Motivation

- Linear regression is arguably the most popular modeling approach across every field in the social sciences.
  1. Very robust technique
  2. Linear regression also provides a basis for more advanced empirical methods.
  3. Transparent and relatively easy to understand technique
  4. Useful for both descriptive and structural analysis

- We’re going to learn linear regression inside and out from an applied perspective
  - focusing on the appropriateness of different assumptions, model building, and interpretation

- This lecture draws heavily from Wooldridge’s undergraduate and graduate texts, as well as Greene’s graduate text.
The simple linear regression model (a.k.a. bivariate linear regression model, 2-variable linear regression model)

\[ y = \alpha + \beta x + u \] (1)

- \( y \) = dependent variable, outcome variable, response variable, explained variable, predicted variable, regressand
- \( x \) = independent variable, explanatory variable, control variable, predictor variable, regressor, covariate
- \( u \) = error term, disturbance
- \( \alpha \) = intercept parameter
- \( \beta \) = slope parameter
Details

- Recall model is

\[ y = \alpha + \beta x + u \]

- \((y, x, u)\) are random variables
- \((y, x)\) are observable (we can sample observations on them)
- \(u\) is unobservable \(\Rightarrow\) no stat tests involving \(u\)
- \((\alpha, \beta)\) are unobserved but estimable under certain cond’s
- Model implies that \(u\) captures everything that determines \(y\) except for \(x\)
- In natural sciences, this often includes frictions, air resistance, etc.
- In social sciences, this often includes a lot of stuff!!!
Assumptions

1. \( E(u) = 0 \)
   - As long as we have an intercept, this assumption is innocuous
   - Imagine \( E(u) = k \neq 0 \). We can rewrite \( u = k + w \implies \)
     \[
     y_i = (\alpha + k) + \beta E(x_i) + w
     \]
   where \( E(\omega) = 0 \). Any non-zero mean is absorbed by the intercept.

2. \( E(u|x) = E(u) \)
   - Assuming \( q \perp u \) (\( \perp = \) orthogonal) is *not enough!* Correlation only measures *linear* dependence
   - **Conditional mean independence**
   - Implied by full independence \( q \parallel u \) (\( \parallel = \) independent)
   - Implies uncorrelated
   - Intuition: avg of \( u \) does *not* depend on the value of \( q \)
   - Can combine with zero mean assumption to get zero conditional mean assumption \( E(u|q) = E(u) = 0 \)
This is the key assumption in most applications

Can we test it?

- Run regression.
- Take residuals $\hat{u} = y - \hat{y}$ & see if avg $\hat{u}$ at each value of $x = 0$?
- Or, see if residuals are uncorrelated with $x$
- Does these exercise make sense?

Can we think about it?

- The assumption says that no matter whether $x$ is low, medium, or high, the unexplained portion of $y$ is, on average, the same (0).
- But, what if agents (firms, etc.) with different values of $x$ are different along other dimensions that matter for $y$?
CMI Example 1: Capital Structure

Consider the regression

\[ \text{Leverage}_i = \alpha + \beta \text{Profitability}_i + u_i \]

CMI \( \implies \) that average \( u \) for each level of \text{Profitability} is the same

But, unprofitable firms tend to have higher bankruptcy risk and should have lower leverage than more profitable firms according to tradeoff theory

Or, unprofitable firms have accumulated fewer profits and may be forced to debt financing, implying higher leverage according to the pecking order

These e.g.’s show that the average \( u \) is likely to vary with the level of profitability

- 1st e.g., low profitable firms will be less levered implies lower avg \( u \) for less profitable firms
- 2nd e.g., low profitable firms will be more levered implies higher avg \( u \) for less profitable firms
Consider the regression

\[ \text{Investment}_{i} = \alpha + \beta q_{i} + u_{i} \]

CMI \implies that average \( u \) for each level of \( q \) is the same

But, firms with low \( q \) may be in distress and invest less

Or, firms with high \( q \) may have difficulty raising sufficient capital to finance their investment

These e.g.’s show that the average \( u \) is likely to vary with the level of \( q \)

1st e.g., low \( q \) firms will invest less implies higher avg \( u \) for low \( q \) firms

2nd e.g., high \( q \) firms will invest less implies higher avg \( u \) for low \( q \) firms
Population Regression Function (PRF)

- PRF is $E(y|x)$. It is fixed but unknown. For simple linear regression:

$$PRF = E(y|x) = \alpha + \beta x$$  \hspace{1cm} (2)

- Intuition: for any value of $x$, distribution of $y$ is centered about $E(y|x)$

![Graph showing PRF]

$E(y|x) = \alpha + \beta x$
OLS Regression Line

- We don’t observe PRF, but we can estimate via OLS
  \[ y_i = \alpha + \beta x_i + u_i \]  
  for each sample point \( i \)
- What is \( u_i \)? It contains all of the factors affecting \( y_i \) other than \( x_i \).
  \[ u_i \text{ contains a lot of stuff!} \]
  - Consider complexity of:
    - \( y \) is individual food expenditures
    - \( y \) is corporate leverage ratios
    - \( y \) is interest rate spread on a bond
- Estimated Regression Line (a.k.a. Sample Regression Function (SRF))
  \[ \hat{y} = \hat{\alpha} + \hat{\beta} x \]  
  Plug in an \( x \) and out comes an estimate of \( y \), \( \hat{y} \)
- Note: Different sample \( \implies \) different \( (\hat{\alpha}, \hat{\beta}) \)
OLS Estimates

- Estimators:

  \[
  \text{Slope} = \hat{\beta} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}
  \]

  \[
  \text{Intercept} = \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}
  \]

- Population analogues

  \[
  \text{Slope} = \frac{\text{Cov}(x, y)}{\text{Var}(x)} = \text{Corr}(x, y) \frac{SD(y)}{SD(x)}
  \]

  \[
  \text{Intercept} = E(y) - \hat{\beta}E(x)
  \]
Linear Regression
Example: CEO Compensation

- Model

\[ salary = \alpha + \beta \text{ROE} + y \]

- Sample 209 CEOs in 1990. Salaries in $000s and ROE in % points.
- SRF

\[ salary = 963.191 + 18.501\text{ROE} \]

- What do the coefficients tell us?
- Is the key CMI assumption likely to be satisfied?
  - Is ROE the only thing that determines salary?
  - Is the relationship linear? \( \Rightarrow \) estimated change is constant across salary and ROE

\[ dy/dx = \beta \text{ indep of salary & ROE} \]

- Is the relationship constant across CEOs?
**Univariate Regression**

**Multivariate Regression**

**Specification Issues**

**Inference**

**Basics**

**Ordinary Least Squares (OLS) Estimates**

**Units of Measurement and Functional Form**

**OLS Estimator Properties**

---

**PRF vs. SRF**

\[ \text{salary} = 963.191 + 18.501 \text{ROE} \]

\[ E(\text{salary} | \text{ROE}) = \alpha + \beta \text{ROE} \]
Goodness-of-Fit ($R^2$)

- R-squared defined as

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

where

- $SSE = \text{Sum of Squares Explained} = \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2$
- $SST = \text{Sum of Squares Total} = \sum_{i=1}^{N} (y_i - \bar{y})^2$
- $SSR = \text{Sum of Squares Residual} = \sum_{i=1}^{N} (\hat{u}_i - \bar{u})^2 = \sum_{i=1}^{N} \hat{u}_i^2$
- $R^2 = [\text{Corr}(y, \hat{y})]^2$
Example: CEO Compensation

- Model

\[ \text{salary} = \alpha + \beta \text{ROE} + y \]

- \( R^2 = 0.0132 \)

- What does this mean?
Scaling the Dependent Variable

- Consider CEO SRF

  \[ salary = 963.191 + 18.501 \times ROE \]

- Change measurement of salary from $000s to $s. What happens?

  \[ salary = 963,191 + 18,501 \times ROE \]

- More generally, multiplying dependent variable by constant \( c \) \( \Rightarrow \) OLS intercept and slope are also multiplied by \( c \)

  \[ y = \alpha + \beta x + u \]

  \[ \iff \quad cy = (c\alpha) + (c\beta)x + cu \]

  (Note: variance of error affected as well.)

- Scaling \( \Rightarrow \) multiplying every observation by same \# 

- No effect on \( R^2 \) - invariant to changes in units
Scaling the Independent Variable

- Consider CEO SRF

\[ salary = 963.191 + 18.501 \times ROE \]

- Change measurement of ROE from percentage to decimal (i.e., multiply every observation’s ROE by $\frac{1}{100}$)

\[ salary = 963.191 + 1,850.1 \times ROE \]

- More generally, multiplying independent variable by constant $c \implies$ OLS intercept is unchanged but slope is divided by $c$

\[ y = \alpha + \beta x + u \]

\[ \iff y = \alpha + \left( \frac{\beta}{c} \right)cx + cu \]

- Scaling $\implies$ multiplying every observation by same #
- No effect on $R^2$ - invariant to changes in units
Changing Units of Both $y$ and $x$

Model:

$$y = \alpha + \beta x + u$$

What happens to intercept and slope when we scale $y$ by $c$ and $x$ by $k$?

$$cy = c\alpha + c\beta x + cu$$

$$cy = (c\alpha) + (c\beta/k)kx + cu$$

- intercept scaled by $c$, slope scaled by $c/k$
Shifting Both $y$ and $x$

- **Model:**
  \[ y = \alpha + \beta x + u \]

- What happens to intercept and slope when we add $c$ and $k$ to $y$ and $x$?
  \[
  \begin{align*}
  c + y &= c + \alpha + \beta x + u \\
  c + y &= c + \alpha + \beta(x + k) - \beta k + u \\
  c + y &= (c + \alpha - \beta k) + \beta(x + k) + u
  \end{align*}
  \]

- Intercept shifted by $\alpha - \beta k$, slope unaffected
Incorporating Nonlinearities

- Consider a traditional wage-education regression

\[ \text{wage} = \alpha + \beta \text{education} + u \]

- This formulation assumes change in wages is constant for all educational levels
- E.g., increasing education from 5 to 6 years leads to the same $ increase in wages as increasing education from 11 to 12, or 15 to 16, etc.
- Better assumption is that each year of education leads to a constant proportionate (i.e., percentage) increase in wages
- Approximation of this intuition captured by

\[ \log(\text{wage}) = \alpha + \beta \text{education} + u \]
Log Dependent Variables

- Percentage change in wage given one unit increase in education is

\[ \% \Delta \text{wage} \approx (100\beta) \Delta \text{educ} \]

- Percent change in wage is constant for each additional year of education

\[ \Rightarrow \text{Change in wage for an extra year of education increases as education increases.} \]

- I.e., increasing return to education (assuming \( \beta > 0 \))

- Log wage is linear in education. Wage is nonlinear

\[ \log(\text{wage}) = \alpha + \beta \text{education} + u \]

\[ \Rightarrow \text{wage} = \exp(\alpha + \beta \text{education} + u) \]
Log Wage Example

- Sample of 526 individuals in 1976. Wages measured in $/hour. Education = years of education.
- SRF:
  
  \[
  \log(wage) = 0.584 + 0.083 \text{education}, \quad R^2 = 0.186 
  \]

  Interpretation:
  - Each additional year of education leads to an 8.3% increase in wages (NOT log(wages)!!!).
  - For someone with no education, their wage is exp(0.584)...this is meaningless because no one in sample has education=0.

- Ignores other nonlinearities. E.g., diploma effects at 12 and 16.
Alter CEO salary model

\[ \log(salary) = \alpha + \beta \log(sales) + u \]

- \( \beta \) is the **elasticity** of salary w.r.t. sales
- SRF

\[ \log(salary) = 4.822 + 0.257 \log(sales), \quad R^2 0.211 \]

- Interpretation: For each 1% increase in sales, salary increase by 0.257%
- Intercept meaningless...no firm has 0 sales.
What happens to intercept and slope if we $\Delta$ units of dependent variable when it’s in log form?

\[
\log(y) = \alpha + \beta x + u
\]

\[\iff \quad \log(c) + \log(y) = \log(c) + \alpha + \beta x + u
\]

\[\iff \quad \log(cy) = (\log(c) + \alpha) + \beta x + u
\]

- Intercept shifted by $\log(c)$, slope unaffected because slope measures proportionate change in log-log model.
Changing Units in Level-Log Model

- What happens to intercept and slope if we $\Delta$ units of independent variable when it’s in log form?

\[
y = \alpha + \beta \log(x) + u
\]
\[
\iff \beta \log(c) + y = \alpha + \beta \log(x) + \beta \log(c) + u
\]
\[
\iff y = (\alpha - \beta \log(c)) + \beta \log(cx) + u
\]

- Slope measures *proportionate* change
Changing Units in Log-Log Model

What happens to intercept and slope if we \( \Delta \) units of dependent variable?

\[
\log(y) = \alpha + \beta \log(x) + u
\]

\[\iff\]

\[
\log(c) + \log(y) = \log(c) + \alpha + \beta \log(x) + u
\]

\[\iff\]

\[
\log(cy) = (\alpha + \log(c)) + \beta \log(x) + u
\]

What happens to intercept and slope if we \( \Delta \) units of independent variable?

\[
\log(y) = \alpha + \beta \log(x) + u
\]

\[\iff\]

\[
\beta \log(c) + \log(y) = \alpha + \beta \log(x) + \beta \log(c) + u
\]

\[\iff\]

\[
\log(y) = (\alpha - \beta \log(c)) + \beta \log(cx) + u
\]
### Log Functional Forms

<table>
<thead>
<tr>
<th>Model</th>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Interpretation of $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level-level</td>
<td>$y$</td>
<td>$x$</td>
<td>$dy = \beta dx$</td>
</tr>
<tr>
<td>Level-log</td>
<td>$y$</td>
<td>log($x$)</td>
<td>$dy = (\beta/100)% dx$</td>
</tr>
<tr>
<td>Log-level</td>
<td>log($y$)</td>
<td>$x$</td>
<td>$%dy = (100\beta)dx$</td>
</tr>
<tr>
<td>Log-log</td>
<td>log($y$)</td>
<td>log($x$)</td>
<td>$%dy = \beta % dx$</td>
</tr>
</tbody>
</table>

- E.g., In Log-level model, $100 \times \beta = \%$ change in $y$ for a 1 unit increase in $x$ ($100\beta =$ **semi-elasticity**)
- E.g., In Log-log model, $\beta = \%$ change in $y$ for a 1% change in $x$ ($\beta =$ **elasticity**)
When is OLS unbiased (i.e., $E(\hat{\beta}) = \beta$)?

1. Model is linear in parameters
2. We have a random sample (e.g., self selection)
3. Sample outcomes on $x$ vary (i.e., no collinearity with intercept)
4. Zero conditional mean of errors (i.e., $E(u|x) = 0$)

Unbiasedness is a feature of sampling distributions of $\hat{\alpha}$ and $\hat{\beta}$.

For a given sample, we hope $\hat{\alpha}$ and $\hat{\beta}$ are close to true values.
Variance of OLS Estimators

- **Homoskedasticity** \(\Rightarrow \ Var(u|x) = \sigma^2\)
- **Heterokedasticity** \(\Rightarrow \ Var(u|x) = f(x) \in \mathbb{R}^+\)
Remember, larger error variance $\implies$ larger $\text{Var}(\beta)$ $\implies$ bigger SEs

Intuition: More variation in unobservables affecting $y$ makes it hard to precisely estimate $\beta$

Relatively more variation in $x$ is our friend!!!

More variation in $x$ means lower SEs for $\beta$

Likewise, larger samples tend to increase variation in $x$ which also means lower SEs for $\beta$

I.e., we like big samples for identifying $\beta$!
• **Multiple Linear Regression Model**

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u \]

• Same notation and terminology as before.

• Similar key identifying assumptions

  1. No perfect collinearity among covariates
  2. \( E(u|x_1, \ldots, x_k) = 0 \implies \) at a minimum no correlation and we have correctly accounted for the functional relationships between \( y \) and \((x_1, \ldots, x_k)\)

• SRF

\[ y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k \]
Interpretation

- Estimated intercept $\hat{\beta}_0$ is predicted value of $y$ when all $x = 0$. Sometimes this makes sense, sometimes it doesn’t.

- Estimated slopes ($\hat{\beta}_1, \ldots, \hat{\beta}_k$) have partial effect interpretations

$$
\Delta \hat{y} = \hat{\beta}_1 \Delta x_1 + \ldots + \hat{\beta}_k \Delta x_k
$$

I.e., given changes in $x_1$ through $x_k$, $(\Delta x_1, \ldots, \Delta x_k)$, we obtain the *predicted* change in $y$.

- When all but one covariate, e.g., $x_1$, is held fixed so $(\Delta x_2, \ldots, \Delta x_k) = (0, \ldots, 0)$ then

$$
\Delta \hat{y} = \hat{\beta}_1 \Delta x_1
$$

I.e., $\hat{\beta}_1$ is the coefficient holding *all else fixed* (ceteris paribus)
Example: College GPA

- SRF of college GPA and high school GPA (4-point scales) and ACT score for $N = 141$ university students

$$\hat{colGPA} = 1.29 + 0.453hsGPA + 0.0094ACT$$

- What do intercept and slopes tell us?
  - Consider two students, Fred and Bob, with identical ACT score but $hsGPA$ of Fred is 1 point higher than that of Bob. Best prediction of Fred’s $colGPA$ is 0.453 points higher than that of Bob.

- SRF without $hsGPA$

$$\hat{colGPA} = 1.29 + 0.0271ACT$$

- What’s different and why? Can we use it to compare 2 people with same $hsGPA$?
Consider prev example. Holding ACT fixed, another point on high school GPA is predicted to inc college GPA by 0.452 points.

If we could collect a sample of individuals with the same high school ACT, we could run a simple regression of college GPA on high school GPA. This holds all else, ACT, fixed.

Multiple regression mimics this scenario without restricting the values of any independent variables.
Each $\beta$ corresponds to the partial effect of its covariate.

What if we want to change more than one variable at the same time?

E.g., What is the effect of increasing the high school GPA by 1 point and the ACT score by 1 point?

$$\Delta\hat{colGPA} = 0.453\Delta hsGPA + 0.0094\Delta ACT = 0.4624$$

E.g., What is the effect of increasing the high school GPA by 2 points and the ACT score by 10 points?

$$\Delta\hat{colGPA} = 0.453\Delta hsGPA + 0.0094\Delta ACT$$

$$= 0.453 \times 2 + 0.0094 \times 10 = 1$$
Fitted Values and Residuals

- Residual = \( \hat{u}_i = y_i - \hat{y}_i \)
- Properties of residuals and fitted values:
  1. Sample avg of residuals = 0 \(\Rightarrow\) \(\hat{y} = \bar{y}\)
  2. Sample cov between each indep variable and residuals = 0
  3. Point of means \((\bar{y}, \bar{x}_1, ..., \bar{x}_k)\) lies on regression line.
Partial Regression

- Consider 2 independent variable model
  \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \]

- What’s the formula for just \( \hat{\beta}_1 \)?
  \[ \hat{\beta}_1 = (\hat{r}_1'\hat{r}_1)^{-1}\hat{r}_1'y \]
  where \( \hat{r}_1 \) are the residuals from a regression of \( x_1 \) on \( x_2 \).

- In other words,
  1. regress \( x_1 \) on \( x_2 \) and save residuals
  2. regress \( y \) on residuals
  3. coefficient on residuals will be identical to \( \hat{\beta}_1 \) in multivariate regression
Frisch-Waugh-Lovell I

- More generally, consider general linear setup
  
  \[ y = XB + u = X_1B_1 + X_2B_2 + u \]

- One can show that
  
  \[ \hat{B}_2 = (X_2' M_1 X_2)^{-1} (X_2' M_1 y) \]  \hspace{1cm} (5)

where

\[ M_1 = (I - P_1) = I - X_1(X_1'X_1)^{-1}X_1' \]

- \( P_1 \) is the projection matrix that takes a vector \( (y) \) and projects it onto the space spanned by columns of \( X_1 \)

- \( M_1 \) is the orthogonal compliment, projecting a vector onto the space orthogonal to that spanned by \( X_1 \)
What does equation (5) mean?

Since $M_1$ is idempotent

$$\hat{B}_2 = (X'_2 M_1 M_1 X_2)^{-1} (X'_2 M_1 M_1 y)$$

$$= (\tilde{X}'_2 \tilde{X}_2)^{-1} (\tilde{X}'_2 \tilde{y})$$

So $\hat{B}_2$ can be obtained by a simple multivariate regression of $\tilde{y}$ on $\tilde{X}_2$

But $\tilde{y}$ and $\tilde{X}_2$ are just the residuals obtained from regressing $y$ and each component of $X_2$ on the $X_1$ matrix
Omitted Variables Bias

- Assume correct model is:
  \[ y = XB + u = X_1B_1 + X_2B_2 + u \]

- Assume we *incorrectly* regress \( y \) on just \( X_1 \). Then
  \[
  \hat{B}_1 = (X'_1X_1)^{-1}X'_1y \\
  = (X'_1X_1)^{-1}X'_1(X_1B_1 + X_2B_2 + u) \\
  = B_1 + (X'_1X_1)^{-1}X'_1X_2B_2 + (X'_1X_1)^{-1}X'_1u
  \]

- Take expectations and we get
  \[
  \hat{B}_1 = B_1 + (X'_1X_1)^{-1}X'_1X_2B_2
  \]

Note \((X'_1X_1)^{-1}X'_1X_2\) is the column of slopes in the OLS regression of each column of \( X_2 \) on the columns of \( X_1 \)

- OLS is biased because of omitted variables and direction is unclear — depending on multiple partial effects
Bivariate Model

• With two variable setup, inference is easier

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \]

• Assume we incorrectly regress \( y \) on just \( x_1 \). Then

\[ \hat{\beta}_1 = \beta_1 + (x'_1 x_1)^{-1} x'_1 x_2 \beta_2 \]
\[ = \beta_1 + \delta \beta_2 \]

• Bias term consists of 2 terms:
  1. \( \delta = \) slope from regression of \( x_2 \) on \( x_1 \)
  2. \( \beta_2 = \) slope on \( x_2 \) from multiple regression of \( y \) on \( (x_1, x_2) \)

• Direction of bias determined by signs of \( \delta \) and \( \beta_2 \).
• Magnitude of bias determined by magnitudes of \( \delta \) and \( \beta_2 \).
Omitted Variable Bias General Thoughts

- Deriving sign of omitted variable bias with multiple regressors in estimated model is hard. Recall general formula

\[ \hat{B}_1 = B_1 + (X'_1 X_1)^{-1} X'_1 X_2 B_2 \]

\((X'_1 X_1)^{-1} X'_1 X_2\) is vector of coefficients.

- Consider a simpler model

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \]

where we omit \(x_3\)

- Note that both \(\hat{\beta}_1\) and \(\hat{\beta}_2\) will be biased because of omission unless both \(x_1\) and \(x_2\) are uncorrelated with \(x_3\).

- The omission will infect every coefficient through correlations
Example: Labor

Consider

\[ \log(wage) = \beta_0 + \beta_1 \text{education} + \beta_2 \text{ability} + u \]

If we can’t measure ability, it’s in the error term and we estimate

\[ \log(wage) = \beta_0 + \beta_1 \text{education} + w \]

What is the likely bias in \( \hat{\beta}_1 \)? recall

\[ \hat{\beta}_1 = \beta_1 + \delta \beta_2 \]

where \( \delta \) is the slope from regression of ability on education.

- Ability and education are likely positively correlated \( \Rightarrow \delta > 0 \)
- Ability and wages are likely positively correlated \( \Rightarrow \beta_2 > 0 \)
- So, bias is likely positive \( \Rightarrow \hat{\beta}_1 \) is too big!
Goodness of Fit

- $R^2$ still equal to squared correlation between $y$ and $\hat{y}$
- Low $R^2$ doesn’t mean model is wrong
- Can have a low $R^2$ yet OLS estimate may be reliable estimates of ceteris paribus effects of each independent variable
- Adjust $R^2$

\[
R_a^2 = 1 - (1 - R^2) \frac{n - 1}{n - k - 1}
\]

where $k = \#$ of regressors excluding intercept
- Adjust $R^2$ corrects for df and it can be < 0
Unbiasedness

- When is OLS unbiased (i.e., $E(\hat{\beta}) = \beta$)?
  1. Model is linear in parameters
  2. We have a random sample (e.g., self selection)
  3. No perfect collinearity
  4. Zero conditional mean of errors (i.e., $E(u|x) = 0$)

- Unbiasedness is a feature of sampling distributions of $\hat{\alpha}$ and $\hat{\beta}$.
- For a given sample, we hope $\hat{\alpha}$ and $\hat{\beta}$ are close to true values.
Irrelevant Regressors

- What happens when we include a regressor that shouldn’t be in the model? (*overspecified*)
- No affect on unbiasedness
- Can affect the variances of the OLS estimator
Variance of OLS Estimators

- Sampling variance of OLS slope

\[
\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2(1 - R_j^2)}
\]

for \( j = 1, \ldots, k \), where \( R_j^2 \) is the \( R^2 \) from regressing \( x_j \) on all other independent variables including the intercept and \( \sigma^2 \) is the variance of the regression error term.

- Note
  - Bigger error variance (\( \sigma^2 \)) \( \implies \) bigger SEs (Add more variables to model, change functional form, improve fit!)
  - More sampling variation in \( x_j \) \( \implies \) smaller SEs (Get a larger sample)
  - Higher collinearity (\( R_j^2 \)) \( \implies \) bigger SEs (Get a larger sample)
Multicollinearity

- Problem of small sample size.
- No implication for bias or consistency, but can inflate SEs
- Consider

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \]

where \( x_2 \) and \( x_3 \) are highly correlated.

- \( \text{Var}(\hat{\beta}_2) \) and \( \text{Var}(\hat{\beta}_3) \) may be large.
- But correlation between \( x_2 \) and \( x_3 \) has no direct effect on \( \text{Var}(\hat{\beta}_1) \)
- If \( x_1 \) is uncorrelated with \( x_2 \) and \( x_3 \), then \( R^2_1 = 0 \) and \( \text{Var}(\hat{\beta}_1) \) is unaffected by correlation between \( x_2 \) and \( x_3 \)
- Make sure included variables are not too highly correlated with the variable of interest

**Variance Inflation Factor (VIF)** = \( 1/(1 - R^2_j) \) above 10 is sometimes cause for concern but this is arbitrary and of limited use
Data Scaling

No one wants to see a coefficient reported as 0.000000456, or 1,234,534,903,875.

Scale the variables for cosmetic purposes:
1. Will effect coefficients & SEs
2. Won’t affect t-stats or inference

Sometimes useful to convert coefficients into comparable units, e.g., SDs.
1. Can standardize \( y \) and \( x \)'s (i.e., subtract sample avg. & divide by sample SD) before running regression.
2. Estimated coefficients \( \implies 1 \text{ SD } \Delta \text{ in } y \text{ given } 1 \text{ SD } \Delta \text{ in } x \).

Can estimate model on original data, then multiply each coef by corresponding SD. This marginal effect \( \implies \Delta \text{ in } y \text{ units for a } 1 \text{ SD } \Delta \text{ in } x \)
Log Functional Forms

Consider

\[
\log(\text{price}) = \beta_0 + \beta_1 \log(\text{pollution}) + \beta_2 \text{rooms} + u
\]

Interpretation

1. \(\beta_1\) is the elasticity of price w.r.t. pollution. I.e., a 1% change in pollution generates an \(100\beta_1\)% change in price.
2. \(\beta_2\) is the semi-elasticity of price w.r.t. rooms. I.e., a 1 unit change in rooms generates an \(100\beta_2\)% change in price.

E.g.,

\[
\log(\text{price}) = 9.23 - 0.718 \log(\text{pollution}) + 0.306 \text{rooms} + u
\]

\[\Rightarrow\] 1% inc. in pollution \[\Rightarrow\] −0.72% dec. in price

\[\Rightarrow\] 1 unit inc. in rooms \[\Rightarrow\] −30.6% inc. in price
Log Approximation

- Note: percentage change interpretation is only \textit{approximate}!
- Approximation error occurs because as $\Delta \log(y)$ becomes larger, approximation \(\%\Delta y \approx 100\Delta \log(y)\) becomes more inaccurate. E.g.,

\[
\log(y) = \hat{\beta}_0 + \hat{\beta}_1 \log(x_1) + \hat{\beta}_2 x_2
\]

- Fixing $x_1$ (i.e., $\Delta x_1 = 0$) $\implies$ $\Delta \log(y) = \Delta \hat{\beta}_2 x_2$

- \textit{Exact} \% change is

\[
\Delta \log(y) = \log(y') - \log(y) = \hat{\beta}_2 \Delta x_2 = \hat{\beta}_2 (x'_2 - x_2)
\]

\[
\log(y'/y) = \hat{\beta}_2 (x'_2 - x_2)
\]

\[
y'/y = \exp(\hat{\beta}_2 (x'_2 - x_2))
\]

\[
[(y' - y)/y] \% = 100 \cdot \left[ \exp(\hat{\beta}_2 (x'_2 - x_2)) - 1 \right]
\]
Approximate % change $y$ : $\Delta \log(y) = \hat{\beta}_2 \Delta x_2$

Exact % change $y$ : $(\Delta y / y)\% = 100 \cdot \left[ \exp(\hat{\beta}_2 \Delta x_2) \right]$
Usefulness of Logs

- Logs lead to coefficients with appealing interpretations.
- Logs allow us to be ignorant about the units of measurement of variables appearing in logs since they’re proportionate changes.
- If \( y > 0 \), log can mitigate (eliminate) skew and heteroskedasticity.
- Logs of \( y \) or \( x \) can mitigate the influence of outliers by narrowing range.
- “Rules of thumb” of when to take logs:
  - positive currency amounts,
  - variable with large integral values (e.g., population, enrollment, etc.)
and when not to take logs:
  - variables measured in years (months),
  - proportions
- If \( y \in [0, \infty) \), can take log(1 + y)
Percentage vs. Percentage Point Change

- Proportionate (or Relative) Change
  \[
  \frac{(x_1 - x_0)}{x_0} = \frac{\Delta x}{x_0}
  \]

- Percentage Change
  \[
  \%\Delta x = 100\left(\frac{\Delta x}{x_0}\right)
  \]

- Percentage Point Change is raw change in percentages.
- E.g., let \( x = \) unemployment rate in \%
- If unemployment goes from 10\% to 9\%, then
  - 1\% percentage point change,
  - \((9-10)/10 = 0.1\) proportionate change,
  - \(100(9-10)/10 = 10\%\) percentage change,
- If you use log of a \% on LHS, take care to distinguish between percentage change and percentage point change.
Models with Quadratics

- Consider

\[ y = \beta_0 + \beta_1 x + \beta_2 x^2 + u \]

- Partial effect of \( x \)

\[ \Delta y = (\beta_1 + 2\beta_2 x)\Delta x \implies dy/dx = \beta_1 + 2\beta_2 x \]

\[ \implies \text{must pick value of } x \text{ to evaluate (e.g., } \bar{x}) \]

- \( \hat{\beta}_1 > 0, \hat{\beta}_2 < 0 \implies \text{parabolic relation} \)
  - Turning point = Maximum = \( \left| \frac{\hat{\beta}_1}{2\hat{\beta}_2} \right| \)
  - \textit{Know where the turning point is!} It may lie outside the range of } x!\text{!
  - Odd values may imply misspecification or be irrelevant (above)

- Extension to higher order straightforward
Models with Interactions

- Consider
  \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u \]

- Partial effect of \( x_1 \)
  \[ \Delta y = (\beta_1 + \beta_3 x_2)\Delta x_1 \implies \frac{dy}{dx_1} = \beta_1 + \beta_3 x_2 \]

- Partial effect of \( x_1 = \beta_1 \iff x_2 = 0 \). Have to ask if this makes sense.
- If not, plug in sensible value for \( x_2 \) (e.g., \( \bar{x}_2 \))
- Or, reparameterize the model:
  \[ y = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 - \mu_1)(x_2 - \mu_2) + u \]
  where \((\mu_1, \mu_2)\) is the population mean of \((x_1, x_2)\)
  \( \delta_2(\delta_1) \) is partial effect of \( x_2(x_1) \) on \( y \) at mean value of \( x_1(x_2) \).
Models with Interactions

- Reparameterized model

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 x_2 + \mu_1 \mu_2 - x_1 \mu_2 - x_2 \mu_1) + u \]

\[ = (\beta_0 + \beta_3 \mu_1 \mu_2) + (\beta_1 + \beta_3 \mu_2) x_1 + (\beta_2 + \beta_3 \mu_1) x_2 + \beta_3 x_1 x_2 + u \]

- For estimation purposes, can use sample mean in place of unknown population mean
- Estimating reparameterized model has two benefits:
  - Provides estimates at average value \((\hat{\delta}_1, \hat{\delta}_2)\)
  - Provides corresponding standard errors
**Predicted Values and SEs I**

- Predicted value:
  \[ \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k \]

- But this is just an estimate with a standard error. I.e.,
  \[ \hat{\theta} = \hat{\beta}_0 + \hat{\beta}_1 c_1 + \ldots + \hat{\beta}_k c_k \]

  where \((c_1, \ldots, c_k)\) is a point of evaluation

- But \(\hat{\theta}\) is just a linear combination of OLS parameters

- We know how to get the SE of this. E.g., \(k = 1\)
  \[
  \text{Var}(\hat{\theta}) = \text{Var}(\hat{\beta}_0 + \hat{\beta}_1 c_1) \\
  = \text{Var}(\hat{\beta}_0) + c_1^2 \text{Var}(\hat{\beta}_1) + 2c_1 \text{Cov}(\hat{\beta}_0, \hat{\beta}_1)
  
  \]

- Take square root and voila’! (Software will do this for you)
Alternatively, reparameterize the regression. Note

$$\hat{\theta} = \hat{\beta}_0 + \hat{\beta}_1 c_1 + \ldots + \hat{\beta}_k c_k \implies \hat{\beta}_0 = \hat{\theta} - \hat{\beta}_1 c_1 - \ldots - \hat{\beta}_k c_k$$

Plug this into the regression

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

to get

$$y = \theta_0 + \beta_1(x_1 - c_1) + \ldots + \beta_k(x_k - c_k) + u$$

I.e., subtract the value $c_j$ from each observation on $x_j$ and then run regression on transformed data.

Look at SE on intercept and that’s the SE of the predicted value of $y$ at the point $(c_1, \ldots, c_k)$

You can form confidence intervals with this too.
Predicting $y$ with $\log(y)$

- **SRF:**
  \[
  \log(y) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k
  \]

- Predicted value of $y$ is not $\exp(\log(y))$

- Recall Jensen’s inequality for convex function, $g$:
  \[
  g \left( \int f \, d\mu \right) \leq \int g \circ f \, d\mu \iff g(E(f)) \leq E(g(f))
  \]

- In our setting, $f = \log(y)$, $g=\exp()$. Jensen $\implies$
  \[
  \exp\{E[\log(y)]\} \leq E[\exp\{\log(y)\}]
  \]

  We will underestimate $y$. 
Predicting $y$ with $\log(y)$ II

- How can we get a consistent (no unbiased) estimate of $y$?
- If $u \perp X$

$$E(y|X) = \alpha_0 \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)$$

where $\alpha_0 = E(\exp(u))$

- With an estimate of $\alpha$, we can predict $y$ as

$$\hat{y} = \hat{\alpha}_0 \exp(\log(y))$$

which requires exponentiating the predicted value from the log model and multiplying by $\hat{\alpha}_0$

- Can estimate $\alpha_0$ with MOM estimator (consistent but biased because of Jensen)

$$\hat{\alpha}_0 = n^{-1} \sum_{i=1}^{n} \exp(\hat{u}_i)$$
Basics

- Qualitative information. Examples,
  1. Sex of individual (Male, Female)
  2. Ownership of an item (Own, don’t own)
  3. Employment status (Employed, Unemployed)

- Code this information using **binary** or **dummy** variables. E.g.,

  
  \[
  Male_i = \begin{cases} 
  1 & \text{if person } i \text{ is Male} \\
  0 & \text{otherwise}
  \end{cases}
  \]

  
  \[
  Own_i = \begin{cases} 
  1 & \text{if person } i \text{ owns item} \\
  0 & \text{otherwise}
  \end{cases}
  \]

  
  \[
  Emp_i = \begin{cases} 
  1 & \text{if person } i \text{ is employed} \\
  0 & \text{otherwise}
  \end{cases}
  \]

- Choice of 0 or 1 is relevant only for interpretation.
Consider

\[ \text{wage} = \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + u \]

- \( \delta_0 \) measures difference in wage between male and female given same level of education (and error term \( u \))

\[
E(\text{wage}|\text{female} = 0, \text{educ}) = \beta_0 + \beta_1 \text{educ} \\
E(\text{wage}|\text{female} = 1, \text{educ}) = \beta_0 + \delta + \beta_1 \text{educ}
\]

\[ \implies \delta = E(\text{wage}|\text{female} = 1, \text{educ}) - E(\text{wage}|\text{female} = 0, \text{educ}) \]

- Intercept for males = \( \beta_0 \), females = \( \delta_0 + \beta_0 \)
Intercept Shift

- Intercept shifts, slope is same.

\[
\begin{align*}
E(wage | female = 0, educ) &= \beta_0 + \beta_1 educ \\
E(wage | female = 1, educ) &= (\beta_0 + \delta_0) + \beta_1 educ
\end{align*}
\]
Wage Example

- SRF with $n = 526, R^2 = 0.364$

\[
\hat{wage} = -1.57 - 1.81 \text{female} + 0.571 \text{educ} + 0.025 \text{exper} + 0.141 \text{tenure}
\]

- Negative intercept is intercept for men...meaningless because other variables are never all = 0
- Females earn $1.81$/hour less than men with the same education, experience, and tenure.

- All else equal is important! Consider SRF with $n = 526, R^2 = 0.116$

\[
\hat{wage} = 7.10 - 2.51 \text{female}
\]

- Female coefficient is picking up differences due to omitted variables.
Log Dependent Variables

- Nothing really new, coefficient has % interpretation.
- E.g., house price model with $N = 88$, $R^2 = 0.649$

\[
\hat{\text{price}} = -1.35 + 0.168 \log(\text{lotsize}) + 0.707 \log(\text{sqrft}) \\
+ 0.027 \text{bdrms} + 0.054 \text{colonial}
\]

- Negative intercept is intercept for non-colonial homes...meaningless because other variables are never all $= 0$
- A colonial style home costs approximately 5.4% more than “otherwise similar” homes

- Remember this is just an approximation. If the percentage change is large, may want to compare with exact formulation
Multiple Binary Independent Variables

Consider

$$\log(\text{wage}) = 0.321 + 0.213\text{marriedMale} - 0.198\text{marriedFemale}$$
$$+ -0.110\text{singleFemale} + 0.079\text{education}$$

The omitted category is single male $\implies$ intercept is intercept for base group (all other vars = 0)

Each binary coefficient represent the estimated difference in intercepts between that group and the base group

E.g., $\text{marriedMale} \implies$ that married males earn approximately 21.3% more than single males, all else equal

E.g., $\text{marriedFemale} \implies$ that married females earn approximately 19.8% less than single males, all else equal
Ordinal Variables

- Consider credit ratings: $CR \in (AAA, AA, ..., C, D)$
- If we want to explain bond interest rates with ratings, we could convert $CR$ to a numeric scale, e.g., $AAA = 1, AA = 2, ...$ and run
  
  $$IR_i = \beta_0 + \beta_1 CR_i + u_i$$

- This assumes a constant linear relation between interest rates and every rating category.
- Moving from AAA to AA produces the same change in interest rates as moving from BBB to BB.
- Could take log interest rate, but is same proportionate change much better?
Converting Ordinal Variables to Binary

- Or we could create an indicator for each rating category, e.g., $CR_{AAA} = 1$ if CR = AAA, 0 otherwise; $CR_{AA} = 1$ if CR = AA, 0 otherwise, etc.

- Run this regression:

$$IR_i = \beta_0 + \beta_1 CR_{AAA} + \beta_2 CR_{AA} + \ldots + \beta_{m-1} CR_C + u_i$$

remembering to exclude one ratings category (e.g., “D”)

- This allows the IR change from each rating category to have a different magnitude

- Each coefficient is the different in IRs between a bond with a certain credit rating (e.g., “AAA”, “BBB”, etc.) and a bond with a rating of “D” (the omitted category).
Interactions Involving Binary Variables I

- Recall the regression with four categories based on (1) marriage status and (2) sex.

\[ \log(\text{wage}) = 0.321 + 0.213\text{marriedMale} - 0.198\text{marriedFemale} \\
+ -0.110\text{singleFemale} + 0.079\text{education} \]

- We can capture the same logic using interactions

\[ \log(\text{wage}) = 0.321 - 0.110\text{female} + 0.213\text{married} \\
+ -0.301\text{female} \times \text{married} + ... \]

- Note excluded category can be found by setting all dummies = 0

\[ \Rightarrow \text{excluded category} = \text{single (married} = 0), \text{male (female} = 0) \]
Interactions Involving Binary Variables II

- Note that the intercepts are all identical to the original regression.
- Intercept for married male
  \[
  \hat{\log(wage)} = 0.321 - 0.110(0) + 0.213(1) - 0.301(0) \times (1) = 0.534
  \]
- Intercept for single female
  \[
  \hat{\log(wage)} = 0.321 - 0.110(1) + 0.213(0) - 0.301(1) \times (0) = 0.211
  \]
- And so on.
- Note that the slopes will be identical as well.
Example: Wages and Computers

- Krueger (1993), $N = 13,379$ from 1989 CPS

\[
\hat{\log(wage)} = \hat{\beta}_0 + 0.177 \text{compwork} + 0.070 \text{comphome} \\
+ 0.017 \text{compwork} \times \text{comphome} + \ldots
\]

(Intercept not reported)

- Base category = people with no computer at work or home
- Using a computer at work is associated with a 17.7% higher wage. (Exact value is $100(\exp(0.177) - 1) = 19.4\%$)
- Using a computer at home but not at work is associated with a 7.0% higher wage.
- Using a computer at home and work is associated with a $100(0.177+0.070+0.017) = 26.4\%$ (Exact value is $100(\exp(0.177+0.070+0.017) - 1) = 30.2\%$)
Different Slopes

- Dummies only shift intercepts for different groups.
- What about slopes? We can interact continuous variables with dummies to get different slopes for different groups. E.g,

\[
\log(wage) = \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + \delta_1 \text{educ} \times \text{female} + u
\]

\[
\log(wage) = (\beta_0 + \delta_0 \text{female}) + (\beta_1 + \delta_1 \text{female}) \text{educ} + u
\]

- Males: Intercept = \(\beta_0\), slope = \(\beta_1\)
- Females: Intercept = \(\beta_0 + \delta_0\), slope = \(\beta_1 + \delta_1\)

\[\Rightarrow\] \(\delta_0\) measures difference in intercepts between males and females

\[\Rightarrow\] \(\delta_1\) measures difference in slopes (return to education) between males and females
\[
\log(wage) = (\beta_0 + \delta_0 \text{female}) + (\beta_1 + \delta_1 \text{female}) \cdot \text{educ} + u
\]

\[
(\delta_0 < 0, \delta_1 < 0)
\]
\[ \log(wage) = (\beta_0 + \delta_0 \text{female}) + (\beta_1 + \delta_1 \text{female}) \text{educ} + u \]

\((\delta_0 < 0, \delta_1 > 0)\)
Interpretation of Figures

- 1st figure: intercept and slope for women are less than those for men
  → women earn less than men at all educational levels
- 2nd figure: intercept for women is less than that for men, but slope is larger
  → women earn less than men at low educational levels but the gap narrows as education increases.
  → at some point, women earn more than men. But, does this point occur within the range of data?
- Point of equality: Set Women eqn = Men eqn

\[
\begin{align*}
\text{Women: } \log(wage) & = (\beta_0 + \delta_0) + (\beta_1 + \delta_1) \text{educ} + u \\
\text{Men: } \log(wage) & = (\beta_0) + \beta_1 \text{educ} + u
\end{align*}
\]

⇒ \( e^* = -\delta_0/\delta_1 \)
Example 1

- Consider $N = 526, R^2 = 0.441$

\[
\hat{\log{(wage)}} = 0.389 - 0.227\text{female} + 0.082\text{educ} \\
- 0.006\text{female} \times \text{educ} + 0.29\text{exper} - 0.0006\text{exper}^2 + \ldots
\]

- Return to education for men = 8.2%, women = 7.6%.
- Women earn 22.7% less than men. But statistically insignif...why?
- Problem is multicollinearity with interaction term.
  - Intuition: coefficient on \textit{female} measure wage differential between men and women when \textit{educ} = 0.
  - Few people have very low levels of \textit{educ} so unsurprising that we can’t estimate this coefficient precisely.
  - More interesting to estimate gender differential at $\bar{educ}$, for example.
  - Just replace \textit{female} $\times$ \textit{educ} with \textit{female} $\times$ (\textit{educ} $- \bar{educ}$) and rerun regression. This will only change coefficient on \textit{female} and its standard error.
Example 2

Consider baseball players salaries $N = 330$, $R^2 = 0.638$

\[
\log(\text{salary}) = 10.34 + 0.0673 \text{years} + 0.009 \text{gamesyr} + \ldots \\
- 0.198 \text{black} - 0.190 \text{hispan} \\
+ 0.0125 \text{black} \times \text{percBlack} + 0.0201 \text{hispan} \times \text{percHisp}
\]

Black players in cities with no blacks ($\text{percBlack} = 0$) earn 19.8% less than otherwise identical whites.

As $\text{percBlack}$ inc ($\implies$ $\text{percWhite}$ dec since $\text{percHisp}$ is fixed), black salaries increase relative to that for whites. E.g., if $\text{percBalck} = 10\% \implies$ blacks earn $-0.198 + 0.0125(10) = -0.073$, 7.3% less than whites in such a city.

When $\text{percBlack} = 20\% \implies$ blacks earn 5.2% more than whites.

Does this $\implies$ discrimination against whites in cities with large black pop? Maybe best black players choose to live in such cities.
Any misspecification in the functional form relating dependent variable to the independent variables will lead to bias.

E.g., assume true model is

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + u \]

but we omit squared term, \( x_2^2 \).

Amount of bias in \((\beta_0, \beta_1, \beta_2)\) depends on size of \(\beta_3\) and correlation among \((x_1, x_2, x_2^2)\).

Incorrect functional form on the LHS will bias results as well (e.g., \(\log(y)\) vs. \(y\)).

This is a minor problem in one sense: we have all the sufficient data, so we can try/test as many different functional forms as we like.

This is different from a situation where we don’t have data for a relevant variable.
Regression Error Specification Test (RESET)

Estimate

\[ y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u \]

Compute predicted values \( \hat{y} \)

Estimate

\[ y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + u \]

(choice of polynomial is arbitrary.)

\[ H_0 : \delta_1 = \delta_2 = 0 \]

Use F-test with \( F \sim F_{2, n-k-3} \)
What if we wanted to test 2 nonnested models? I.e., we can’t simply restrict parameters in one model to obtain the other.

E.g.,

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \]

vs.

\[ y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + u \]

E.g.,

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \]

vs.

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 z + u \]
Davidson-MacKinnon Test

- Test

Model 1: \[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \]

Model 2: \[ y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + u \]

- If 1st model is correct, then fitted values from 2nd model, (\( \hat{y} \)), should be insignificant in 1st model.
- Look at t-stat on \( \theta_1 \) in

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \theta_1 \hat{y} + u \]

- Significant \( \theta_1 \) \( \implies \) rejection of 1st model.
- Then do reverse, look at t-stat on \( \theta_1 \) in

\[ y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \theta_1 \hat{y} + u \]

where \( \hat{y} \) are predicted values from 1st model.
- Significant \( \theta_1 \) \( \implies \) rejection of 2nd model.
Davidson-MacKinnon Test: Comments

- Clear winner need not emerge. Both models could be rejected or neither could be rejected.
- In latter case, could use $R^2$ to choose.
- Practically speaking, if the effects of key independent variables on $y$ are not very different, the it doesn’t really matter which model is used.
- Rejecting one model does not imply that the other model is correct.
Consider

\[ \log(wage) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{ability} + u \]

We don’t observe or can’t measure ability.

⇒ coefficients are unbiased.

What can we do?

Find a **proxy variable**, which is correlated with the unobserved variable. E.g., IQ.
Proxy Variables

- Consider

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3^* + u \]

- \( x_3^* \) is unobserved but we have proxy, \( x_3 \)
- \( x_3 \) should be related to \( x_3^* \):

\[ x_3^* = \delta_0 + \delta_1 x_3 + \nu_3 \]

where \( \nu_3 \) is error associated with the proxy’s imperfect representation of \( x_3^* \)

- Intercept is just there to account for different scales (e.g., ability may have a different average value than IQ)
Can we just substitute $x_3$ for $x_3^*$? (and run)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

Depends on the assumptions on $u$ and $v_3$.

1. $E(u|x_1, x_2, x_3^*) = 0$ (Common assumption). In addition, $E(u|x_3) = 0 \implies x_3$ is irrelevant once we control for $(x_1, x_2, x_3^*)$ (Need this but not controversial given 1st assumption and status of $x_3$ as a proxy)

2. $E(v_3|x_1, x_2, x_3) = 0$. This requires $x_3$ to be a good proxy for $x_3^*$

$$E(x_3^*|x_1, x_2, x_3) = E(x_3^*|x_3) = \delta_0 + \delta_1 x_3$$

Once we control for $x_3$, $x_3^*$ doesn’t depend on $x_1$ or $x_2$
Recall true model

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3^* + u \]

Substitute for \( x_3^* \) in terms of proxy

\[
\begin{align*}
    y &= \left( \beta_0 + \beta_3 \delta_0 \right) + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \delta_3 x_3 + u + \beta_3 v_3 \\
    &= \alpha_0 + \beta_1 x_1 + \beta_2 x_2 + \alpha_3 x_3 + e
\end{align*}
\]

Assumptions 1 & 2 on prev slide \( \implies E(e|x_1, x_2, x_3) = 0 \implies \) we can est.

\[ y = \alpha_0 + \beta_1 x_1 + \beta_2 x_2 + \alpha_3 x_3 + e \]

Note: we get unbiased (or at least consistent) estimators of \((\alpha_0, \beta_1, \beta_2, \alpha_3)\).

\((\beta_0, \beta_3)\) not identified.
Example 1: Plug-In Solution

- In wage example where IQ is a proxy for ability, the 2nd assumption is

\[ E(\text{ability}|\text{educ, exper, IQ}) = E(\text{ability}|IQ) = \delta_0 + \delta_3 IQ \]

- This means that the average level of ability only changes with IQ, not with education or experience.

- Is this true? Can’t test but must think about it.
Example 1: Cont.

- If proxy variable doesn’t satisfy the assumptions 1 & 2, we’ll get biased estimates
- Suppose
  \[ x^*_3 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + v_3 \]
  where \( E(v_3|x_1, x_2, x_3) = 0 \).
- Substitute into structural eqn
  \[ y = (\beta_0 + \beta_3 \delta_0) + (\beta_1 + \beta_3 \delta_1)x_1 + (\beta_2 + \beta_3 \delta_2)x_2 + \beta_3 \delta_3 x_3 + u + \beta_3 v_3 \]
- So when we estimate the regression:
  \[ y = \alpha_0 + \beta_1 x_1 + \beta_2 x_2 + \alpha_3 x_3 + e \]
  we get consistent estimates of \( (\beta_0 + \beta_3 \delta_0) \), \( (\beta_1 + \beta_3 \delta_1) \), \( (\beta_2 + \beta_3 \delta_2) \), and \( \beta_3 \delta_3 \) assuming \( E(u + \beta_3 v_3|x_1, x_2, x_3) = 0 \).
- Original parameters are not identified.
Example 2: Plug-In Solution

- Consider q-theory of investment
  \[ \ln v = \beta_0 + \beta_1 q + u \]
- Can’t measure q so use proxy, market-to-book (MB),
  \[ q = \delta_0 + \delta_1 MB + \nu \]
- Think about identifying assumptions
  1. \( E(u|q) = 0 \) theory say q is sufficient statistic for inv
  2. \( E(q|MB) = \delta_0 + \delta_1 MB \implies \) avg level of q changes only with MB
- Even if assumption 2 true, we’re not estimating \( \beta_1 \) in
  \[ \ln v = \alpha_0 + \alpha_1 MB + e \]
- We’re estimating \((\alpha_0, \alpha_1)\) where
  \[ \ln v = \left( \beta_0 + \beta_1 \delta_0 \right) + \beta_1 \delta_1 MB + e \]
  with
  \[ \alpha_0 \] and \[ \alpha_1 \]
Let’s say we have no idea how to proxy for an omitted variable. One way to address is to use the lagged dependent variable, which captures inertial effects of all factors that affect $y$. This is unlikely to solve the problem, especially if we only have one cross-section. But, we can conduct the experiment of comparing to observations with the same value for the outcome variable last period. This is imperfect, but it can help when we don’t have panel data.
Model 1

- Consider an extension to the basic model

\[ y_i = \alpha_i + \beta_i x_i \]

where \( \alpha_i \) is an unobserved intercept and the return to education differs for each person.

- This model is unidentified: more parameters \((2n)\) than observations \((n)\)

- But we can hope to identify avg intercept, \( E(\alpha_i) = \alpha \), and avg slope, \( E(\beta_i) = \beta \) (a.k.a., **Average Partial Effect (APE)**).

\[ \alpha_i = \alpha + c_i, \beta_i = \beta + d_i \]

where \( c_i \) and \( d_i \) are the individual specific deviation from average effects.

\[ \Rightarrow E(c_i) = E(d_i) = 0 \]
Model II

- Substitute coefficient specification into model

\[ y_i = \alpha + \beta x_i + c_i + d_i x_i \equiv \alpha + \beta x_i + u_i \]

- What we need for unbiasedness is \( E(u_i|x_i) = 0 \)

\[ E(u_i|x_i) = E(c_i + d_i x_i|x_i) \]

- This amounts to requiring

  1. \( E(c_i|x_i) = E(c_i) = 0 \implies E(\alpha_i|x_i) = E(\alpha_i) \)
  2. \( E(d_i|x_i) = E(d_i) = 0 \implies E(\beta_i|x_i) = E(\beta_i) \)

- Understand these assumptions!!!! In order for OLS to consistently estimate the mean slope and intercept, the slopes and intercepts must be mean independent (at least uncorrelated) of the explanatory variable.
When we use an imprecise measure of an economic variable in a regression, our model contains measurement error (ME).

- The market-to-book ratio is a noisy measure of “q”
- Altman’s Z-score is a noisy measure of the probability of default
- Average tax rate is a noisy measure of marginal tax rate
- Reported income is noisy measure of actual income

Similar statistical structure to omitted variable-proxy variable solution but conceptually different

- Proxy variable case we need variable that is associated with unobserved variable (e.g., IQ proxy for ability)
- Measurement error case the variable we don’t observe has a well-defined, quantitative meaning but our recorded measure contains error
Let $y$ be observed measure of $y^*$

$$y^* = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u$$

Measurement error defined as $e_0 = y - y^*$

Estimable model is:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u + e_0$$

If mean of ME $\neq 0$, intercept is biased so assume mean $= 0$

If ME independent of $X$, then OLS is unbiased and consistent and usual inference valid.

If $e_0$ and $u$ uncorrelated than $Var(u + e_0) > Var(u) \implies$ measurement error in dependent variable results in larger error variance and larger coef SEs
Measurement Error in Log Dependent Variable

- When $\log(y^*)$ is dependent variable, we assume

  $$\log(y) = \log(y^*) + e_0$$

- This follows from multiplicative ME

  $$y = y^* a_0$$

  where

  $$a_0 > 0$$

  $$e_0 = \log(a_0)$$
Measurement Error in Independent Variable

- Model
  \[ y = \beta_0 + \beta_1 x_1^* + u \]

- ME defined as \( e_1 = x_1 - x_1^\star \)

- Assume
  - Mean ME = 0
  - \( u \perp x_1^*, x_1 \), or \( E(y|x_1^*, x_1) = E(y|x_1^*) \) (i.e., \( x_1 \) doesn’t affect \( y \) after controlling for \( x_1^* \))

- What are implications of ME for OLS properties?
  - Depends crucially on assumptions on \( e_1 \)
  - Econometrics has focused on 2 assumptions
Assumption 1: $e_1 \perp x_1$

- $1^{st}$ assumption is ME uncorrelated with observed measure
- Since $e_1 = x_1 - x_1^*$, this implies $e_1 \perp x_1^*$
- Substitute into regression

$$y = \beta_0 + \beta_1 x_1 + (u - \beta_1 e_1)$$

- We assumed $u$ and $e_1$ have mean 0 and are $\perp$ with $x_1$

$\Rightarrow$ $(u - \beta_1 e_1)$ is uncorrelated with $x_1$.

$\Rightarrow$ OLS with $x_1$ produces consistent estimator of coef’s

$\Rightarrow$ OLS error variance is $\sigma_u^2 + \beta_1^2 \sigma_{e_1}^2$

- ME increases error variance but doesn’t affect any OLS properties (except coef SEs are bigger)
Assumption 2: $e_1 \perp x_1^*$

- This is the **Classical Errors-in-Variables (CEV)** assumption and comes from representation:

  $$x_1 = x_1^* + e_1$$

- (Still maintain 0 correlation between $u$ and $e_1$)

- Note $e_1 \perp x_1^* \implies$

  $$\text{Cov}(x_1, e_1) = E(x_1 e_1) = E(x_1^* e_1) + E(e_1^2) = \sigma_{e_1}^2$$

- This covariance causes problems when we use $x_1$ in place of $x_1^*$ since

  $$y = \beta_0 + \beta_1 x_1 + (u - \beta_1 e_1)$$ and

  $$\text{Cov}(x_1, u - \beta_1 e_1) = -\beta_1 \sigma_{e_1}^2$$

- I.e., indep var is correlated with error $\implies$ bias and inconsistent OLS estimates
Assumption 2: $e_1 \perp x_1^*$ (Cont.)

- Amount of inconsistency in OLS

\[
\text{plim}(\hat{\beta}_1) = \beta_1 + \frac{\text{Cov}(x_1, u - \beta_1 e_1)}{\text{Var}(x_1)} \\
= \beta_1 + \frac{\beta_1 \sigma_{e_1}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2} \\
= \beta_1 \left(1 - \frac{\sigma_{e_1}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2}\right) \\
= \beta_1 \left(\frac{\sigma_{x_1^*}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2}\right)
\]
CEV asymptotic bias

- From previous slide:
  \[ \text{plim}(\hat{\beta}_1) = \beta_1 \left( \frac{\sigma_{x_1}^2}{\sigma_{x_1}^2 + \sigma_{e_1}^2} \right) \]

- Scale factor is always \(< 1 \implies\) asymptotic bias attenuates estimated effect (attenuation bias)
- If variance of error \(\sigma_{e_1}^2\) is small relative to variance of unobserved factor, then bias is small.
- More than 1 explanatory variable and bias is less clear
- Correlation between \(e_1\) and \(x_1\) creates problem. If \(x_1\) correlated with other variables, bias infects everything.
- Generally, measurement error in a single variable causes inconsistency in all estimators. Sizes and even directions of the biases are not obvious or easily derived.
Counterexample to CEV Assumption

Consider

\[ \text{colGPA} = \beta_0 + \beta_1 \text{smoked}^* + \beta_2 \text{hsGPA} + u \]
\[ \text{smoked} = \text{smoked}^* + e_1 \]

where \( \text{smoked}^* \) is actual \# of times student smoked marijuana and \( \text{smoked} \) is reported

- For \( \text{smoked}^* = 0 \) report is likely to be 0 \( \implies e_1 = 0 \)
- For \( \text{smoked}^* > 0 \) report is likely to be off \( \implies e_1 \neq 0 \)

\( \implies e_1 \) and \( \text{smoked}^* \) are correlated estimated effect (attenuation bias)

I.e., CEV Assumption does not hold

Tough to figure out implications in this scenario
At a basic level, regression is just math (linear algebra and projection methods)

We don’t need statistics to run a regression (i.e., compute coefficients, standard errors, sums-of-squares, $R^2$, etc.)

What we need statistics for is the interpretation of these quantities (i.e., for statistical inference).

From the regression equation, the statistical properties of $y$ come from those of $X$ and $u$
What is heteroskedasticity (HSK)?

- Non-constant variance, that’s it.
- HSK has no effect on bias or consistency properties of OLS estimators
- HSK means OLS estimates are no longer BLUE
- HSK means OLS estimates of standard errors are incorrect
- We need an HSK-robust estimator of the variance of the coefficients.
Eicker (1967), Huber (1967), and White (1980) suggest:

$$\text{Var}(\hat{\beta}_j) = \frac{\sum_{i=1}^{N} \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}$$

where \( \hat{r}_{ij}^2 \) is the \( i \)th residual from regressing \( x_j \) on all other independent variables, and \( SSR_j \) is the sum of square residuals from this regression.

- Use this in computation of t-stas to get an HSK-robust t-statistic
- Why use non-HSK-robust SEs at all?
- With small sample sizes robust t-stats can have very different distributions (non “t”)

Michael R. Roberts
Linear Regression 106/129
HSK-Robust LM-Statistics

The recipe:

1. Get residuals from restricted model \( \tilde{u} \)
2. Regress each independent variable excluded under null on all of the included independent variables; \( q \) excluded variables \( \Rightarrow (\tilde{r}_1, ..., \tilde{r}_q) \)
3. Compute the products between each vector \( \tilde{r}_j \) and \( \tilde{u} \)
4. Regression of 1 (a constant “1” for each observation) on all of the products \( \tilde{r}_j \tilde{u} \) without an intercept
5. HSK-robust LM statistic, \( LM \), is \( N - SSR_1 \), where \( SSR_1 \) is the sum of squared residuals from this last regression.
6. \( LM \) is asymptotically distributed \( \chi^2_q \)
Testing for HSK

- The model
  \[ y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u \]

- Test \( H_0 : \text{Var}(y|x_1, \ldots, x_k) = \sigma^2 \)
  \[ E(u|x_1, \ldots, x_k) = 0 \implies \text{this hypothesis is equivalent to} \]
  \[ H_0 : E(u^2|x_1, \ldots, x_k) = \sigma^2 \] (i.e., is \( u^2 \) related to any explanatory variables?)
  \[ \hat{u}^2 = \delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k + u \]

- Test null \( H_0 : \delta_1 = \ldots = \delta_k = 0 \)
  \[
  \text{F-test} : \quad F = \frac{R_{\hat{u}^2}^2}{1 - R_{\hat{u}^2}^2/(n - k - 1)}
  \]
  \[
  \text{LM-test} : \quad LM = N \times R_{\hat{u}^2}^2 \quad \text{(BP-test sort of)}
  \]
Weighted Least Squares (WLS)

- Pre HSK-robust statistics, we did WLS - more efficient than OLS if correctly specified variance form

\[ \text{Var}(u|X) = \sigma^2 h(X), \ h(X) > 0 \forall X \]

- E.g., \( h(X) = x_1^2 \) or \( h(x) = \exp(x) \)

- WLS just normalizes all of the variables by the square root of the variance fxn \( (\sqrt{h(X)}) \) and runs OLS on transformed data.

\[
\begin{align*}
  y_i/\sqrt{h(X_i)} &= \beta_0/\sqrt{h(X_i)} + \beta_1/(x_{i1}/\sqrt{h(X_i)}) + \ldots \\
  &\quad + \beta_k/(x_{ik}/\sqrt{h(X_i)}) + u_i/\sqrt{h(X_i)} \\
  y_i^* &= \beta_0 x_0^* + \beta_1 x_1^* + \ldots + \beta_k x_k^* + u^* \\
\end{align*}
\]

where \( x_0^* = 1/\sqrt{h(X_i)} \)
WLS is an example of a **Generalized Least Squares** Estimator

Consider

\[ \text{Var}(u|X) = \sigma^2 \exp\delta_0 + \delta x_1 \]

We need to estimate variance parameters. Using estimates gives us FGLS
Feasible Generalized Least Squares (FGLS) Recipe

- Consider variance form:
  \[
  \text{Var}(u|X) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k)
  \]

- FGLS to correct for HSK:
  1. Regress \( y \) on \( X \) and get residuals \( \hat{u} \)
  2. Regress \( \log(\hat{u}^2) \) on \( X \) and get fitted values \( \hat{g} \)
  3. Estimate by WLS

\[
\hat{y} = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u
\]

  with weights \( 1/exp(\hat{g}) \), or transform each variable (including intercept) by multiplying by \( 1/exp(\hat{g}) \) and estimate via OLS

- FGLS estimate is biased but consistent and more efficient than OLS.
If coefficient estimates are very different across OLS and WLS, it’s likely \( E(y|x) \) is misspecified.

If we get variance form wrong in WLS then

1. WLS estimates are still unbiased and consistent
2. WLS standard errors and test statistics are invalid even in large samples
3. WLS may not be more efficient than OLS
Single Parameter Tests

- **Model**

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u
\]

- Under certain assumptions

\[
t(\hat{\beta}_j) = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}
\]

- Under other assumptions, asymptotically \( t \sim N(0, 1) \)

- **Intuition:** \( t(\hat{\beta}_j) \) tells us how far – in standard deviations – our estimate \( \hat{\beta}_j \) is from the hypothesized value \( (\beta_j) \)

- **E.g.,** \( H_0 : \beta_j = 0 \quad \Rightarrow \quad t = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \)

- **E.g.,** \( H_0 : \beta_j = 4 \quad \Rightarrow \quad t = \frac{(\hat{\beta}_j - 4)}{se(\hat{\beta}_j)} \)
Statistical vs. Economic Significance

- These are not the same thing
- We can have a statistically insignificant coefficient but it may be economically large.
  - Maybe we just have a power problem due to a small sample size, or little variation in the covariate
- We can have a statistically significant coefficient but it may be economically irrelevant.
  - Maybe we have a very large sample size, or we have a lot of variation in the covariate (outliers)
- You need to think about both statistical and economic significance when discussing your results.
Model

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u \]

Are two parameters the same? I.e.,

\[ H_0 : \beta_1 = \beta_2 \iff (\beta_1 - \beta_2) = 0 \]

The usual statistic can be slightly modified

\[
  t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{se(\hat{\beta}_1 - \hat{\beta}_2)} \sim t_{n-k-1}
\]

Careful: when computing the SE of difference not to forget covariance term

\[
  se(\hat{\beta}_1 - \hat{\beta}_2) = \left( se(\hat{\beta}_1)^2 + se(\hat{\beta}_2)^2 - 2Cov(\hat{\beta}_1, \hat{\beta}_2) \right)^{1/2}
\]
Instead of dealing with computing the SE of difference, can reparameterize the regression and just check a t-stat.

E.g., define $\theta = \beta_1 - \beta_2 \implies \beta_1 = \theta + \beta_2$ and

$$y = \beta_0 + (\theta + \beta_2)x_1 + \beta_2x_2 + \ldots + \beta_kx_k + u$$

$$= \beta_0 + \theta x_1 + \beta_2(x_1 + x_2) + \ldots + \beta_kx_k + u$$

Just run a t-test of new null, $H_0 : \theta = 0$ same as previous slide.

This strategy always works.
Testing Multiple Linear Restrictions

Consider $H_0: \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$ (a.k.a., exclusion restrictions), $H_1: H_0$ not true

To test this, we need a **joint hypothesis test**

One such test is as follows:

1. **Estimate the Unrestricted Model**
   
   $$ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u $$

2. **Estimate the Restricted Model**
   
   $$ y = \beta_0 + \beta_4 x_4 + \beta_5 x_5 + \ldots + \beta_k x_k + u $$

3. **Compute $F$-statistic**
   
   $$ F = \frac{SSR_R - SSR_U}{q} \left/ \frac{SSR_U}{(n - k - 1)} \right. 
   \sim F_{q, n - k - 1} $$

   where $q =$ degrees of freedom (df) in numerator $= df_R - df_U$, $n - k - 1 =$ df in denominator $= df_U,$
Relationship Between $F$ and $t$ Statistics

- $t^2_{n-k-1}$ has an $F_{1,n-k-1}$ distribution.
- All coefficients being individually statistically significant (significant $t$-stats) does not imply that they are jointly significant.
- All coefficients being individually statistically insignificant (insignificant $t$-stats) does not imply that they are jointly insignificant.
- $R^2$ form of the F-stat:

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

(Equivalent to previous formula.)

- “Regression F-Stat” tests $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$
Testing General Linear Restrictions I

Can write any set of linear restrictions as follows

\[ H_0 : R\beta - q = 0 \]
\[ H_1 : R\beta - q \neq 0 \]

\( \text{dim}(R) = \# \text{ of restrictions} \times \# \text{ of parameters.} \) E.g.,

\[ H_0 : \beta_j = 0 \implies R = [0, 0, \ldots, 1, 0, \ldots, 0], q = 0 \]
\[ H_0 : \beta_j = \beta_k \implies R = [0, 0, 1, \ldots, -1, 0, \ldots, 0], q = 0 \]
\[ H_0 : \beta_1 + \beta_2 + \beta_3 = 1 \implies R = [1, 1, 1, 0, \ldots, 0], q = 1 \]
\[ H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0 \implies \]
\[
R = \begin{bmatrix}
1 & 0 & 0 & 0 & \ldots & 0 \\
0 & 1 & 0 & 0 & \ldots & 0 \\
0 & 0 & 1 & 0 & \ldots & 0
\end{bmatrix},
q = \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}
\]
Testing General Linear Restrictions II

- Note that _under the null hypothesis_

\[
E(R\hat{\beta} - q|X) = R\beta_0 - q = 0
\]
\[
Var(R\hat{\beta} - q|X) = RVar(\hat{\beta}|X)R' = \sigma^2 R(X'X)^{-1}R'
\]

- Wald criterion:

\[
W = (R\hat{\beta} - q)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q) \sim \chi^2_J
\]

where \( J \) is the degrees of freedom under the null (i.e., the \# of restrictions, the \# of rows in \( R \))

- Must estimate \( \sigma^2 \), this changes distribution

\[
F = (R\hat{\beta} - q)'[\hat{\sigma}^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q) \sim F_{J,n-k-1}
\]

where the \( n - k - 1 \) are df of the denominator \( (\sigma^2) \)
Differences in Regression Function Across Groups I

- Consider

\[ \text{cumgpa} = \beta_0 + \beta_1 \text{sat} + \beta_2 \text{hsperc} + \beta_3 \text{tothrs} + u \]

where \( \text{sat} = \) SAT score, \( \text{hsperc} = \) high school rank percentile, \( \text{tothrs} = \) total hours of college courses.

- Does this model describe the college GPA for male \textit{and} females?

- Can allow intercept and slopes to vary by sex as follows:

\[
\text{cumgpa} = \beta_0 + \delta_0 \text{female} + \beta_1 \text{sat} + \delta_1 \text{sat} \times \text{female} \\
+ \beta_2 \text{hsperc} + \delta_2 \text{hsperc} \times \text{female} \\
+ \beta_3 \text{tothrs} + \delta_3 \text{tothrs} \times \text{female} + u
\]

- \( H_0 : \delta_0 = \delta_1 = \delta_2 = \delta_3 = 0 \), \( H_1 : \) At least one \( \delta \) is non-zero.
Differences in Regression Function Across Groups II

- We can estimate the interaction model and compute the corresponding F-test using the statistic from above
  \[ F = (R\hat{\beta} - q)'[\hat{\sigma}^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q) \sim F_{J,n-k-1} \]

- We can estimate the restricted (assume female = 0) and unrestricted versions of the model. Compute F-statistic as (will be identical)
  \[ F = \frac{SSR_R - SSR_U}{SSR_U} \frac{n - 2(J)}{J} \]

  where \( SSR_R = \) sum of squares of restricted model, \( SSR_U = \) sum of squares of unrestricted model, \( n = \) total \# of obs, \( k = \) total \# of explanatory variables \textit{excluding} intercept, \( J = k + 1 \) total \# of restrictions (we restrict all \( k \) slopes and intercept).

- \( H_0 : \delta_0 = \delta_1 = \delta_2 = \delta_3 = 0, \ H_1 : \) At least one \( \delta \) is non-zero.
Chow Test

- What if we have a lot of explanatory variables? Unrestricted model will have a lot of terms.
- Imagine we have two groups, \( g = 1, 2 \)
- Test whether intercept and slopes are same across two groups.
  Model is:
  \[
  y = \beta_{g,0} + \beta_{g,1}x_1 + \ldots + \beta_{g,k}x_k + u
  \]
  - \( H_0 : \beta_{1,0} = \beta_{2,0}, \beta_{1,1} = \beta_{2,1}, \ldots, \beta_{1,k} = \beta_{2,k} \)
  - Null \( \iff \) \( k + 1 \) restrictions (slopes + intercept). E.g., in GPA example, \( k = 3 \)
Chow Test Recipe

- Chow test form of F-stat from above:

\[ F = \frac{SSR_P - (SSR_1 + SSR_2)}{SSR_1 + SSR_2} \cdot \frac{n - 2(k + 1)}{k + 1} \]

1. Estimate pooled (i.e., restricted) model with no interactions and save \( SSR_P \)
2. Estimate model on group 1 and save \( SSR_1 \)
3. Estimate model on group 2 and save \( SSR_2 \)
4. Plug into F-stat formula.

- Often used to detect a structural break across time periods.
- Requires homoskedasticity.
If

1. $u$ are i.i.d. with mean 0 and variance $\sigma^2$, and
2. $x$ meet Grenander conditions (look it up), then

$$\hat{\beta} \xrightarrow{a} N \left[ \beta, \frac{\sigma^2}{n} Q^{-1} \right]$$

where $Q = \text{plim}(X'X/n)$

- Basically, under fairly weak conditions, OLS estimates are asymptotically normal and centered around the true parameter values.
The Delta Method

- How do we compute variance of nonlinear function of random variables? Use a Taylor expansion around the expectation.

- If \( \sqrt{n}(z_n - \mu) \xrightarrow{d} N(0, \sigma^2) \) and \( g(z_n) \) is continuous function not involving \( n \), then

\[
\sqrt{n}(g(z_n) - g(\mu)) \xrightarrow{d} N(0, g'(\mu)^2 \sigma^2)
\]

- If \( Z_n \) is \( K \times 1 \) sequence of vector-valued random variables:

\[
\sqrt{n}(Z_n - M) \xrightarrow{d} N(0, \Sigma) \text{ and } C(Z_n) \text{ is a set of } J \text{ continuous functions not involving } n,
\]

then

\[
\sqrt{n}(C(Z_n) - C(M)) \xrightarrow{d} N(0, G(M)\Sigma G(M)')
\]

where \( G(M) \) is the \( J \times K \) matrix \( \partial C(M)/\partial M' \). The \( j \)th row of \( G(M) \) is the vector of partial derivatives of the \( j \)th fxn with respect to \( M' \).
The Delta Method in Action

- Consider two estimators \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) of \( \beta_1 \) and \( \beta_2 \):

\[
\begin{bmatrix}
\hat{\beta}_1 \\
\hat{\beta}_2 \\
\end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\
0 \\
\end{bmatrix}, \Sigma \right)
\]

where \( \Sigma = \begin{pmatrix}
\sigma_{11} & \sigma_{12} \\
\sigma_{21} & \sigma_{22}
\end{pmatrix} \)

- What is asymptotic distribution of \( f(\hat{\beta}_1, \hat{\beta}_2) = \hat{\beta}_1/(1 - \hat{\beta}_2) \)

\[
\begin{align*}
\frac{\partial f}{\partial \beta_1} &= \frac{1}{1 - \beta_2} \\
\frac{\partial f}{\partial \beta_2} &= \frac{\beta_1}{(1 - \beta_2)^2}
\end{align*}
\]

\[
\text{AVar } f(\hat{\beta}_1, \hat{\beta}_2) = \left( \frac{1}{1 - \beta_2} \frac{\beta_1}{(1 - \beta_2)^2} \right) \Sigma \left( \frac{1}{1 - \beta_2} \frac{\beta_1}{(1 - \beta_2)^2} \right)
\]
Reporting Regression Results

A table of OLS regression output should show the following:

1. the dependent variable,
2. the independent variables (or a subsample and description of the other variables),
3. the corresponding estimated coefficients,
4. the corresponding standard errors (or t-stats),
5. stars by the coefficient to indicate the level of statistical significance, if any (1 star for 5%, 2 stars for 1%),
6. the adjusted $R^2$, and
7. the number of observations used in the regression.

In the body of paper, focus discussion on variable(s) of interest: sign, magnitude, statistical & economic significance, economic interpretation.

Discuss “other” coefficients if they are “strange” (e.g., wrong sign, huge magnitude, etc.)
### Example: Reporting Regression Results

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry Avg. Leverage</td>
<td>0.067**</td>
<td>0.053**</td>
<td>0.018**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35.179)</td>
<td>(25.531)</td>
<td>(7.111)</td>
<td></td>
</tr>
<tr>
<td>Log(Sales)</td>
<td>0.022**</td>
<td>0.017**</td>
<td>0.018**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(11.861)</td>
<td>(8.996)</td>
<td>(9.036)</td>
<td></td>
</tr>
<tr>
<td>Market-to-Book</td>
<td>-0.024**</td>
<td>-0.017**</td>
<td>-0.018**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-17.156)</td>
<td>(-12.175)</td>
<td>(-12.479)</td>
<td></td>
</tr>
<tr>
<td>EBITDA / Assets</td>
<td>-0.035**</td>
<td>-0.035**</td>
<td>-0.036**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-20.664)</td>
<td>(-20.672)</td>
<td>(-20.955)</td>
<td></td>
</tr>
<tr>
<td>Net PPE / Assets</td>
<td>0.049**</td>
<td>0.031**</td>
<td>0.045**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(24.729)</td>
<td>(15.607)</td>
<td>(16.484)</td>
<td></td>
</tr>
<tr>
<td>Firm Fixed Effects</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Industry Fixed Effects</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Year Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Obs</td>
<td>77,328</td>
<td>78,189</td>
<td>77,328</td>
<td>77,328</td>
</tr>
<tr>
<td>Adj. R²</td>
<td>0.118</td>
<td>0.113</td>
<td>0.166</td>
<td>0.187</td>
</tr>
</tbody>
</table>