_i = \beta_0 + \beta_1 \ln X_{i1} + \beta_2 \ln X_{i2} + \epsilon_i \end{aligned}$$

are “non-nested alternatives”

Which fits better?

Testing Nonnested Alternatives

One alternative:

Construct a "super-model" containing both [A] and [B] as restricted cases

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 \ln X_{i1} + \beta_4 \ln X_{i2} + \epsilon_i$$

and test for coefficient significance

Multicollinearity problems

Testing Nonnested Alternatives

Another alternative:

- fit both models separately
- collect their fitted values
- include each fitted value series as a regressor in the other specification

$$\begin{aligned} &[\text{A'}] \quad Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \delta_1 \widehat{Y}^B_{i,ols} + \epsilon_i \\ \text{and} \quad &[\text{B'}] \quad Y_i = \beta_0 + \beta_1 \ln X_{i1} + \beta_2 \ln X_{i2} + \delta_2 \widehat{Y}^A_{i,ols} + \epsilon_i \end{aligned}$$

and test (separately) if fitted values are statistically significant

Testing Nonnested Alternatives

Example

$$
\begin{aligned}
&[A] \quad \ln(earn_i) = \beta_0 + \beta_1 wexp_i + \beta_2 tenure_i + \epsilon_i \,, \\
\text{and} \quad &[B] \quad \ln(earn_i) = \beta_0 + \beta_1 \ln(wexp_i) + \beta_2 \ln(tenure_i) + \epsilon_i \,.
\end{aligned}
$$

```
mdlA <- lm(learn100~wexp+tenure, data=dat1)
mdlB <- lm(learn100~lwexp+ltenure, data=dat1)
dat1$yhatA <- fitted(mdlA)
dat1$yhatB <- fitted(mdlB)
```

Testing Nonnested Alternatives

```
cat("Model A plus yhatB:\n")
mdlAplusB <- lm(learn100~wexp+tenure+yhatB, data=dat1)
round(summary(mdlAplusB)$coefficients,4)
```

```
Model A plus yhatB:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  28.2510    33.3054  0.8482   0.3963
wexp         -0.1189     0.1192 -0.9975   0.3186
tenure        0.2166     0.2272  0.9532   0.3406
yhatB         0.9075     0.1101  8.2430   0.0000
```

Testing Nonnested Alternatives

```
cat("\nModel B plus yhatA:\n")
mdlBplusA <- lm(learn100~lwexp+ltenure+yhatA, data=dat1)
round(summary(mdlBplusA)$coefficients,4)
```

```
Model B plus yhatA:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 255.5660    36.0583  7.0876   0.0000
lwexp        -3.7265     1.0257 -3.6330   0.0003
ltenure      15.7694     1.9183  8.2204   0.0000
yhatA         0.1219     0.1215  1.0033   0.3158
```

It appears that the specification [B] is preferred over specification [A].

Testing Normality of Noise Terms

ϵ_i normally distributed?

If $X \sim \text{Normal}(\mu, \sigma^2)$, then

$$\text{Skewness} \quad S = E((X-\mu)^3)/\sigma^3 = 0$$

$$\text{Kurtosis} \quad K = E((X-\mu)^4)/\sigma^4 = 3$$

Estimate by

$$\widehat{S} = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^3}{\left[\frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2\right]^{3/2}} \qquad \text{and} \qquad \widehat{K} = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^4}{\left[\frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2\right]^{2}}$$

Testing Normality of Noise Terms

The Jarque-Bera Test applies this to regression residuals

$$JB = \frac{n-k-1}{6}\left(\widehat{S}^2 + \frac{1}{4}(\widehat{K}-3)^2\right) \quad \overset{a}{\sim} \quad \chi^2_{(2)}$$

We test for normality of the residuals in the regression

$$100\ln(earn) = \beta_0 + \beta_1 \ln(wexp) + \beta_2 \ln(tenure) + \epsilon$$

Testing Normality of Noise Terms

```
Skew <- function(x){
  # Returns Skewness Coefficient
  return(mean((x-mean(x))^3)/(mean((x-mean(x))^2)^(3/2)))
}
Kurt <- function(x){
  # Returns Kurtosis Coefficient
  return(mean((x-mean(x))^4)/(mean((x-mean(x))^2)^2))
}
JB <- function(mdl){
  # requires lm object, returns JB Stat, p-val, Skewness and Kurtosis Coef.
  nobs <- nobs(mdl)
  nbetas <- summary(mdl)$df[1]
  ehat <- residuals(mdl)
  JBSkew <- Skew(ehat)
  JBKurt <- Kurt(ehat)
  JBstat <- ((nobs-nbetas)/6*(JBSkew^2 + (1/4)*(JBKurt-3)^2))
  JBpval <- 1-pchisq(JBstat,2)
  return(list("JBstat"=JBstat, "JBpval"=JBpval, "Skewness"=JBSkew, "Kurtosis"=JBKurt))
}
```

Testing Normality of Noise Terms

```
mdlJB <- lm(learn100~lwexp+ltenure, data=dat1)
JBtest <- JB(mdlJB)
fmt <- function(x){format(round(x,4),nsmall=4)}
cat("JB:", fmt(JBtest$JBstat),
    "  p-val:", fmt(JBtest$JBpval),
    "  Skewness:", fmt(JBtest$Skewness),
    "  Kurtosis:", fmt(JBtest$Kurtosis),"\n")
```

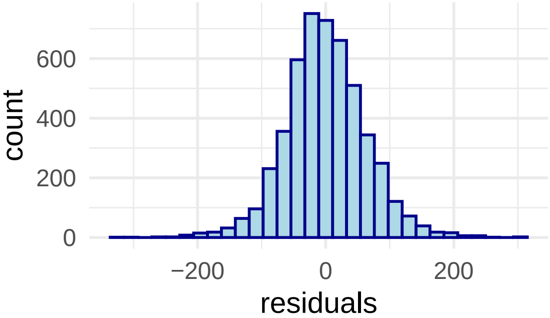
```
JB: 338.6182   p-val: 0.0000   Skewness: 0.0816   Kurtosis: 4.2718
```

The null of normality is rejected

Testing Normality of Noise Terms

Distribution of residuals

- probably no skewness
- ‘excess kurtosis’ (kurtosis in excess of 3)



Roadmap

- (Previous) Session 1: Statistics Review
- (Previous) Session 2: Simple Linear Regression
- (Previous) Session 3: Estimator Standard Errors; Multiple Linear Regression
- (Previous) Session 4: Matrix Algebra
- (Previous) This Session 5: OLS using Matrix Algebra
- **This Session 6: Hypothesis Testing**
- *Next Session 7: Prediction*
- Session 8: Instrumental Variable Regression
- Session 9: Logistic and Other Regressions
- Session 10: Panel Data Regressions
- Session 11: Introduction to Time Series
- Session 12: Time Series Regressions

Appendix

Objective: to show (from slide 19)

$$\hat{\epsilon}_{res}^{\mathrm{T}}\hat{\epsilon}_{res} - \hat{\epsilon}_{ols}^{\mathrm{T}}\hat{\epsilon}_{ols} = (R\hat{\beta}^{ols} - r_0)^{\mathrm{T}}[R(X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}]^{-1}(R\hat{\beta}^{ols} - r_0)$$

- Regression model: $y = X\beta + \epsilon$, and let
- OLS estimator subject to restriction $R\beta = r_0$: $\hat{\beta}^{res}$

We first show

$$\hat{\beta}^{res} = \hat{\beta}^{ols} + (X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}[R(X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}]^{-1}(r_0 - R\hat{\beta}^{ols})$$

where $\hat{\beta}^{ols}$ is the usual unrestricted OLS estimator

Appendix

Restricted SSR minimization problem:

$$\hat{\beta}^{res} = \text{argmin}_{\hat{\beta}}(y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta}) \text{ subject to } R\hat{\beta} - r_0 = 0$$

The Lagrangian is

$$L = (y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta}) + 2(r_0^{\mathrm{T}} - \hat{\beta}^{\mathrm{T}}R^{\mathrm{T}})\lambda$$

The FOC is

$$\left.\frac{\partial L}{\partial \hat{\beta}}\right|_{\hat{\beta}^{res},\hat{\lambda}} = -2X^{\mathrm{T}}y + 2X^{\mathrm{T}}X\hat{\beta}^{res} - 2R^{\mathrm{T}}\hat{\lambda} = 0$$

$$\left.\frac{\partial L}{\partial \hat{\lambda}}\right|_{\hat{\beta}^{res},\hat{\lambda}} = \quad 2(r_0 - R\hat{\beta}^{r}) = 0$$

Appendix

First equation in FOC implies

$$\hat{\beta}^{res} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y + (X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}\hat{\lambda} = \hat{\beta}^{ols} + (X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}\hat{\lambda}$$

Multiplying throughout by R gives

$$R\hat{\beta}^{res} = R\hat{\beta}^{ols} + R(X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}\hat{\lambda}$$

If follows that

$$\begin{aligned}\hat{\lambda} &= [R(X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}]^{-1}(R\hat{\beta}^{res} - R\hat{\beta}^{ols}) \\ &= [R(X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}]^{-1}(r_0 - R\hat{\beta}^{ols})\end{aligned}$$

and therefore

$$\hat{\beta}^{res} = \hat{\beta}^{ols} + (X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}[R(X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}]^{-1}(r_0 - R\hat{\beta}^{ols}) \tag{1}$$

Appendix

Now let

$$\begin{aligned}\hat{\epsilon}_{res} &= y - X\hat{\beta}^{res} \\ &= y - X\hat{\beta}^{ols} + X\hat{\beta}^{ols} - X\hat{\beta}^{res} \\ &= \hat{\epsilon}_{ols} + X(\hat{\beta}^{ols} - \hat{\beta}^{res})\end{aligned} \tag{2}$$

Since (unrestricted) OLS residuals are orthogonal to the regressors, we have

$$\hat{\epsilon}_{ols}^{\mathrm{T}}\hat{\epsilon}_{r} = \hat{\epsilon}_{ols}^{\mathrm{T}}\hat{\epsilon}_{ols} + \hat{\epsilon}_{ols}^{\mathrm{T}}X(\hat{\beta}^{ols} - \hat{\beta}^{res}) = \hat{\epsilon}_{ols}^{\mathrm{T}}\hat{\epsilon}_{ols}$$

Therefore

$$(\hat{\epsilon}_{res} - \hat{\epsilon}_{ols})^{\mathrm{T}}(\hat{\epsilon}_{res} - \hat{\epsilon}_{ols}) = \hat{\epsilon}_{res}^{\mathrm{T}}\hat{\epsilon}_{res} - \hat{\epsilon}_{ols}^{\mathrm{T}}\hat{\epsilon}_{ols} \tag{3}$$

Appendix

Finally, use (1), (2) and (3) to show desired equality

$$
\begin{aligned}
&\hat{\epsilon}_{res}^{\mathrm{T}}\hat{\epsilon}_{res} - \hat{\epsilon}_{ols}^{\mathrm{T}}\hat{\epsilon}_{ols} \\
&= (\hat{\epsilon}_{res} - \hat{\epsilon}_{ols})^{\mathrm{T}}(\hat{\epsilon}_{res} - \hat{\epsilon}_{ols}) \\
&= (\hat{\beta}^{ols} - \hat{\beta}^{res})^{\mathrm{T}} X^{\mathrm{T}} X (\hat{\beta}^{ols} - \hat{\beta}^{res}) \\
&= (R\hat{\beta}^{ols} - r_0)^{\mathrm{T}}[R(X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}]^{-1}R(X^{\mathrm{T}}X)^{-1}(X^{\mathrm{T}}X)(X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}[R(X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}]^{-1} \\
&\qquad \times (R\hat{\beta}^{ols} - r_0) \\
&= (R\hat{\beta}^{ols} - r_0)^{\mathrm{T}}[R(X^{\mathrm{T}}X)^{-1}R^{\mathrm{T}}]^{-1}(R\hat{\beta}^{ols} - r_0)
\end{aligned}
$$