EC 508: Econometrics

Midterm Study Guide

Alex Hoagland, Boston University

March 7, 2019

Model:

$$\underbrace{y_{it}}_{\text{Dependent variable}} = \underbrace{\beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \ldots + \beta_k x_{kit} + u_{it}}_{Independent variables/regressors}$$

Model:

$$\underbrace{y_{it}}_{\text{Dependent variable}} = \underbrace{\beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \ldots + \beta_k x_{kit} + u_{it}}_{Independent variables/regressors}$$

- N observations—**randomly sampled**
  - Why do we need random sampling?
  - How might a sampling procedure violate random sampling?
- T periods—time series data
  - If $T = 1$ (fixed point in time), the data is **cross-sectional data**

Model:

$$\underbrace{y_{it}}_{\text{Dependent variable}} = \underbrace{\beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \ldots + \beta_k x_{kit} + u_{it}}_{Independent variables/regressors}$$

- N observations—**randomly sampled**
  - Why do we need random sampling?
  - How might a sampling procedure violate random sampling?
- T periods—time series data
  - If $T = 1$ (fixed point in time), the data is **cross-sectional data**
- **Pooled cross-sections** versus **panel data**:
  - Both have repeated variables across observations i and periods t
  - If observations are **the same** over time, it's a panel
  - If not, repeated/pooled cross section

Chapter 2: Simple Linear Regression

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Assumptions

- $\mathbb{E}[u] = 0$ (WLOG as long as β_0 is in the regression)
- $\mathbb{E}[u|x] = 0$ (mean independence/zero condition mean)
  - Allows us to write the *population regression function* $\mathbb{E}[y|x] = \beta_0 + \beta_1 x$
  - Implies that $\text{Cov}(x, u)$ and $\mathbb{E}[xu]$ are both 0

$$\mathbb{E}[u] = 0 \Rightarrow \mathbb{E}[y - \beta_0 - \beta_1 x] = 0 \tag{2.12}$$
$$\mathbb{E}[xu] = 0 \Rightarrow \mathbb{E}[x(y - \beta_0 - \beta_1 x)] = 0 \tag{2.13}$$

- Two equations, two unknowns (β_0, β_1)

$$\mathbb{E}[u] = 0 \Rightarrow \mathbb{E}[y - \beta_0 - \beta_1 x] = 0 \tag{2.12}$$

$$\mathbb{E}[xu] = 0 \Rightarrow \mathbb{E}[x(y - \beta_0 - \beta_1 x)] = 0 \tag{2.13}$$

- Two equations, two unknowns (β_0, β_1)
- Solve in sample (matching moments):

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{2.14}$$

$$\frac{1}{n}\sum_{i=1}^{n}x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{2.15}$$

3. Deriving OLS Estimators

$$\mathbb{E}[u] = 0 \Rightarrow \mathbb{E}[y - \beta_0 - \beta_1 x] = 0 \tag{2.12}$$

$$\mathbb{E}[xu] = 0 \Rightarrow \mathbb{E}[x(y - \beta_0 - \beta_1 x)] = 0 \tag{2.13}$$

- Two equations, two unknowns (β_0, β_1)
- Solve in sample (matching moments):

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{2.14}$$

$$\frac{1}{n}\sum_{i=1}^{n}x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{2.15}$$

- Can rewrite (2.14) to get:

$$\overline{y} - \hat{\beta}_0 - \hat{\beta}_1\overline{x} = 0 \Rightarrow \hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$$

Plugging equation for $\hat{\beta}_0$ into (2.15):

$$\sum_{i=1}^{n} x_i \left[y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i \right] = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i (y_i - \bar{y}) = \beta_1 \sum_{i=1}^{n} x_i (x_i - \bar{x})$$

Hence, we can write $\hat{\beta}_1$ as:

$$\hat{\beta}_1 = \frac{\sum_i x_i (y_i - \bar{y})}{\sum_i x_i (x_i - \bar{x})} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i (x_i - \bar{x}) y_i}{\sum_i (x_i - \bar{x})^2}$$

4. Properties of OLS Estimators

- Regression line *must* go through $(\overline{x}, \overline{y})$
- $\sum_i \hat{u}_i = \sum_i x_i \hat{u}_i = 0$
- SST = SSE $+$ SSR, where
  - SST = $\sum_i (y_i - \overline{y})^2$
  - SSE = $\sum_i (\hat{y}_i - \overline{y})^2$ (sometimes called regression/model sum of squares)
  - SSR = $\sum_i \hat{u}_i^2$ (sometimes called error sum of squares)
- OLS is responsive to changes in units of measurement, but in sensible ways. How does it respond?

Log-level models:

$$\log(y) = \beta_0 + \beta_1 x$$

Coefficients multiplied by 100 are *approximately* **percentage changes** in y

Log-level models:

$$\log(y) = \beta_0 + \beta_1 x$$

Coefficients multiplied by 100 are *approximately* **percentage changes** in y

Log-log models:

$$\log(y) = \beta_0 + \beta_1 \log(x)$$

Coefficients are *exactly* **elasticities** of y with respect to x

Can you prove these?

Log-level models:

$$\log(y) = \beta_0 + \beta_1 x$$

Coefficients multiplied by 100 are *approximately* **percentage changes** in y

Log-log models:

$$\log(y) = \beta_0 + \beta_1 \log(x)$$

Coefficients are *exactly* **elasticities** of y with respect to x

Can you prove these?

- **Linear estimators:** linear function of data
- **Linear regression:** linear function of parameters β

7. & 8. Proofs of Unbiasedness

$$\hat{\beta}_1 = \frac{\sum_i (x_i - \overline{x}) y_i}{\sum_i (x_i - \overline{x})^2} = \frac{\sum_i (x_i - \overline{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_i (x_i - \overline{x})^2},$$

Hence, $\mathbb{E}[\hat{\beta}_1] = \beta_1$ (can show β_0 is unbiased easily from this).

$$\begin{aligned}\hat{\beta}_1 = \frac{\sum_i (x_i - \overline{x}) y_i}{\sum_i (x_i - \overline{x})^2} &= \frac{\sum_i (x_i - \overline{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_i (x_i - \overline{x})^2}, \\ &= \frac{\beta_0 \sum_i (x_i - \overline{x}) + \beta_1 \sum_i x_i (x_i - \overline{x}) + \sum_i u_i (x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2},\end{aligned}$$

Hence, $\mathbb{E}[\hat{\beta}_1] = \beta_1$ (can show β_0 is unbiased easily from this).

$$\begin{aligned}
\hat{\beta}_1 = \frac{\sum_i (x_i - \overline{x}) y_i}{\sum_i (x_i - \overline{x})^2} &= \frac{\sum_i (x_i - \overline{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_i (x_i - \overline{x})^2}, \\
&= \frac{\beta_0 \sum_i (x_i - \overline{x}) + \beta_1 \sum_i x_i (x_i - \overline{x}) + \sum_i u_i (x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}, \\
&= \frac{0 + \beta_1 \sum_i (x_i - \overline{x})^2 + \sum_i u_i (x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2},
\end{aligned}$$

Hence, $\mathbb{E}[\hat{\beta}_1] = \beta_1$ (can show β_0 is unbiased easily from this).

$$\begin{aligned}
\hat{\beta}_1 = \frac{\sum_i (x_i - \overline{x}) y_i}{\sum_i (x_i - \overline{x})^2} &= \frac{\sum_i (x_i - \overline{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_i (x_i - \overline{x})^2}, \\
&= \frac{\beta_0 \sum_i (x_i - \overline{x}) + \beta_1 \sum_i x_i (x_i - \overline{x}) + \sum_i u_i (x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}, \\
&= \frac{0 + \beta_1 \sum_i (x_i - \overline{x})^2 + \sum_i u_i (x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}, \\
&= \beta_1 + \frac{\sum_i u_i (x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}
\end{aligned}$$

Hence, $\mathbb{E}[\hat{\beta}_1] = \beta_1$ (can show β_0 is unbiased easily from this).

8. General proofs of (un)biasedness

1. Choose a good formula for regression coefficient (partialled out, etc.)
2. Plug in true linear model
3. Distribute summation
4. Simplify using properties of sums/residuals
5. Take expectation

7. & 9. Deriving Variance of β_1

Homoskedasticity Assumption: $\text{Var}(u|x) = \sigma^2$
Using above algebra,

$$\mathbb{V}[\hat{\beta}_1] = \mathbb{V}\left[\beta_1 + \frac{\sum_i u_i(x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}\right],$$

7. & 9. Deriving Variance of β_1

Homoskedasticity Assumption: $\text{Var}(u|x) = \sigma^2$
Using above algebra,

$$\begin{aligned}
\mathbb{V}[\hat{\beta}_1] &= \mathbb{V}\left[\beta_1 + \frac{\sum_i u_i(x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}\right], \\
&= 0 + \mathbb{V}\left[\frac{\sum_i u_i(x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}\right],
\end{aligned}$$

Homoskedasticity Assumption: $\text{Var}(u|x) = \sigma^2$
Using above algebra,

$$\begin{aligned}
\mathbb{V}[\hat{\beta}_1] &= \mathbb{V}\left[\beta_1 + \frac{\sum_i u_i(x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}\right], \\
&= 0 + \mathbb{V}\left[\frac{\sum_i u_i(x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}\right], \\
&= \frac{\sum_i \mathbb{V}\left[u_i\right](x_i - \overline{x})^2}{\left(\sum_i (x_i - \overline{x})^2\right)^2},
\end{aligned}$$

7. & 9. Deriving Variance of β_1

Homoskedasticity Assumption: $\text{Var}(u|x) = \sigma^2$
Using above algebra,

$$\begin{aligned}
\mathbb{V}[\hat{\beta}_1] &= \mathbb{V}\left[\beta_1 + \frac{\sum_i u_i(x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}\right], \\
&= 0 + \mathbb{V}\left[\frac{\sum_i u_i(x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2}\right], \\
&= \frac{\sum_i \mathbb{V}\left[u_i\right](x_i - \overline{x})^2}{\left(\sum_i (x_i - \overline{x})^2\right)^2}, \\
&= \frac{\sigma^2 \sum_i (x_i - \overline{x})^2}{\left(\sum_i (x_i - \overline{x})^2\right)^2} = \frac{\sigma^2}{\sum_i (x_i - \overline{x})^2}
\end{aligned}$$

10. Estimating σ^2

An unbiased estimator of $\mathbb{V}[u]$ would be $\frac{1}{n}\sum_i u_i^2$, but these are unobserved.

Instead:

1. Use residuals instead of sample errors: $\hat{\sigma}^2 = \frac{1}{n}\sum_i \hat{u}_i^2$ (this is biased)
2. Correct with a degree of freedom adjustment: $s^2 = \frac{1}{n-2}\sum_i \hat{u}_i^2$

Standard Errors:

- Standard error of the regression: $s = \sqrt{s^2}$ (RMSE)
  - Estimates standard deviation of unobservables affecting y or $\text{sd}(y|x)$
- Standard error of coefficients: $\text{se}(\hat{\beta}_1) = \frac{s}{\sqrt{SST_x}}$ (comes from square root of previous slide)
  - What do we learn from standard errors?

What if we ignore the constant term?

$$\tilde{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

- Biased if $\beta_0 \neq 0$
- May reduce variance of $\tilde{\beta}_1$
- R^2 may be negative here—What would this mean?

Chapter 3: Multiple Regression Analysis

$$y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_k x_{ki} + u_i$$

- How do you interpret β_i?
- Estimators derived in the same way (easier with matrix algebra)
- Key assumption: $\mathbb{E}[u|x_1, ..., x_k] = 0$
  - Other factors affecting y are not related (on average) to x's

$$y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_k x_{ki} + u_i$$

- How do you interpret β_i?
- Estimators derived in the same way (easier with matrix algebra)
- Key assumption: $\mathbb{E}[u|x_1, ..., x_k] = 0$
  - Other factors affecting y are not related (on average) to x's

Partialled out formula for β_1:

$$\hat{\beta}_1 = \frac{\sum_i \hat{r}_{i1} y_i}{\sum_i \hat{r}_{i1}^2}, \qquad (3.22)$$

where $\hat{r}_{i1}$ are the residuals from regressing x_1 on **all** other x's

Can define **SST**, **SSE**, and **SSR** in the same way. Then

$$R^2 = \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\text{SSR}}{\text{SST}}$$

What happens to R^2 when you add variables to a regression?

Can define **SST**, **SSE**, and **SSR** in the same way. Then

$$R^2 = \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\text{SSR}}{\text{SST}}$$

What happens to R^2 when you add variables to a regression?

- How does this affect the interpretation of R^2 in multiple regression?
- Should you judge a regression based on its R^2?

15. Unbiasedness in MLR

$$\hat{\beta}_1 = \frac{\sum_i \hat{r}_{i1}(\beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + u_i)}{\sum_i \hat{r}_{i1}^2},$$

15. Unbiasedness in MLR

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_i \hat{r}_{i1}(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i)}{\sum_i \hat{r}_{i1}^2}, \\
&= \frac{\beta_0 \sum_i \hat{r}_{i1} + \beta_1 \sum_i \hat{r}_{i1} x_{i1} + \ldots + \beta_k \sum_i \hat{r}_{i1} x_{ik} + \sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2},
\end{aligned}
$$

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_i \hat{r}_{i1}(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i)}{\sum_i \hat{r}_{i1}^2}, \\
&= \frac{\beta_0 \sum_i \hat{r}_{i1} + \beta_1 \sum_i \hat{r}_{i1} x_{i1} + \ldots + \beta_k \sum_i \hat{r}_{i1} x_{ik} + \sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2}, \\
&= \frac{0 + \beta_1 \sum_i \hat{r}_{i1}{}^2 + 0 + \sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2},
\end{aligned}
$$

15. Unbiasedness in MLR

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_i \hat{r}_{i1}(\beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + u_i)}{\sum_i \hat{r}_{i1}^2}, \\
&= \frac{\beta_0 \sum_i \hat{r}_{i1} + \beta_1 \sum_i \hat{r}_{i1} x_{i1} + ... + \beta_k \sum_i \hat{r}_{i1} x_{ik} + \sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2}, \\
&= \frac{0 + \beta_1 \sum_i \hat{r}_{i1}{}^2 + 0 + \sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2}, \\
&= \beta_1 + \frac{\sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2},
\end{aligned}
$$

15. Unbiasedness in MLR

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_i \hat{r}_{i1}(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i)}{\sum_i \hat{r}_{i1}^2}, \\
&= \frac{\beta_0 \sum_i \hat{r}_{i1} + \beta_1 \sum_i \hat{r}_{i1} x_{i1} + \ldots + \beta_k \sum_i \hat{r}_{i1} x_{ik} + \sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2}, \\
&= \frac{0 + \beta_1 \sum_i \hat{r}_{i1}{}^2 + 0 + \sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2}, \\
&= \beta_1 + \frac{\sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2}, \\
\Rightarrow & \mathbb{E}[\hat{\beta}_1] = \beta_1
\end{aligned}
$$

Gauss-Markov Assumptions:

- Conditional Mean Assumption: $\mathbb{E}[u|x_1, ..., x_k] = 0$
- Linear in Parameters: The true model is $y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u$
- Random Sampling: Need a representative sample
- No Linear Relationships in Data: (no perfect multicolinearity)
- Homoskedasticity: $\mathbb{V}[u] = \sigma^2$

If a variable x_j is **omitted** from the regression:

$$\hat{\beta}_1 = \frac{\sum_i \hat{r}_{i1}(\beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + u_i)}{\sum_i \hat{r}_{i1}^2},$$

In general, $\tilde{\beta}_i = \hat{\beta}_i + \hat{\beta}_2 \tilde{\delta}$

If a variable x_j is **omitted** from the regression:

$$\hat{\beta}_1 = \frac{\sum_i \hat{r}_{i1}(\beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + u_i)}{\sum_i \hat{r}_{i1}^2},$$

$$= \frac{0 + \beta_1 \sum_i \hat{r}_{i1}{}^2 + \beta_j \sum_i \hat{r}_{i1} x_{ij} + \sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2},$$

In general, $\tilde{\beta}_i = \hat{\beta}_i + \hat{\beta}_2 \tilde{\delta}$

If a variable x_j is **omitted** from the regression:

$$\hat{\beta}_1 = \frac{\sum_i \hat{r}_{i1}(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i)}{\sum_i \hat{r}_{i1}^2},$$

$$= \frac{0 + \beta_1 \sum_i \hat{r}_{i1}{}^2 + \beta_j \sum_i \hat{r}_{i1} x_{ij} + \sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2},$$

$$= \beta_1 + \frac{\beta_j \sum_i \hat{r}_{i1} x_{ij} + \sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2},$$

In general, $\tilde{\beta}_i = \hat{\beta}_i + \hat{\beta}_2 \tilde{\delta}$

If a variable x_j is **omitted** from the regression:

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_i \hat{r}_{i1}(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i)}{\sum_i \hat{r}_{i1}^2}, \\
&= \frac{0 + \beta_1 \sum_i \hat{r}_{i1}{}^2 + \beta_j \sum_i \hat{r}_{i1} x_{ij} + \sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2}, \\
&= \beta_1 + \frac{\beta_j \sum_i \hat{r}_{i1} x_{ij} + \sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2}, \\
\Rightarrow \mathbb{E}[\hat{\beta}_1] &= \beta_1 + \beta_j \frac{\sum_i \hat{r}_{i1} x_{ij}}{\sum_i \hat{r}_{i1}^2}
\end{aligned}
$$

In general, $\tilde{\beta}_i = \hat{\beta}_i + \hat{\beta}_2 \tilde{\delta}$

18. Variance of β_i in MLR

$$\mathbb{V}[\hat{\beta}_j] = \frac{\sigma^2}{SST_j(1 - R_j^2)},$$

where $SST_j = \sum_i (x_{ij} - \overline{x}_j)^2$ and R_j^2 is from the regression of x_j on all other x's

If you add more regressors, how does this variance change?

18. Variance of β_i in MLR

$$\mathbb{V}[\hat{\beta}_j] = \frac{\sigma^2}{SST_j(1 - R_j^2)},$$

where $SST_j = \sum_i (x_{ij} - \overline{x}_j)^2$ and R_j^2 is from the regression of x_j on all other x's

If you add more regressors, how does this variance change?

- Variance of the regression, σ^2 should go down
- SST_j won't change (sidenote: what do we want this to look like?)
- R_j^2 will weakly increase
- Hence both numerator and denominator decrease, so variance may change **in either direction**

Formally, $R_j^2 = 1$ violates assumptions. Informally, $R_j^2 \to 1$ causes what bad things?.

- Perfect multicolinearity only exists with algebraic relationships
- High colinearity can always be solved by obtaining more data ("micronumerosity")
- Some stats exist to "diagnose", but are **almost always misused**
- Conclusion: don't use these in your own analysis, only for the test

Variance Inflation Factor:

$$\text{VIF}_j = \frac{1}{1 - R_j^2} = \mathbb{V}[\hat{\beta}_1]\frac{SST_j}{\sigma^2}$$

When does the VIF matter?

1. If we need x_j to infer causality of x_i on y, have to deal with colinearity
2. If parameter of interest is β_i, then the VIF's of β_j's don't matter
3. Arbitrary VIF cutoffs don't help because they are highly dependent on sample size and sample variation

Conclusion: VIF is rarely used in practice (for the test, $VIF > 10$ is good rule)

21. Estimating σ^2 in MLR

An unbiased estimator of $\mathbb{V}[u]$ would be $\frac{1}{n}\sum_i u_i^2$, but these are unobserved.

Instead:

1. Use residuals instead of sample errors: $\hat{\sigma}^2 = \frac{1}{n}\sum_i \hat{u}_i^2$ (this is biased)
2. Correct with a degree of freedom adjustment: $s^2 = \frac{1}{n-k-1}\sum_i \hat{u}_i^2$

21. Estimating σ^2 in MLR

An unbiased estimator of $\mathbb{V}[u]$ would be $\frac{1}{n}\sum_i u_i^2$, but these are unobserved.

Instead:

1. Use residuals instead of sample errors: $\hat{\sigma}^2 = \frac{1}{n}\sum_i \hat{u}_i^2$ (this is biased)
2. Correct with a degree of freedom adjustment: $s^2 = \frac{1}{n-k-1}\sum_i \hat{u}_i^2$

Standard Errors:

- Standard error of the regression: $s = \sqrt{s^2}$ (RMSE)
  - Can **increase or decrease** when adding regressors based on how SSR changes relative to degrees of freedom
- Standard error of coefficients: $\text{se}(\hat{\beta}_1) = \frac{s}{\sqrt{SST_j(1-R_j^2)}}$
  - Standard errors are **inconsistent** under heteroskedasticity

OLS is the **best linear unbiased estimator (BLUE)**:

- What is a linear estimator?
- Unbiased: $\mathbb{E}[\tilde{\beta}_j] = \beta_j$
- Best: *having the smallest asymptotic variance*
  - Why is this desirable?

Chapter 4: Inference for Linear Regression

23. Sampling distributions of OLS

Normality assumption: $u \sim \mathcal{N}(0, \sigma^2)$

- Simplifying assumption—not needed in practice with large N
- Why make it?

Normality assumption: $u \sim \mathcal{N}(0, \sigma^2)$

- Simplifying assumption—not needed in practice with large N
- Why make it?

Under the normality assumption:

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \mathbb{V}[\beta_j])$$
$$\Rightarrow \left(\frac{\hat{\beta}_j - \beta_j}{\sqrt{\mathbb{V}[\beta_j]}} \right) \sim \mathcal{N}(0, 1)$$

- Any linear combination of β's is normally distributed
- Any subset of β's is jointly normally distributed

24. Testing a single coefficient in MLR

If we use estimate of σ^2 (swapping sd for standard error):

$$\left(\frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)}\right) \sim t_{n-k-1}$$

Hence we can test hypotheses of the form $\mathcal{H}_0 : \beta_j = c$ against $\mathcal{H}_1 : \beta_j \neq c$ using the following:

1. Specify null and alternative hypotheses (1- or 2-sided?) and α
2. Calculate test statistic above
3. Find p-value from computer or comparable t-stat from table
4. Interpret result (do we reject? What can we conclude about β_j?)

24. Testing a single coefficient continued

TABLE G.2 Critical Values of the *t* Distribution

	Significance Level				
1-Tailed:	**.10**	**.05**	**.025**	**.01**	**.005**
2-Tailed:	**.20**	**.10**	**.05**	**.02**	**.01**
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355

25. Confidence intervals for β_i

Dual version of hypothesis testing:

$$\mathrm{CI}(\hat{\beta}_j) = \hat{\beta}_j \pm c_\alpha * \mathrm{se}(\hat{\beta}_j)$$

- c_α comes from table in previous slide
- If confidence interval overlaps with c, cannot reject $\mathcal{H}_0$ at specified α
- How do we interpret a confidence interval?

26. Testing linear combinations

Can we expand tests to $\mathcal{H}_0 : \theta = 0$, where θ is a **linear combination of β's**?

1. Specify $\mathcal{H}_0$, $\mathcal{H}_1$, and α (WLOG can set $\theta = 0$)
2. Estimate se(θ) with transformed regression
   1. Rewrite $\mathcal{H}_0$ in terms of one β
   2. Plug in to original regression equation and simplify
   3. Estimate transformed model to get correct se(θ)
3. Evaluate test stat: $\hat{\theta}/\text{se}(\hat{\theta})$
4. Find corresponding *p*-value (or *t*-stat from table)
5. Interpret result (do we reject? How do we unpack θ?)

27. & 28. Testing multiple linear restrictions

Now, we want to test **multiple** parameters at once (not just linear combo)

e.x.: $\mathcal{H}_0 : \beta_3 = \beta_4 = \beta_5 = 0$

1. Specify $\mathcal{H}_0$, α, and $\mathcal{H}_1$ ($\beta_i \neq 0$ for at least one i)
2. Estimate restricted and unrestricted models
3. Compute F-statistic:

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} = \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n-k-1)},$$

   where q is the number of restrictions

   - Note: $F \geq 0$ always—something's off if you get a negative stat
4. Look up F stat using table; df: $(q, n-k-1)$
5. Interpret accordingly (careful of what rejection implies)

Why are we testing different things?

- Square of t stat is the F stat—hence, the two approaches are **equivalent** in single tests
- t-stat allows for one-sided tests directly (more flexible)
- Plus they're less work to calculate
- One caveat: it is possible to have $\beta_1 \neq 0$ and $\beta_1 = ... = \beta_k = 0$ if you throw in enough garbage
- Testing should be **well-motivated**

30. F-stats for overall regression

If the model includes k regressors, we are testing

$$\mathcal{H}_0 : \beta_1 = ... = \beta_k = 0$$
$$\mathcal{H}_1 : \beta_i \neq 0 \text{ for at least one } i$$

1. Specify α
2. Estimate unrestricted model
   - Restricted model is just $\overline{y}$, which has $R^2 = 0$
3. Calculate F test stat:

$$F = \frac{R^2/k}{(1 - R^2)/(n - k - 1)}$$

4. Compare with F-stat from table
5. Interpret

31. Testing general linear restrictions

- Comes down to correct estimation of restricted model
- Same procedure as before

Questions?