Intermediate Data Analysis & Econometrics

Lecture7

Edward Vytlacil

Yale University, Department of Economics

March 7, 2023

Overview for Today:

Today:

- Examples;
- Asymptotic Normality of OLS Estimator;
- Asymptotic Standard Errors,
- Implementation.
  - Standard errors;
  - Inference on single OLS coefficient;
  - Inference on linear combination of OLS coefficients;
    - using linear algebra
    - by reparameterization
  - Inference on joint null (Wald test).

Future lecture will cover inference using bootstrap,
including for inference on nonlinear functions of coefficients.

Example 1: Treatment with Interactions

Consider the following regression model using results from an RCT:

$$Y_i = \beta_0 + \beta_1 \text{Treat}_i + \beta_2 \text{Fem}_i + \beta_3 \text{Treat}_i \times \text{Fem}_i + \epsilon_i,$$

where

- Treat_i is a dummy variable for receipt of treatment,
- Fem_i is a dummy variable for being female.

For example:

- Treat_i dummy variable denoting receipt of PROGRESA,
- Y_i denoting school enrollment.

Example 1: Treatment with Interactions

Consider the following regression model using results from an RCT:

$$Y_i = \beta_0 + \beta_1 \text{Treat}_i + \beta_2 \text{Fem}_i + \beta_3 \text{Treat}_i \times \text{Fem}_i + \epsilon_i.$$

Can show that OLS estimate of . . .

- $\hat{\beta}_0$ is equivalent to sample mean of Y among untreated men,
- $\hat{\beta}_1$, $\hat{\beta}_2$ is equivalent to differences in sample means,
- $\hat{\beta}_3$ is equivalent to differences in difference in sample means,

OLS regression here equivalent to sample means and difference in sample means.

This example closely related to upcoming diff-in-diff, "Triple Difference" analysis for "Ban the Box: Criminal Records, and Racial Discrimination: A Field Experiment.".

Example 1: Treatment with Interactions

Consider the following regression model using results from an RCT:

$$Y_i = \beta_0 + \beta_1 \text{Treat}_i + \beta_2 \text{Fem}_i + \beta_3 \text{Treat}_i \times \text{Fem}_i + \epsilon_i.$$

Average treatment effect:

- for men: β_1,
- for women: $\beta_1 + \beta_3$,

Gender differences:

- without treatment: β_2,
- with treatment: $\beta_2 + \beta_3$,

Example 1: Treatment with Interactions

Consider the following regression model using results from an RCT:

$$Y_i = \beta_0 + \beta_1 \text{Treat}_i + \beta_2 \text{Fem}_i + \beta_3 \text{Treat}_i \times \text{Fem}_i + \epsilon_i.$$

Consider testing null hypotheses of no treatment effect on

- men, $H_0 : \beta_1 = 0$,
- women, $H_0 : \beta_1 + \beta_3 = 0$,
- men or on women: $H_0 : \beta_1 = 0$ and $\beta_3 = 0$.

Also consider corresponding asymptotic C.I. on β_1, on $\beta_1 + \beta_3$.

Example 1: Treatment with Interactions

Consider the following regression model using results from an RCT:

$$Y_i = \beta_0 + \beta_1 \text{Treat}_i + \beta_2 \text{Fem}_i + \beta_3 \text{Treat}_i \times \text{Fem}_i + \epsilon_i.$$

Consider testing null hypotheses of no gender difference

- ... without treatment, $H_0 : \beta_2 = 0$,
- ... with treatment, $H_0 : \beta_2 + \beta_3 = 0$,
- with or without treatment, $H_0 : \beta_2 = 0$ and $\beta_3 = 0$.

Also consider corresponding asymptotic C.I. on β_2, on $\beta_2 + \beta_3$.

Example 2: Cost Function

Consider the following cost function for electric companies (Nerlove 1963):

$$\log C_i = \beta_0 + \beta_1 \log Q_i + \beta_2 \log PL_i + \beta_3 \log PK_i + \beta_4 \log PF_i + e_i,$$

where

- C_i is total cost,
- Q_i is output,
- PL is unit price of labor,
- PK is unit price of capital,
- PF is unit price of fuel.

How does entering outcome and covariates in logs change interpretation of coefficients?

Example 2: Cost Function

Consider the following cost function for electric companies (Nerlove 1963):

$$\log C_i = \beta_0 + \beta_1 \log Q_i + \beta_2 \log PL_i + \beta_3 \log PK_i + \beta_4 \log PF_i + e_i.$$

Consider $H_0 : \beta_2 + \beta_3 + \beta_4 = 1$ vs $H_1 : \beta_2 + \beta_3 + \beta_4 \neq 1$,

- What is economic meaning of H_0?
- How to test H_0?

Example 3: Mincer Wage Equation

Consider the following model of wages (Mincer 1958):

$$\texttt{ln(wage)}_i = \beta_0 + \beta_1 \texttt{educ}_i + \beta_2 \texttt{exp}_i + \beta_3 \texttt{exp}_i^2 + e_i,$$

where

- $\texttt{educ}_i$ is years of schooling,
- $\texttt{exp}_i$ is years of work experience.

How does entering outcome variable in logs change interpretation of coefficients?

Example 3: Mincer Wage Equation

Consider the following model of wages (Mincer 1958):

$$\mathtt{ln(wage)}_i = \beta_0 + \beta_1 \mathtt{educ}_i + \beta_2 \mathtt{exp}_i + \beta_3 \mathtt{exp}_i^2 + e_i \ .$$

- Model linear in parameters (βs) but non-linear in experience.
- Expect $\beta_2 > 0$, $\beta_3 < 0$, concave function.
- Marginal effect of experience: $\beta_2 + 2\beta_3 \mathtt{exp}_i$.
    - Marginal effect
        - if $\mathtt{exp}_i = 0$ is β_2;
        - if $\mathtt{exp}_i = 10$ is $\beta_2 + 20\,\beta_3$;
        - on average: $\beta_2 + 2\,\beta_3\, E[\mathtt{exp}_i]$.
    - Assuming $\beta_2 > 0$, $\beta_3 < 0$, predicted income highest at $\mathtt{exp}^* = \left|\frac{\beta_2}{2\beta_3}\right|$.

Example 3: Mincer Wage Equation

Consider the following model of wages (Mincer 1958):

$$\texttt{ln(wage)}_i = \beta_0 + \beta_1 \texttt{educ}_i + \beta_2 \texttt{exp}_i + \beta_3 \texttt{exp}_i^2 + e_i \quad .$$

How to:

- Test for linearity in experience, $H_0 : \beta_3 = 0$?
- Test for no effect of experience, $H_0 : \beta_2 = 0$ and $\beta_3 = 0$?
- Confidence interval on marginal effect of experience for given level of experience, $\beta_2 + 2\beta_3 \texttt{exp}$?
- Confidence interval on average marginal effect of experience, $\beta_2 + 2\,\beta_3\, E[\texttt{exp}_i]$?
- Confidence interval on $\texttt{exp}^* = |\frac{\beta_2}{2\beta_3}|$?

Examples

To analyze these examples, we need to be able to perform inference on

- individual coefficients, β_j,
- linear combinations of coefficients,
- multiple restrictions on coefficients (joint inference).

We will use that the OLS estimator is asymptotically normal to apply previous analysis to inference on individual coefficients and on linear combination of coefficients. We will use implication of asymptotic normality for joint inference.

We will cover bootstrap inference including on nonlinear function of coefficients in a future lecture.

Linear Combinations of Coefficients

Previous examples include instances of linear combinations of coefficients. Let

$$a = \begin{pmatrix} a_0, \\ a_1, \\ \vdots \\ a_K \end{pmatrix}, \qquad \beta = \begin{pmatrix} \beta_0, \\ \beta_1, \\ \vdots \\ \beta_K \end{pmatrix},$$

so that

$$a'\beta = \begin{pmatrix} a_0 & a_1 & \dots & a_K \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}$$

$$= a_0\beta_0 + a_1\beta_1 + a_2\beta_2 + \ldots + a_K\beta_K.$$

Linear Combinations of Coefficients

$$a'\beta = a_0\beta_0 + a_1\beta_1 + a_2\beta_2 + \ldots + a_K\beta_K.$$

Many of the previous examples can be expressed as inference on $a'\beta$ for proper choice of a.

For example:

- In example 1,
  - $\beta_2 = a'\beta$ for $a = (0, 0, 1, 0)$,
  - $\beta_2 + \beta_3 = a'\beta$ for $a = (0, 0, 1, 1)$.
- In example 2,
  - $\beta_2 + \beta_3 + \beta_4 = a'\beta$ for $a = (0, 0, 1, 1, 1)$.
- In example 3
  - $\beta_2 + 2\,\beta_3 \text{exp} = a'\beta$ for $a = (0, 0, 1, 2\,\text{exp})$.
  - $\beta_2 + 2\,\beta_3 E[\text{exp}] = a'\beta$ for $a = (0, 0, 1, 2\,E[\text{exp}])$.

Linear Combinations of Coefficients

$$a'\beta = a_0\beta_0 + a_1\beta_1 + a_2\beta_2 + \ldots + a_K\beta_K.$$

- Distinction for inference:
  1. Whether a is known, for example, $a = (0, 0, 1, 1, 1)'$, or
  2. Whether a has to be estimated, for example, $a = (0, 0, 1, 2\,E[\exp])'$.
- For now, we will only consider linear combinations of coefficients, $a'\beta$, for known a.
- We will hold off on analyzing $a'\beta$ when a itself needs to be estimated. We will likewise hold off on nonlinear functions of parameters, as in example 3 with $\left|\frac{\beta_2}{2\beta_3}\right|$. Such problems can be analyzed using the Delta method. We will instead use bootstrap inference later in the course.

Asymptotic Normality of OLS

Assumption:

A1. $Y_i = X_i^T \beta + e_i, \qquad E[X_i \epsilon_i] = 0.$

A2. (X_i, Y_i) are independent and identically distributed ($i.i.d.$).

A3. $0 < E[X_{1i}^4] < \infty$,..., $0 < E[X_{Ki}^4] < \infty$, and $0 < E[Y_i^4] < \infty$.

A4. No perfect multicollinearity.

Theorem

Let $\hat{\beta}_N$ denote the OLS estimator. If Assumptions A1 – A4 hold, then

$$\sqrt{N}\left(\hat{\beta}_N - \beta\right) \xrightarrow{d} N(0, \Sigma).$$

Asymptotic Normality of OLS

Theorem

Let $\hat{\beta}_N$ denote the OLS estimator. If Assumptions A1 – A4 hold, then

$$\sqrt{N}\left(\hat{\beta}_N - \beta\right) \xrightarrow{d} N(0, \Sigma).$$

- The Theorem provides a result for the vector $\hat{\beta}_N$, allows us to approximate distribution of $\hat{\beta}_N$ as a vector.
- How to translate the result into an approximation for one component of $\hat{\beta}_N$? For a linear combination of elements of $\hat{\beta}_N$, i.e., for $a'\hat{\beta}_N$?

Digression: Bivariate Normal

Suppose $Z = (Z_1, Z_2)'$ is *Bivariate Normal*,

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \right).$$

Then, for any constants a_1, a_2,

$$a_1 Z_1 + a_2 Z_2 \sim N\left(a_1\mu_1 + a_2\mu_2, a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\sigma_{12}\right).$$

For example:

- $Z_1 \sim N(\mu_1, \sigma_1^2)$,
- $Z_2 \sim N(\mu_2, \sigma_2^2)$,
- $Z_2 - Z_1 \sim N(\mu_2 - \mu_1, \sigma_2^2 + \sigma_1^2 - 2\sigma_{12})$,
- $Z_2 + Z_1 \sim N(\mu_2 + \mu_1, \sigma_2^2 + \sigma_1^2 + 2\sigma_{12})$.

Digression: Bivariate Normal

Suppose $Z = (Z_1, Z_2)'$ is *Bivariate Normal*, in vector notation,

$$\mathbf{z} = \sim N(\mu, \Sigma) .$$

Then, for any $a = (a_1, a_2)'$,

$$a'Z \sim N(a'\mu, a'\Sigma a) .$$

For example:

- $Z_1 \sim N(\mu_1, \sigma_1^2)$ special case with $a = (1, 0)'$,
- $Z_2 \sim N(\mu_2, \sigma_2^2)$ special case with $a = (0, 1)'$,
- $Z_2 - Z_1 \sim N(\mu_2 - \mu_1, \sigma_2^2 + \sigma_1^2 - 2\sigma_{12})$ special case with $a = (-1, 1)'$,
- $Z_2 + Z_1 \sim N(\mu_2 + \mu_1, \sigma_2^2 + \sigma_1^2 + 2\sigma_{12})$ special case with $a = (1, 1)'$.

Digression: Multivariate Normal

Suppose $Z = (Z_1, Z_2, ..., Z_K)'$ is *Multivariate Normal*, in vector notation,

$$\mathbf{Z} \sim N(\mu, \Sigma).$$

Z multivariate normal implies that, for $a = (a_1, a_2, ..., a_K)'$,

$$a'Z \sim N(a'\mu, a'\Sigma a).$$

Digression: Multivariate Normal

Suppose $Z = (Z_1, Z_2, \ldots, Z_K)'$ is *Multivariate Normal*, in vector notation,

$$\mathbf{z} \sim N(\mu, \Sigma).$$

Z multivariate normal implies that, for $a = (a_1, a_2, \ldots, a_K)'$,

$$a'Z \sim N(a'\mu, a'\Sigma a).$$

Can likewise show

$$\sqrt{N}\left(\hat{\beta}_N - \beta\right) \xrightarrow{d} N(0, \Sigma)$$

implies that

$$\sqrt{N}\left(a'\hat{\beta}_N - a'\beta\right) \xrightarrow{d} N(0, a'\Sigma a).$$

Digression: Multivariate Normal

$$\sqrt{N}\left(a'\hat{\beta}_N - a'\beta\right) \xrightarrow{d} N(0, a'\Sigma a)$$

- Thus, can conduct asymptotic inference on both individual components of β and any linear combination of components of β just as we did for sample mean based on CLT, just need valid standard errors.
- Let $\hat{\Sigma}_N$ denote a consistent estimator of Σ, i.e., such that $\hat{\Sigma}_N \xrightarrow{p} \Sigma$.
- Approximate $\text{Var}(\hat{\beta}_N)$ by $\hat{\Sigma}/N$.
- Approximate $\text{Var}(a'\hat{\beta}_N)$ by $a'\hat{\Sigma}a/N$.
- Construct standard errors on $a'\hat{\beta}_N$ by $\frac{\hat{\omega}_N}{\sqrt{N}}$ where $\hat{\omega}_N^2 = a'\hat{\Sigma}_N a$.
- How to construct $\hat{\Sigma}$?

Estimating OLS Asymptotic Variance Under Homoscedasticity

- Homoscedasticity a very strong assumption that typically is not plausible in economics.
- The default estimator of the OLS asymptotic variance and thus the resulting standard errors in most statistical software (e.g. Stata's `reg` and R's `lm` command) are based on the assumption of *homoscedasticity*.
- If homoscedasticity is violated, than conventional s.e. based on homoscedasticity will not be consistent.

Homoscedasticity vs Heteroscedastic Robust s.e.

- If homoscedasticity is violated, we say that the model is heteroscedastic.
- We can use alternative (non-default) estimators of Σ that do not rely on homoscedasticity. We say that resulting standard errors are "robust" to heteroscedasticity.
- There is a tradeoff:
  - heteroscedastic robust s.e. require larger sample sizes to be reasonably accurate;
  - if sample size is relatively small and homoscedasticity is not implausible, may prefer to use default s.e. instead of robust s.e..

Example 1

Model: $Y_i = \beta_0 + \beta_1 \text{Treat}_i + \beta_2 \text{Fem}_i + \beta_3 \text{Treat}_i \times \text{Fem}_i + \epsilon_i$.

```
> reg.2 <- lm(school ~ treat * sex1, data = subset(df,wave==5))
> summary(reg.2)

Call:
lm(formula = school ~ treat * sex1, data = subset(df, wave == 5))

Residuals:
    Min      1Q  Median      3Q     Max
-0.7889  0.2111  0.2168  0.2382  0.2710

Coefficients:
             Estimate Std. Error t value            Pr(>|t|)
(Intercept)  0.766323   0.007728  99.157 < 0.0000000000000002 ***
treat        0.021652   0.009712   2.229             0.02581 *
sex1        -0.033434   0.010988  -3.043             0.00235 **
treat:sex1   0.038429   0.013886   2.767             0.00566 **
---
```

These default standard errors are only valid under homoscedasticity.

How to get heterscedastic robust s.e.?

Variance Estimation

By default, R estimates Σ and thus Σ/N using the `vcov` function which relies on homoscedasticity. We typically instead use the `vcovHC` function of the `sandwich` package which provides heteroscedastic-robust results.

```
> library(sandwich)
> vcovHC(reg.2)
               (Intercept)          treat           sex1      treat:sex1
(Intercept)  0.00006161803 -0.00006161803 -0.00006161803  0.00006161803
treat       -0.00006161803  0.00009512944  0.00006161803 -0.00009512944
sex1        -0.00006161803  0.00006161803  0.00013037510 -0.00013037510
treat:sex1   0.00006161803 -0.00009512944 -0.00013037510  0.00019955386

> se.robust.2 <- sqrt(diag(vcovHC(reg.2)))
> se.robust.2
(Intercept)       treat        sex1  treat:sex1
0.007849716 0.009753432 0.011418192 0.014126353
```

Note that `vcovHC` is providing estimates of Σ/N, not estimates of Σ.

Variance Estimation

Alternatively, use `coeftest` from the package *lmtest* with `vcovHC` option to use heteroscedastic-robust s.e..

```
> library(sandwich)
> library(lmtest)
> reg.test.2 <- coeftest(reg.2,vcov = vcovHC)
> reg.test.2

t test of coefficients:

              Estimate  Std. Error t value              Pr(>|t|)
(Intercept)  0.7663230  0.0078472 97.6552 < 0.00000000000000022 ***
treat        0.0216523  0.0097393  2.2232              0.026217 *
sex1        -0.0334343  0.0114167 -2.9286              0.003410 **
treat:sex1   0.0384286  0.0141079  2.7239              0.006459 **
---

> se.robust.2 <- reg.test.2[,2]
> se.robust.2
(Intercept)       treat        sex1  treat:sex1
0.007847235 0.009739271 0.011416663 0.014107936
```

How to get s.e. on estimated effect for girls, i.e., on $\hat{\beta}_1 + \hat{\beta}_3$?

Variance Estimation

How to get s.e. on estimated effect for girls?
One way: use formula $\text{Var}(\hat{\beta}_1 + \hat{\beta}_3) = \text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_3) + 2\text{Cov}(\hat{\beta}_1, \hat{\beta}_3)$,
approximate by $\hat{\sigma}_2^2 + \hat{\sigma}_4^2 + 2\hat{\sigma}_{2,4}$.

```
> vcovHC(reg.2)
               (Intercept)           treat            sex1      treat:sex1
(Intercept)  0.00006161803 -0.00006161803 -0.00006161803  0.00006161803
treat       -0.00006161803  0.00009512944  0.00006161803 -0.00009512944
sex1        -0.00006161803  0.00006161803  0.00013037510 -0.00013037510
treat:sex1   0.00006161803 -0.00009512944 -0.00013037510  0.00019955386

 > sqrt(vcovHC(reg.2)[2,2]+vcovHC(reg.2)[4,4] + 2 * vcovHC(reg.2)[2,4])
[1] 0.01021883
```

Variance Estimation

Alternatively, same math with linear algebra notation, approximate $\mathrm{Var}\left(a'\hat{\beta}_N\right)$ by $a'\hat{\Sigma}a$ with $a = \left(0, 1, 0, 1\right)'$.

```
> vcovHC(reg.2)
               (Intercept)           treat            sex1      treat:sex1
(Intercept)  0.00006161803 -0.00006161803 -0.00006161803  0.00006161803
treat       -0.00006161803  0.00009512944  0.00006161803 -0.00009512944
sex1        -0.00006161803  0.00006161803  0.00013037510 -0.00013037510
treat:sex1   0.00006161803 -0.00009512944 -0.00013037510  0.00019955386
> a <- c(0,1,0,1)
> as.numeric(sqrt(t(a)%*%vcovHC(reg.2)%*%a))
[1] 0.01021883
```

Variance Estimation

Alternatively, can "reparameterize" model to get one coefficient equal to the effect for girls.
Including a dummy for boys and not girls, coefficient on "treat" is now estimated effect for girls.

```
> reg.3 <- lm(school ~ treat * boy, data = subset(df, wave==5))
> reg.test.3 <- coeftest(reg.3,vcov = vcovHC)
> reg.test.3

t test of coefficients:

              Estimate Std. Error t value              Pr(>|t|)
(Intercept)  0.7328887  0.0082922 88.3826 < 0.00000000000000022 ***
treat        0.0600809  0.0102069  5.8863        0.000000004031 ***
boy          0.0334343  0.0114167  2.9286              0.003410 **
treat:boy   -0.0384286  0.0141079 -2.7239              0.006459 **
---
```

how to interpret each coefficient in reparameterized model?

Reporting Results

Reporting results

```
> varlabels.3 <- c("Intercept","Treatment","Girl","Girl*Treatment","Boy",
+                             "Boy*Treatment")
> stargazer(reg.test.1,reg.test.2,reg.test.3,#using output from coeftest, het. robust
+           intercept.bottom = FALSE, #intercept at top, not bottom
+           dep.var.labels="Enrollment", #label dependent variable
+           covariate.labels=varlabels.3, #label regressors with labels defined above
+           type="latex", #output  "latex", can also use "html" or plain text
+           keep.stat=c("n","rsq"),  # what statistics to print
+           notes.append = FALSE, notes.align = "l",
+           notes = "Reporting heteroscedastic-robust standard errors in parenthesis")
```

- Note that the first line of `stargzer` inputing the heteroscedastic-robust output from `coeftest`, not directly the output from `lm`.
- See "Stargazer Handout." for more details on making regression tables with `stargazer`.

Reporting Results

Table 7.1

	Dependent variable:		
	Enrollment		
	(1)	(2)	(3)
Intercept	0.750^{***}	0.766^{***}	0.733^{***}
	(0.006)	(0.008)	(0.008)
Treatment	0.041^{***}	0.022^{**}	0.060^{***}
	(0.007)	(0.010)	(0.010)
Girl		-0.033^{***}	
		(0.011)	
Girl*Treatment		0.038^{***}	
		(0.014)	
Boy			0.033^{***}
			(0.011)
Boy*Treatment			-0.038^{***}
			(0.014)
Note:	Reporting heteroscedastic-robust standard errors in parenthesis		

Reporting Results

- Conventional to report p-values corresponding to null hypothesis that each coefficient equals zero versus two-sided alternative.
- Default p-values from `lm` use default s.e., not valid if model is heteroscedastic.
- Can use `coeftest` with `vcovHC` option to use heteroscedastic-robust s.e.. and p-values based on those heteroscedastic-robust s.e..
- Note inclusion of ‘p‘ in option `report=('vc*sp')` in `stargazer` to obtain p-values

```
> reg.test.1<-coeftest(reg.1,vcov = vcovHC)
> reg.test.2<-coeftest(reg.2,vcov = vcovHC)
> reg.test.3<-coeftest(reg.3,vcov = vcovHC)
> stargazer(reg.test.1,reg.test.2,reg.test.3,
+           intercept.bottom = FALSE, #intercept at top, not bottom
+           report=('vc*sp'), #v is variable name, c is coefficient with **, s is s.e.
    and p is pvalue
+           dep.var.labels="Enrollment", #label dependent variable
+           covariate.labels=varlabels.3,  #label regressors with labels defined above
+           type="latex", #output  "latex", can also use "html" or plain text
+           keep.stat=c("n","rsq"),  # what statistics to print
+           notes.append = FALSE, notes.align = "l",
+           notes = "Reporting heteroscedastic-robust standard errors in parenthesis."
   )
```

Reporting Results

Table 7.2

	Dependent variable:		
	Enrollment		
	(1)	(2)	(3)
Intercept	0.750***	0.766***	0.733***
	(0.006)	(0.008)	(0.008)
	p = 0.000	p = 0.000	p = 0.000
Treatment	0.041***	0.022**	0.060***
	(0.007)	(0.010)	(0.010)
	p = 0.000	p = 0.027	p = 0.000
Girl		$-$0.033***	
		(0.011)	
		p = 0.004	
Girl*Treatment		0.038***	
		(0.014)	
		p = 0.007	
Boy			0.033***
			(0.011)
			p = 0.004
Boy*Treatment			$-$0.038***
			(0.014)
			p = 0.007

Example 1: Test Null of Zero Treatment Effect on either Girls or Boys

$$Y_i = \beta_0 + \beta_1 \text{Treat}_i + \beta_2 \text{Fem}_i + \beta_3 \text{Treat}_i \times \text{Fem}_i + \epsilon_i.$$

Consider testing null hypotheses of no treatment effect on either men or on women, vs the alternative that it has an effect on men or an effect on women (or both):

- $H_0 : \beta_1 = 0$ **and** $\beta_3 = 0,$ vs.
- $H_1 : \beta_1 \neq 0$ **or** $\beta_3 \neq 0$ **or** both.

This is an example of a *joint null hypothesis*.

Related to, but distinct from, multiple hypothesis testing.

Ex. 1: Test Null of Zero Treatment Effect on either Girls or Boys

- H_0 : $\beta_1 = 0$ **and** $\beta_3 = 0$, vs.
- H_1 : $\beta_1 \neq 0$ **or** $\beta_3 \neq 0$ **or** both.

What if we wished to test the null hypothesis with significance level 5%?

Ex. 1: Test Null of Zero Treatment Effect on either Girls or Boys

- $H_0 : \beta_1 = 0$ **and** $\beta_3 = 0,$ vs.
- $H_1 : \beta_1 \neq 0$ **or** $\beta_3 \neq 0$ **or** both.

What if we wished to test the null hypothesis with significance level 5%?

Naive, and incorrect strategy:

- Test null $H_{01} : \beta_1 = 0$, vs. $H_{11} : \beta_1 \neq 0$, at significance level 5%.
- Test null $H_{03} : \beta_3 = 0$, vs. $H_{13} : \beta_3 \neq 0$, at significance level 5%.
- Reject joint null if you reject either $\beta_1 = 0$ at 5% level, or reject null $\beta_3 = 0$ at 5% level.

Resulting test does *NOT* have significance level 5% in general, typically probability of rejecting at least one null when both are true will be strictly more than 0.05.

Joint Null Hypothesis: Wald Test

- For inference on joint null hypotheses, can use Wald test.
- Wald test based on asymptotic normality of OLS coefficients, and using estimated variance, but has a different test statistic (a quadratic form for test statistic) with different asymptotic distribution (chi-square distribution).
- We will not cover the theory for a Wald test, but how to implement in **R** using `linearHypothesis` from package *car*.

Example 1: Joint Null of No Effect on Girls of Boys, Using Wald Test

$$Y_i = \beta_0 + \beta_1 \text{Treat}_i + \beta_2 \text{Fem}_i + \beta_3 \text{Treat}_i \times \text{Fem}_i + \epsilon_i,$$

```
> library(car)
> linearHypothesis(reg.2, c("treat=0","treat:sex1 = 0"), test="Chisq",  vcov=vcovHC)

Linear hypothesis test

Hypothesis:
treat = 0
treat:sex1 = 0

Model 1: restricted model
Model 2: school ~ treat * sex1

Note: Coefficient covariance matrix supplied.

  Res.Df Df  Chisq     Pr(>Chisq)
1  15417
2  15415  2 39.591 0.000000002529 ***
---
```

Notice using heteroscedastic-robust standard errors for test by specifying option `vcov=vcovHC`.

Wald Test for Joint Null

- Possible to reject an individual null at 5% level without rejecting joint null at 5% level.
- Possible to reject an joint null at 5% level without rejecting any individual null at 5% level.
- How to interpret?

Example 2: Cost Function

$$\log C_i = \beta_0 + \beta_1 \log Q_i + \beta_2 \log PL_i + \beta_3 \log PK_i + \beta_4 \log PF_i + e_i.$$

```
> library("haven")
> df.N <- read_dta("Nerlove1963.dta")
> reg.c <- lm(log(cost) ~ log(output) + log(Plabor) + log(Pcapital) +log(Pfuel),
                       data=df.N)
> coeftest(reg.c,vcov = vcovHC)

t test of coefficients:

               Estimate Std. Error t value               Pr(>|t|)
(Intercept)   -3.526503   1.794226 -1.9655                0.05134 .
log(output)    0.720394   0.034052 21.1559 < 0.0000000000000000022 ***
log(Plabor)    0.436341   0.254044  1.7176                0.08808 .
log(Pcapital) -0.219888   0.337508 -0.6515                0.51579
log(Pfuel)     0.426517   0.078022  5.4666           0.0000002044 ***
---
```

How to test $H_0 : \beta_2 + \beta_3 + \beta_4 = 1$?

Example 2: Cost Function

Test:

- $H_0 : \beta_2 + \beta_3 + \beta_4 = 1,$ vs
- $H_1 : \beta_2 + \beta_3 + \beta_4 \neq 1.$

```
> a <- c(0,0,1,1,1)
> theta.hat<-as.numeric(t(a)%*%summary(reg.c)$coefficients[,1])
> se.theta.hat <- as.numeric(sqrt(t(a)%*%vcovHC(reg.c)%*%a))
> test_stat<- abs((theta.hat-1)/se.theta.hat)
> #p-value, two-sided alternative
> 2 * (1 - pnorm(test_stat))
[1] 0.4419
```

Example 2: Reparameterization

- We saw in example 1 that we could directly obtain the estimated effect for girls and corresponding standard errors/p-values/CI by including a dummy variable for being a boy instead of for being a girl.
- Doing so *reparameterized* the model.
- Can we do so here?

Example 2: Reparameterization

$$\log c_i = \beta_0 + \beta_1 \log Q_i + \beta_2 \log PL_i + \beta_3 \log PK_i + \beta_4 \log PF_i + e_i.$$

- Want to test
  - $H_0 : \beta_2 + \beta_3 + \beta_4 = 1$, vs
  - $H_1 : \beta_2 + \beta_3 + \beta_4 \neq 1$.

Example 2: Reparameterization

$$\log c_i = \beta_0 + \beta_1 \log Q_i + \beta_2 \log PL_i + \beta_3 \log PK_i + \beta_4 \log PF_i + e_i.$$

- Want to test
  - $H_0 : \beta_2 + \beta_3 + \beta_4 = 1$, vs
  - $H_1 : \beta_2 + \beta_3 + \beta_4 \neq 1$.
- Define $\theta = (\beta_2 + \beta_3 + \beta_4) - 1$.
- Can restate our null and alternative hypothesis as:
  - $H_0 : \theta = 0$,
  - $H_1 : \theta \neq 0$.

Example 2: Reparameterization

$$\log c_i = \beta_0 + \beta_1 \log Q_i + \beta_2 \log PL_i + \beta_3 \log PK_i + \beta_4 \log PF_i + e_i.$$

- Want to test
  - $H_0 : \beta_2 + \beta_3 + \beta_4 = 1$, vs
  - $H_1 : \beta_2 + \beta_3 + \beta_4 \neq 1$.
- Define $\theta = (\beta_2 + \beta_3 + \beta_4) - 1$.
- Can restate our null and alternative hypothesis as:
  - $H_0 : \theta = 0$,
  - $H_1 : \theta \neq 0$.
- Can we write a new, equivalent regression such that θ directly appears as a coefficient?

Example 2: Reparameterization

Can show

$$\Rightarrow \log C_i - \log PL_i = \beta_0 + \beta_1 \log Q_i + \theta \log PL_i + \beta_3(\log PK_i - \log PL_i) + \beta_4(\log PF_i - \log PL_i) + e_i.$$

Proof? blackboard intermission . . .

Example 2: Reparameterization

Can show

$$\Rightarrow \log C_i - \log PL_i = \beta_0 + \beta_1 \log Q_i + \theta \log PL_i + \beta_3(\log PK_i - \log PL_i) + \beta_4(\log PF_i - \log PL_i) + e_i.$$

Proof? blackboard intermission . . .

Suggests test null $H_0 : \beta_2 + \beta_3 + \beta_4 = 1$ by testing whether coefficient on $\log PL_i$ is zero in reparameterized regression.

- Important: Interpretation of each coefficient depends on what other regressors are in the model!

Example 2: Reparameterization

$$\log C_i - \log PL_i = \beta_0 + \beta_1 \log Q_i + \theta \log PL_i + \beta_3(\log PK_i - \log PL_i) + \beta_4(\log PF_i - \log PL_i) + e_i.$$

```
> #reparameterized regression
> reg.r <- lm(I(log(cost)-log(Plabor)) ~ log(output) + log(Plabor)
                     + I(log(Pcapital)-log(Plabor)) +  I(log(Pfuel)-Plabor),
                       data=df.N)
> reg.r

Call:
lm(formula = I(log(cost) - log(Plabor)) ~ log(output) + log(Plabor) +
    I(log(Pcapital) - log(Plabor)) + I(log(Pfuel) - Plabor),   data = df.N)

Coefficients:
                   (Intercept)                     log(output)
                      -3.29578                         0.72119

                   log(Plabor)  I(log(Pcapital) - log(Plabor))
                       0.04695                        -0.21000

        I(log(Pfuel) - Plabor)
                       0.42898
```

Example 2: Reparameterization

Use `coeftest`, report results.

```
> reg.test.c<-coeftest(reg.c,vcov = vcovHC)
> reg.test.r<-coeftest(reg.r,vcov = vcovHC)
>
> varlabels.c <- c("Intercept","log(Output)","log(PriceLabor)","log(PriceCapital)","
    log(PriceFuel)","log(PriceCapital)-log(PriceLabor)","log(PriceFuel)-log(
    PriceLabor)")
>
> stargazer(reg.test.c,reg.test.r,
+           intercept.bottom = FALSE, #intercept at top, not bottom
+           report=('vc*sp'), #v is variable name, c is coefficient with **, s is s.e.
     and p is pvalue
+           dep.var.labels=c("log(Cost)", "log(Cost) - log(PriceLabor)"), #label
    dependent variable
+           covariate.labels=varlabels.c,  #label regressors with labels defined above
+           type="latex", #output  "latex",
+           keep.stat=c("n","rsq"),  # what statistics to print
+           notes.append = FALSE, notes.align = "l",
+           notes = "Reporting heteroscedastic-robust standard errors in parenthesis."
    )
```

How to use resulting table to test $H_0 : \beta_1 + \beta_2 + \beta_3 = 1$?

Reporting Results

	Dependent variable:	
	log(Cost)	log(Cost) - log(PriceLabor)
	(1)	(2)
Intercept	-3.527^{*}	-3.527^{*}
	(1.794)	(1.794)
	p = 0.052	p = 0.052
log(Output)	0.720^{***}	0.720^{***}
	(0.034)	(0.034)
	p = 0.000	p = 0.000
log(PriceLabor)	0.436^{*}	-0.357
	(0.254)	(0.464)
	p = 0.089	p = 0.444
log(PriceCapital)	-0.220	
	(0.338)	
	p = 0.516	
log(PriceFuel)	0.427^{***}	
	(0.078)	
	p = 0.00000	
log(PriceCapital)-log(PriceLabor)		-0.220
		(0.338)
		p = 0.516
log(PriceFuel)-log(PriceLabor)		0.427^{***}
		(0.078)
		p = 0.00000
Observations	145	145
R^2	0.926	0.925
Note:	Reporting heteroscedastic-robust standard errors in parenthesis.	

Example 3: Mincer Wage Equation

$$\mathtt{ln(wage)}_i = \beta_0 + \beta_1 \mathtt{educ}_i + \beta_2 \mathtt{exp}_i + \beta_3 \mathtt{exp}_i^2 + e_i.$$

```
> library(AER)
> data(CPS1988)
> df.cps <-CPS1988
> reg.c0 <- lm( log(wage) ~ education + experience + I(experience^2), data = df.cps )
> coeftest(reg.c0,vcov = vcovHC)

t test of coefficients:

                   Estimate    Std. Error t value              Pr(>|t|)
(Intercept)      4.278089739  0.020532241 208.360 < 0.00000000000000022 ***
education        0.087441104  0.001377010  63.501 < 0.00000000000000022 ***
experience       0.077520255  0.001017524  76.185 < 0.00000000000000022 ***
I(experience^2) -0.001315969  0.000023434 -56.156 < 0.00000000000000022 ***
```

Estimate average effect of experience, $\beta_2 + 2\beta_3 \mathbb{E}[\text{exp}]$?

Estimate experience level that maximizes wages, $\left|\frac{\beta_2}{2\beta_3}\right|$?

Example 3: Mincer Wage Equation

Use $\hat{\beta}_2 + 2\hat{\beta}_3\overline{\text{exp}}$ to estimate average effect of experience, $\beta_2 + 2\beta_3\mathbb{E}[\text{exp}]$:

```
> mean.exp <- mean(df.cps$experience)
> #estimated average effect experience
> avgeff.exp <- reg.c0$coefficients[[3]]+2*reg.c0$coefficients[[4]]*mean.exp
> avgeff.exp
[1] 0.02961916
```

$\hat{\beta}_2 + 2\hat{\beta}_3\overline{\text{exp}}$ is a consistent estimator of $\beta_2 + 2\beta_3\mathbb{E}[\text{exp}]$.

(why?)

Example 3: Mincer Wage Equation

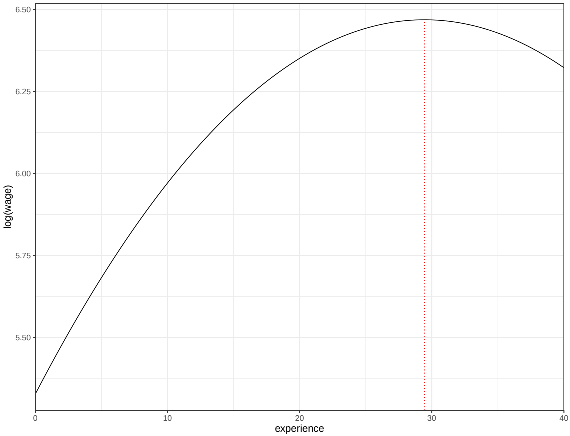
Use $\left|\frac{\hat{\beta}_2}{2\hat{\beta}_3}\right|$ to estimate experience level that maximizes wages, $\left|\frac{\beta_2}{2\beta_3}\right|$:

```
> maxexp <- abs(reg.c0$coefficients[[3]]/(2*reg.c0$coefficients[[4]]))
> maxexp
[1] 29.45367
> ggplot(data.frame(x = c(0, 40)), aes(x = x))  +
+     ylab("log(wage)")+ theme_bw(base_size = 14) +
+     scale_x_continuous(name="experience",
+                        expand = c(0, 0) )+
+     scale_y_continuous(name="log(wage)",
+                        limits=c(f0(0)-.05,f0(maxexp)+.05),
+                        expand = c(0,0) )+
+     stat_function(fun = f0)+
+     geom_segment(mapping=aes(x=maxexp, y=f0(0)-.05,
+                       xend=maxexp, yend=f0(maxexp)),linetype="dotted",
+                      color = "red")
```

$\left|\frac{\hat{\beta}_2}{2\hat{\beta}_3}\right|$ is a consistent estimator of $\left|\frac{\beta_2}{2\beta_3}\right|$ as long as $\beta_3 \neq 0$.

(why?)

Example 3: Mincer Wage Equation



Log(Wage) as function of experience from estimated Mincer Wage Equation using CPS 1988 Data.

Example 3: Mincer Wage Equation

- Experience level that maximizes wages, $\left|\frac{\beta_2}{2\beta_3}\right|$, is a *non-linear* function of coefficients.
  - Can use Delta method for inference (not covered in this course).
  - Can use bootstrap method for inference (we will cover in this course).

Average effect of experience, $\beta_2 + 2\beta_3\mathbb{E}[\text{exp}]$.

- Can again use Delta method (not covered) or bootstrap method (we will cover) for inference accounting for estimation noise in $\mathbb{E}[\text{exp}]$.
- For now, consider inference ignoring estimation noise in $\mathbb{E}[\text{exp}]$. We consider using reparameterization, although we could alternatively use $a'\hat{\Sigma}a$ type formula with $a = (0, 0, 1, 2\overline{\text{exp}})$.

Example 3: Reparamerterization

- Estimate average effect of experience to be $\hat{\beta}_2 + 2\hat{\beta}_3\, \overline{\text{exp}} = 0.029619$.
- Can reparameterize model:

$$\ln(\text{wage})_i = \delta_0 + \delta_1 \text{educ}_i + \delta_2(\text{exp}_i - \overline{\text{exp}}) + \delta_3(\text{exp}_i - \overline{\text{exp}})^2 + e_i.$$

where OLS $\hat{\delta}_2 = \hat{\beta}_2 + 2\hat{\beta}_3\, \overline{\text{exp}} = 0.029619$.

```
> reg.c1 <-
    lm( log(wage) ~ education + I(experience-mean.exp) + I((experience-mean.exp) ^2),
    data = df.cps )
> reg.c1

Call:
lm(formula = log(wage) ~ education + I(experience - mean.exp) +
    I((experience - mean.exp)^2), data = df.cps)

Coefficients:
  (Intercept)      education  I(experience - mean.exp)  I((experience - mean.exp)^2)
     5.253055       0.087441                  0.029619                     -0.001316
```

Example 3: Reparamerterization

- Estimate average effect of experience to be $\hat{\beta}_2 + 2\hat{\beta}_3\,\bar{\exp} = 0.029619$.
- Can reparameterize model:

$$\texttt{ln(wage)}_i = \delta_0 + \delta_1 \texttt{educ}_i + \delta_2(\texttt{exp}_i - \bar{\texttt{exp}}) + \delta_3(\texttt{exp}_i - \bar{\texttt{exp}})^2 + e_i.$$

where OLS $\hat{\delta}_2 = \hat{\beta}_2 + 2\hat{\beta}_3\,\bar{\exp} = 0.029619$.

- In reparameterized model, $\hat{\delta}_2$ gives estimated average effect of experience. However, resulting s.e./p-value/CI do not incorporate estimation error of $\bar{\exp}$.

```
> coeftest(reg.c1,vcov = vcovHC)
t test of coefficients:
                            Estimate   Std. Error t value                 Pr(>|t|)
(Intercept)              5.253054654  0.018948894 277.222 < 0.0000000000000002
education                0.087441104  0.001377010  63.501 < 0.0000000000000002
I(experience - mean.exp)  0.029619165  0.000332123  89.181 < 0.0000000000000002
I((experience - mean.exp)^2) -0.001315969  0.000023434 -56.156 < 0.0000000000000002
```

Example 3: Reparamerterization

	Dependent variable:	
	log Wage	
	(1)	(2)
Intercept	4.278^{***}	5.253^{***}
	(0.021)	(0.019)
	p = 0.000	p = 0.000
Education	0.087^{***}	0.087^{***}
	(0.001)	(0.001)
	p = 0.000	p = 0.000
Experience	0.078^{***}	
	(0.001)	
	p = 0.000	
Experience-Squared	-0.001^{***}	
	(0.00002)	
	p = 0.000	
Experience-Mean(Experience)		0.030^{***}
		(0.0003)
		p = 0.000
(Experience-Mean(Experience))-Squared		-0.001^{***}
		(0.00002)
		p = 0.000
Avg Effect Experience	0.030	0.030
Observations	28,155	28,155
R^2	0.326	0.326
Note:	Reporting heteroscedastic-robust standard errors in parenthesis. Inference does not adjust for estimation error in sample mean of experience.	

Example 3: Mincer Wage Equation

$$\texttt{ln(wage)}_i = \beta_0 + \beta_1 \texttt{educ}_i + \beta_2 \texttt{exp}_i + \beta_3 \texttt{exp}_i^2 + e_i.$$

Test joint null that experience has no effect, i.e., that $\beta_2 = \beta_3 = 0$?

```
 >  linearHypothesis(reg.c2, c("experience=0","I(experience^2) = 0"), test = "Chisq",
      vcov=vcovHC)

Linear hypothesis test

Hypothesis:
experience = 0
I(experience^2) = 0

Model 1: restricted model
Model 2: log(wage) ~ education + experience + I(experience^2)

Note: Coefficient covariance matrix supplied.

  Res.Df Df   Chisq             Pr(>Chisq)
1   28150
2   28148   2 8701.1 < 0.0000000000000022 ***
---
```

Summary

- Under appropriate assumptions, OLS is an asymptotically normal estimator, can perform inference as with sample mean based on asymptotic normality.

Summary

- Under appropriate assumptions, OLS is an asymptotically normal estimator, can perform inference as with sample mean based on asymptotic normality.
- In **R**, use `lm` to estimate linear regression model, but default standard errors, p-values, and confidence intervals are based on homoscedasticity. Typically want to use heteroscedastic-robust standard errors, and corresponding p-values, and confidence intervals.

Summary

- Under appropriate assumptions, OLS is an asymptotically normal estimator, can perform inference as with sample mean based on asymptotic normality.
- In **R**, use `lm` to estimate linear regression model, but default standard errors, p-values, and confidence intervals are based on homoscedasticity. Typically want to use heteroscedastic-robust standard errors, and corresponding p-values, and confidence intervals.
- For single coefficients:
  - can find heteroscedastic-robust s.e. using `vcovHC`;
  - Can compute asymptotic p-values directly, or, for null that coefficient is zero, can use `coeftest`;
  - Can compute asymptotic confidence interval directly, or can use `coefci`.

Summary (cont'd)

- For nulls of a linear combination of coefficients,
  - Using `vcovHC` and linear algebra, can directly calculate s.e., p-value, and CI.
  - Can alternatively run reparametrerized regression model in which one coefficient of new regression model equals desired linear combination of coefficients from original model.
  - Doing either of the above when the weights on linear combination of coefficients need to be estimated does not account for estimation error in the weights, we will return to this issue.

Summary (cont'd)

- For nulls of a linear combination of coefficients,
  - Using `vcovHC` and linear algebra, can directly calculate s.e., p-value, and CI.
  - Can alternatively run reparametrerized regression model in which one coefficient of new regression model equals desired linear combination of coefficients from original model.
  - Doing either of the above when the weights on linear combination of coefficients need to be estimated does not account for estimation error in the weights, we will return to this issue.
- For joint hypothesis tests can use `linearHypothesis` from package *car*, again using `vcovHC` for heteroscedastic-robust standard errors.

Summary (cont'd)

- For non-linear functions of coefficients, such as $\left|\frac{\hat{\beta}_2}{2\beta_3}\right|$, natural plug-in estimator is consistent, but standard errors/p-values/CI require alternative methods (Delta method or bootstrap), we will return to this issue when we cover bootstrap.

- Additional resource for this lecture:
  "Stargazer Handout."