Aug 30 → Sept 2

$Var[\hat{\beta}] = (x'x)^{-1}x' \cdot \sigma^2 \cancel{I} x (x'x)^{-1}$

const. error term var

$= \sigma^2 \cancel{(x'x)^{-1} x'x} (x'x)^{-1}$

$Var[\hat{\beta}] = \sigma^2 (x'x)^{-1}$

* How computer will calc unless otherwise specified

* dont know σ^2

$\hat{\sigma}^2 = \sum_{i=1}^{n} \cancel{\frac{\hat{u}_i - \hat{u}_i}{n-1}} \left(\frac{\hat{u}_i - \hat{u}_i}{n-k}\right)^2 \to 0$

$= \sum_{i=1}^{n} \frac{(\hat{u}_i)^2}{n-k}$

$= \frac{RSS}{n-k}$

$= \frac{\hat{u}'\hat{u}}{n-k}$

- can derive matrix by scaler
- inverse of matrix = division in scaler

$\boxed{\text{So}}$

$\widehat{Var}[\hat{\beta}] = \hat{\sigma}^2 (x'x)^{-1}$

$$\widehat{Var}[\hat{\beta}] = \begin{bmatrix} \widehat{Var}[\hat{\beta}_1] & & \\ & \widehat{Var}[\hat{\beta}_2] & cov \\ cov & & \widehat{Var}[\hat{\beta}_k] \end{bmatrix}$$

standard error est:

$\sqrt{Var[\hat{\beta}]}$

* we now have $\hat{\beta}$ + standard error est. w/ $\hat{\beta}$

- $Var\ \hat{\beta}_1 = E[\hat{\beta}_1 - E[\hat{\beta}]]^2]$
  - expected sqd deviation of $\hat{\beta}_1$ from mean $\hat{\beta}_1 \to \beta_1$
  - Dispersion of $\hat{\beta}_1$ from β_1
  - measure of precision

- $\hat{\beta}$ is linear function of y

Aug 30 → Sept 2

Covariance matrix:

$V[x_1] = E[(x_1 - E(x_1))^2] = \sigma_1^2$

↳ "measure of dispersal"

est $\hat{\sigma}_1^2 = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}{n-1} \rightarrow E[\hat{\sigma}_1^2] = \sigma_1^2$

· unbias estimator

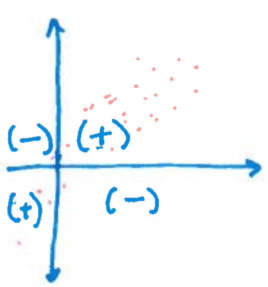
$\text{cov}(x_1, x_2) = E[(x_1 - \underbrace{E(x_1)}_{\mu_1})(x_2 - \underbrace{E(x_2)}_{\mu_2})]$

↳ deviation of x_1 from mean and x_2 from mean

· cov. of itself = variance

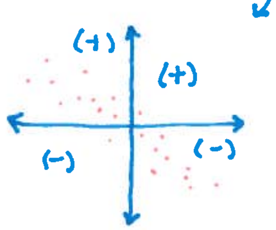
est $\hat{\sigma}_{12}^2 = \sum_{i=1}^{n} \frac{(x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{(n-1)} \rightarrow E[\hat{\sigma}_{12}^2] = \sigma_{12}^2$

equations



- if most values in Σ are pos. expected value of deviation is pos.

- if negitive correlation → negitive covariance



- units are x_1 unit · x_2 unit

↳ so → $\$/\text{lb} + \$/\text{lb} \rightarrow \$/\text{lb}^2$

↳ value of cov depend on x_1/x_2 units
the value itself isn't meaningful

[Correlation fixes this!]

$\text{corr}[x_1, x_2] = \frac{\sigma_{12}^2}{\sigma_1 \sigma_2} = \frac{\sigma_{12}^2}{\sqrt{\sigma_1^2 \sigma_2^2}} = \rho_{12}$

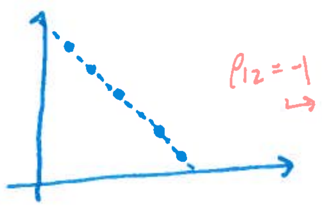
est $\hat{\rho}_{12}$: $\frac{\hat{\sigma}_{12}^2}{\sqrt{\hat{\sigma}_1^2 \hat{\sigma}_2^2}} - E[\hat{\rho}_{12}] = \rho_{12}$

*same sign as covariance
*gets rid of measurement [No units]

Aug 30 – Sept 2 cont

· Correlation coef does not Δ w/ Δ units.

↳ Bounded b/w 1 & -1



$\rho_{12} = -1$

↳ slope doesn't matter

· Measures dispersion of x_1 & x_2 from line

↳ more dispersed? closer to $\emptyset$

↳ No pattern → $\rho = \emptyset$

<u>Apply</u>

Cov Matrix:

Var[U] → AKA covariance matrix for U

↳ $\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}$

$$\text{Var}[U] = E\Big[\underbrace{(U - E[U])}_{(n \times 1)}\underbrace{(U - E[U])'}_{(1 \times n)}\Big] \quad (n \times n)$$

$\boxed{\text{So}}$

$$E\left[\begin{bmatrix} U_1 - E[U_1] \\ U_2 - E[U_2] \\ \vdots \\ U_n - E[U_n] \end{bmatrix}\begin{bmatrix} U_1 - E[U_1], & U_2 - E[U_2] \cdots & U_n - E[U_n] \end{bmatrix}\right]$$

$\boxed{\text{Gives you}}$

$$\text{Var}[U] = E\begin{bmatrix} \sigma_{u_1}^2 & \text{cov}\,u_1u_2 \cdots & \text{cov}\,u_1u_n \\ \text{cov}\,u_2u_1 & \sigma_{u_2}^2 & \\ \vdots & & \\ \text{cov}\,u_nu_1 & & \sigma_n^2 \end{bmatrix}$$

same

same

*symmetric matrix

*No auto correlation? → all covar are $\emptyset$

*Yes Homoskidisity? → all var are same

$\begin{bmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ 0 & & \sigma^2 \end{bmatrix}$ → $\sigma^2 I$ if error term is iid

iid: independently (no autocorr) + identically distributed (no heteroskd)

Aug 30 - Sept 2 cont.

Var of v: error term cov matrix

- var on diag; off diag is covar
- iid: var same + cov is $\emptyset$

Calc OLS standard error

- Cov Matrix $\hat{B}$ v. $\hat{U}$ [ ]
- $\hat{B}$ is random var
  - ↳ $\hat{B} = (x'x)^{-1}x'y \rightarrow y = x\overset{*}{B} + u \rightarrow$ error term is a random var; x is fixed;
  - $B = (k \times 1)$: so y is random so $\hat{B}$ is Random
    - ↳ dif data, dif $\hat{B}$
- Variance of $\hat{B} = E[(\hat{B} - E[\hat{B}])(\hat{B} - E[\hat{B}])']$

$$\text{Var}\,\hat{B} = E\left[\begin{bmatrix}\hat{B}_1 - E[\hat{B}_1] \\ \hat{B}_2 - E[\hat{B}_2] \\ \vdots \\ \hat{B}_k - E[\hat{B}_k]\end{bmatrix}\left[\hat{B}_1 - E[\hat{B}_1],\ \hat{B}_2 - E[\hat{B}_2] \ldots \hat{B}_k - E[\hat{B}_k]\right]\right]$$ *symmetric

$$\text{Var}\,\hat{B} = E\begin{bmatrix}\text{Var}\,\hat{B}_1 & \text{cov}\,\hat{B}_1\hat{B}_2 & \rightarrow & \text{cov} \\ \text{cov}\,\hat{B}_2\hat{B}_1 & \text{Var}\,\hat{B}_2 & & \downarrow \\ \downarrow & & \searrow & \\ \text{cov} & \rightarrow & & \text{Var}\,B_k\end{bmatrix}$$

Really interested in variances:

$$\text{Var}\,\hat{B} = E[(\hat{B} - \underset{B}{E[\hat{B}]})(\hat{B} - \underset{B}{E[\hat{B}]})']$$

$\hat{B} = (x'x)^{-1}x'y$

$E[\hat{B}] = B$

$\hat{B} - B = (x'x)^{-1}x'u$

*WHERE DOES U COME From?

$\text{Var}\,\hat{B} = E[\overbrace{(x'x)^{-1}x'u}\ \overbrace{u'x(x'x)^{-1}}]$

$\text{Var}\,\hat{B} = (x'x)^{-1}x'\ \underbrace{E[uu']}\ x(x'x)^{-1}$

- ↳ var u → if don't know $\overset{\text{var}}{u}$ then stuck
- if violate 3/4 → have to plug in whole matrix
- if meet assumptions → var $u = \sigma^2 I$

Sept 6-8

Properties of OLS estimators

- if all OLS ~~est.~~ assump. hold → estimator is BLUE
  - Best, Linear, Unbias, Estimator:
    - Best in sense it is min variance
    - Linear unbias - in that is only best among all possible unbias est. → also linear functions of Y
- However → if v_i is Normally distributed → OLS estm. has min var. among all possible UNBIAS est.

TOPIC III: Measures of Goodness of fit

- Measures of goodness of fit
  - OLS method → Best fitting model as in it has smallest possible RSS → possibly BLUE
  - HOWEVER → How ~~can~~ well can this best fitting model predict the behavior of Y remains unanswered ↳ use R^2 to quantify
- The R^2
  - Based on notion that each Y observation (Y_i) can be decomposed into total, explained, and residual variation (wrt $\bar{Y}$)
  - Total variation: Square of the difference between observed Y (Y_i) and mean of Y ($\bar{Y}$) ↳ $[(Y_i - \bar{Y})^2]$
  - Explained variation: square difference between the predicted value of Y ($\hat{Y}_i$) and mean of $\bar{Y}$ ↳ $[(\hat{Y}_i - \bar{Y})^2]$
  - Residual variation: Square of difference between the observed Y (Y_i) and predicted value of y ($\hat{Y}_i$) ↳ $[(Y_i - \hat{Y}_i)^2]$
  - Specifically: TSS: ($TSS = \Sigma(Y_i - \bar{Y})^2$)
    - Measures total amt of variation in observed values and dependent variable
    - numerator of variance

    ESS: ($ESS = \Sigma(\hat{Y}_i - \bar{Y})^2$)
    - Measures how much of total variation is being explained by the estimated regression model

    RSS: ($RSS = \Sigma(Y_i - \hat{Y}_i)^2$) $= \Sigma(\hat{u}_i)^2$
    - Measures how much of total amt of variation is not explained by estimated model

- $TSS = ESS + RSS \Rightarrow (Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + (Y_i - \bar{Y})$

$$\Sigma(Y_i - \bar{Y})^2 = \Sigma[(\hat{Y}_i - \bar{Y})^2 + (Y_i - \hat{Y}_i)^2]$$

$$= \Sigma[(\hat{Y}_i - \bar{Y})^2 + (Y_i - \hat{Y}_i)^2 + 2(\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i)]$$

$$= \underbrace{\Sigma(Y_i - \bar{Y})^2}_{ESS} + \underbrace{\Sigma(Y_i - \hat{Y}_i)^2}_{RSS} + \underbrace{2(\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i)}_{\text{Has to equal } 0}$$

Sept 6-8 cont.

$0 = \Sigma(\hat{y}_i - \bar{y})(y_i - \hat{y}_i)$

$= \Sigma \hat{y}_i y_i - \Sigma \hat{y}_i^2 - \Sigma \bar{y} y_i + \Sigma \bar{y} \hat{y}_i$

$\Sigma \hat{y}_i y_i - \Sigma \hat{y}_i^2 - \bar{y}\Sigma y_i + \bar{y}\Sigma \hat{y}_i \rightarrow \bar{y}(\cancel{\Sigma \hat{y}_i - \Sigma y_i}) = -\bar{y}\Sigma(y_i - \bar{y}\Sigma \hat{y}_i)$ → 0

$= \cancel{\Sigma \hat{y}_i}(\hat{y}_i + \hat{u}_i) - \cancel{\Sigma y_i}$

$= \Sigma \hat{y}_i \hat{u}_i$

uncorrelated so = to 0

* Mathmatical properties have to hold
  → if $\hat{B}_1 = 0 \Rightarrow$ 1st normal equation doesn't hold

- R^2 calculated as:
  - $R^2 = \Sigma(\hat{y}_i - \bar{y})^2 / \Sigma(y_i - \bar{y})^2 = ESS/TSS$ → Proportion of total variation in dep var explained by model

    [OR]

  - $R^2 = 1 - \{\Sigma \hat{u}_i^2 / \Sigma(y_i - \bar{y})^2\} = 1 - RSS/TSS$ → 1 - the proportion that is ~~not~~ not explained by model
- SO: R^2 measures ~~variation~~ portion of total variation in y that is explained by Model
  - R^2 is Ratio → no units; so Δ units y is measured in ~~doesn't~~ DOES NOT Δ R^2
  - Takes only value b/w 0+1
    - $R^2 = 0$ → (RSS = TSS, ESS = 0) → No fit @ all
    - $R^2 = 1$ → (ESS = TSS); RSS = 0) → perfect fit
    - ↑ R^2 = better fit
    - in simple Regression → $R^2 = (r_{yx})^2$ [square of correl coef b/w y+x]
- Value of R^2 must be assessed in light of type of data being analyzed
  - time series data: $R^2 > .80$
  - cross sectional data: $R^2 > .5$
  - R^2 is only a ~~model~~ Measure of models capasity to predict y → can have low R^2 but Precise Parameters
  - ** R^2 is only valid if model includes intercept → $\Sigma y_i \neq \Sigma \hat{y}_i$ → $TSS \neq RSS + ESS$ ∴
- Disadvantage of R^2 → R^2 ↑ as indep var are added (even if var does not affect y)
  - when OLS ~~est~~ ~~est~~ coef → any added indep var = smaller RSS → ~~RSS~~ R^2↑
- An increase in R^2 as a result of adding indep var ≠ expanded model is better or that var. affects y

Adjusted R^2

- Denoted $\bar{R}^2$ → better to asses ~~the~~ if the addition of an indepvar likely to ↑ models ability to predict y

Sept 6-8 cont. #3

Adj R^2:

$\bar{R}^2 = 1 - \hat{var}(u) / \hat{var}(y) = 1 - [\underbrace{(RSS/TSS)}(n-1/n-k)]$

$\bar{R}^2 = 1 - \frac{RSS}{TSS}$ → sum $u_i / n-k$ [°of freedom] → $\hat{\sigma}^2$; TSS → sum $(y_i - \bar{y})^2 / n-1 = \hat{\sigma}_y^2$

Penalty for # of expl. var

These two terms have to override to Δ $\bar{R}^2$

- is less R^2 unless $k=1$ or $R^2=1$
  - No model
  - $k \geq 2$
  - $\frac{RSS}{TSS} = \emptyset$
  - $\frac{n-1}{n-k}$ wouldn't matter
- $\bar{R}^2$ does not have same straight forward interpretation as R^2; can be (-) [BAD]
- Some argue $\bar{R}^2$ provides better measure of goodness of fit when model is est w/ lots of variables + few observations

· Final Note on Model Specification

- Any var. believed to DIRECTLY and NOTABLY affect y, and does not hold const. value should be included in model
- excluding this var cause est. of remaining parameters to be bias + inconsistent → $E[U] \neq 0$
  - *The tradeoff → always favor putting in
- Potential conseq. of indvd irrelevant var in model is less serious
- if var affect y indirectly through another indep var; doesn't have to be included

Topic 4/5 → Normal Error model

- Normal error model:
  - in addition to normal error term assump:
    - $cov[U, x_j] = 0$
    - $E[U_i] = 0$ for all i
    - $Var[U_i] = \sigma^2$ for all i
    - $cov[u_i, u_j] = 0$ for all $i \neq j$

    Normality
  - error term is also Normally distributed → $U \sim N(0, \sigma^2)$
  - Supposedly supported by central limit theroum
    - central limit therom: sum of any large set r.v. [if iid] will be normal
      - if not iid have to be indep.
  - Recall Assum 1+2 → unbias + consist
  - in addition if U_i is iid; $var[U] = \sigma^2 I$ + $var[\hat{B}] = \sigma^2 (x'x)^{-1}$
  - if all assump hold + error term Normal distrib:
    - $\hat{B} \sim N(B, \sigma^2 (x'x)^{-1})$ thus.
    - $\hat{B}_j \sim N(B_j, var[\hat{B}_j])$ $j = 1 \ldots k$ where $var[\hat{B}_j] =$ j^{th} diagonal element of $\sigma^2 (x'x)^{-1}$

Sept 13 - 15

[Hypothesis testing: t tests]

*95% CI → 5% chance that the B is out of that Range

- $t = (\hat{B}_j - B_j) / \hat{se}(\hat{B}_j)$
- only follows t distribution if guess correct B_j
  - ↳ more B_j value departs from true B_j the more the t ratio differs from
- if compute t over several sets? Sets of t follow t distribution w/ n-k degrees of freedom

[
- Null Hypo: $H_0 \Rightarrow B_j = C$ (const value)
- Alt Hypo: $H_a \Rightarrow B_j \neq C$
]

- ascerne how likely is it that a particular t value is drawn from a $t_{(n-k)}$
  - ↳ pick α → usually .05 → prob of being wrong if reject H_0
    - ↳ we know nothing about fail to reject
  - ↳ Find t value for that level of signif [t^*]
  - ↳ if $|t| > t^*$ reject H_0; favor $H_a \to B_j \neq$ w/ α > Prob. of being mistaken
  - ↳ if $|t| \leq t^*$ cannot reject H_0 BUT probability that H_0 is correct is unknown

[Statistical Hypothesis testing]

- t test only holds if 4 OLS assumptions hold AND error term is normally distrib
- HOWEVER if sample size is large (100+) t test can be used even if U not Normal
  - ↳ w/ large sample → convergence to t distrib
- Most common use of t stat: test if parameter is statistically different from $\varnothing$ → implies x will affect y
  - $H_0: B_j = 0$ (x has no effect)
  - $H_a: B_j \neq 0$ (x has affect)
  - two tail → pos or neg: allows for pos or neg effect only

Pay attention to t and situation

- Rule of thumb: Ratio $\hat{B}/\hat{se}[\hat{B}] > |2|$ tend to reject H_0
- α of .1, .05, .01 usually → mean 3 levels of statistical certainty when rejecting H_0: 90, 95, 99 respectively
- P-value: exact/lowest α for which H_0 can be rejected in favor of TWO TAILED HYPOTHESIS → $H_a: B_j \neq 0$
- one tail tests usually have knowledge that affect isnt in one of the tails → an x cant have neg effect or something

·1 tail makes easier to reject null

- Assume all assumptions hold; if take a var out have to recalc; assume all units are similar
- if opt for one tail dont have to ~~recalc~~ Δ process t value → critical table value Δ b/c adj for idea that one of the tails is left out
- if same critical value is used $\alpha \to \alpha/2$

Sept 13-15

[Statistical Hypothesis testing]

- when reporting results → report se or t value
- also include test of each model parameter
  ideally include pvalue → state 1 or 2 tail

$\frac{(SE)}{(S^2)}$ * based on signif level

[Confidence interval for y_0]

- x_0 = ~~xx~~ $1 \times k$ vector of particular xs
- y_0 = value taken by y when x_0 y_i → observed
- $\hat{y}_0 = x_0\hat{B}$ predicted value given x_0
- $\hat{U}_0 = y_0 - \hat{y}_0$; predicted error @ x_0
- can be shown that $E[\hat{U}_0] = 0$
- can be shown that $Var[\hat{U}_0] = \sigma^2 + \sigma^2 x_0 (x'x)^{-1} x_0'$

* · + if U normal distrib then $\hat{U}_0$ is usually normal

- thus $\hat{U}_0 / se[\hat{U}_0] \sim N(0,1)$ + $\hat{U}_0 / \hat{se}[\hat{U}_0] \sim t_{n-k}$ $E[\hat{U}_0] = 0$
- confidence interval for y_0 → $\hat{y}_0 \pm t_{(n-k)} \hat{se}[\hat{U}_0]$
- ci for $E[y_0] = x_0 B$?
  also Normal so w/ $\hat{se}[\hat{y}_0]$ → $t_{(n-k)}$
  - with $E[y_0]$ + $\hat{se}[\hat{y}_0]$ makes ci narrower

$E[y_0] = \hat{y}_0 \pm \hat{se}[y_0] t_{(\alpha/2, n-k)}$

$E[\hat{U}_0] = 0 \rightarrow E[y_0 - \hat{y}_0] = E[y_0] - E[\hat{y}_0] =$
$E[x_0 B] - E[\hat{y}_0] = 0$
$E[\hat{U}_0] = 0$
$E[x_0\hat{B}]$

$Var[\hat{U}_0] = var[y_0 - \hat{y}_0]$
$= var[y_0] + var[\hat{y}_0]$
$= var[x_0 B + U_0] + var[x_0\hat{B}]$
$= \underbrace{var[x_0 B]}_{0} + \underbrace{var[U_0]}_{\sigma^2} + x_0 var[\hat{B}] x_0'$
$= \sigma^2 + x_0 \sigma (x'x)^{-1} x_0'$

$\hat{U}_0 = y_0 - \hat{y}_0 = x_0 B + U_0 - x_0\hat{B}$
$= x_0 B + U_0 - x_0 (x'x)^{-1} x'y$
$= x_0 B + U_0 - x_0 (x'x)^{-1} x' (xB + U)$
$= x_0 B + U_0 - x_0 (x'x)^{-1} x'x B - x_0 (x'x)^{-1} x' U$
$= x_0 B - x_0 B - x_0 (x'x)^{-1} x' U$
$\hat{U}_0 = U_0 - x_0 (x'x)^{-1} x' U$

Quiz Sept 20 - Oct 3

Regression through origin:

- usually used when theory dictates $Y = 0$ if all $x = 0$
  - ↳ ie price comodity
- estimated using OLS by excluding 1st column of 1s
- will Δ slope parameter estimates
- R^2 not useable → no longer bounded @ 0 → $TSS \neq ESS + RSS$
  - ↳ can use "raw" R^2 [corr($Y\hat{Y}$)] but interp not the same
- only use through 0 if there is STRONG evidence

Scaling + units of Measure:

- Δ in scaling of any explanitory var by a fixed multiplicitive factor of W_x only Δ value of [corresponding parameter] + [se by a factor] of $\left[\frac{1}{W_x}\right]$
- Δ in scaling of dependent var by fixed multiplicitive factor W_y changes value of all Model Parameters + correspond se by factor of $[W_y]$

Standard Regression:

- regression w/ all var (Y+X) being standardized to mean 0 and variance 1
- each var: subtract μ + divide by σ

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y}$$

$$B_2^* = \frac{sd_x}{sd_y} \cdot \hat{B}_2$$

- Parameter est in this Regression are "Beta coef"
- B/c all var have mean of 0; intercept is 0
- interpretation of β coef follows from both Y + X are now measured in stand. dev.
  - ↳ ie they measure the sd Δ in Y when x Δ by 1 sd
    - no unit of measure
    - used for relitive importance of xs but doesn't account for rate of Δs in x
- a larger β means the x contributes to explaining var in Y more

$$Y_i^* = \frac{Y_i}{S_Y} - \frac{\bar{Y}}{S_Y}$$

$$W_x = \frac{1}{S_Y}$$

$$\frac{1}{W_x} = S_x$$

Lagged variables

- Sometimes the value of Y → depends on value of x from previous time period.
- Can use OLS ⇒ must rearrange data so that Y_t correspond w/ x @ t-1
- lose one var observation

Quiz Oct 3

First difference Variable

- The Δ in value from one period to the next

$$Y_t = B_1 + B_2(x_t - x_{t-1}) + U_t \text{ or } Y_t = B_1 + B_2 \Delta X_t - U_t$$

- What affects y? value of x or Δ in x
- lose one observation
- used to forcast
- will Δ Bs so do this before OLS

Alt Model specifications:

- So far only looked at linear models
- ARE x + y linearly related? → OLS: linear in parameters but not in explan/dep var

LOG LINEAR:

- AKA: LOG LOG or double log
- can accomidate: ↑@ decr rate; ↑@ incr rate + ↓@ decr rate

$$\hat{Y}_i = e^{\hat{B}_1} x_2^{\hat{B}_2} x_3^{\hat{B}_3} \ldots x_k^{\hat{B}_k}$$

$$E_{yx_2} = \frac{\partial \hat{y}_i}{dx_{i2}} \times \frac{x_{i2}}{\hat{y}_i} = \frac{\hat{B}_2 e^{\hat{B}_1} x_{i2}^{\hat{B}_2 - 1} x_{i3}^{\hat{B}_3} x_{i2}}{e^{\hat{B}_1} x_{i2}^{\hat{B}_2} x_{i3}^{\hat{B}_3}} = \hat{B}_2$$

- $\hat{B}_j < 0$; as x_j ↑; y ↓ @ decr rate
- $\hat{B}_j = 1$; as x_j ↑; y ↑ @ const rate
- $0 < \hat{B}_j < 1$; as x_j ↑; Y ↑ @ decr rate
- $\hat{B}_j > 1$; as x_j ↑; Y ↑ @ incr rate

$$\ln Y = \ln e^{\hat{B}_1} \ln x_2^{\hat{B}_2} \ln x_3^{\hat{B}_3} \ldots$$

$$\ln Y = \hat{B}_1 + \hat{B}_2 \ln x + \hat{B}_3 \ln x_3 \ldots + u$$

- So use normal OLS but all apply to $\ln x_i$ + y_i
- disadvantage: all x_j–y relations assumed non lin
- all x + y val must be pos → ln(−) not dfn → can have 0 but not recommended
- $\hat{B}_j$ directly measures elast of Y wrt x_j → % Δ in Y when x_j Δ 1%
- slope varies but elasticity is const
- since ln y → R^2 not the same

$$corr[x_1, x_2] = \frac{\sigma_{12}^2}{\sigma_1 \sigma_2} = \rho(12)$$

- use rho R^2
- Better goodness of fit → square of the correlation coef b/w observed + predicted val of Y
- Note: ~~est~~ predictions from org. equation not est model

Semi log: Log Lin

- log of dep. var but explan in original form
- B measure rel Δ in y when x Δ by 1 unit
  - ↳ if ·100 → % Δ Y for 1 unit Δx
- as x ↑ Y ↑ @ incr rate ($B_j > 0$) or ↓ @ decr rate ($B_j < 0$)
- Y can't be neg

Quiz oct 8

Semi log: Lin log

- Log of explan var but not dep. var
- B_j measure unit ΔY for ~~% ΔX~~ rel ΔX
  - ↳ ÷100? → unit ΔY for $X \Delta 1\%$
- as X ↑ Y ↑ @ decr rate ($B_j > 0$) decr @ decr rate ($B_j < 0$)
- Y can be neg but not X

Reciprocal Specification

$$Y_i = \hat{B}_1 + \hat{B}_2\left(\frac{1}{X_2}\right) + \dots$$

- $\hat{B}$ pos → Y ↓ @ decr rate (Has asyntope)
  - as ~~y~~ ↑ approch intercept
- if $\hat{B}$ neg → Y ~~decr~~ ↑ @ decr rate; intercept is max val
- Slope op. sign of $\hat{B}$
- slope: $\frac{d\hat{Y}_i}{dX_j} = -\hat{B}_j\left(\frac{1}{X_{ij}}\right)^2$
  - dif slope dep on value of x
- can have reciprocal + original in same model

Note on OLS Estimation

- all (log lin; lin log, log log, + reciprocal) can be est w/ normal ols
- Make sure all logs + reciprocals taken + used before est ols

Log Reciprocal:

- $\ln y$ + $\frac{1}{x}$
- get s shape
- ↑@ incr then inc @ decr
- (−) B ↗
- (+) B: decr @ decr rate

Polynomial Specification:

- $\hat{Y}_i = \hat{B}_1 + \hat{B}_{21}X_{12} + \hat{B}_{22}X_{12}^2 + \hat{B}_3 \dots \hat{B}_K X_{1K}$
- incr @ decr rate
  or incr @ incr rate
- decr @ incr rate or decr @ decr rate
- very flexable
- Might enentualy employ reversal of relationship ∴
- can combine linear + non linear
- used to test non linearity of x/y
- if parameter of x_j^2 is not stat dif. than 0; H_0 = linear cannot be rejected
- can have multi colinearity problem

Quiz Oct 3

Choice of functional form:

- usually use theory or knowledge of subject
- Also consider:
  - signs of parameters is according to theory
  - level of ~~signif~~ stat signif of param. in other models
  - R^2 or $\bar{R}^2$ as long as dep var same across model choices

- Residual analysis:
  - compute $\hat{e}_i = \alpha_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_2^2 + \alpha_5 x_3^2 + \underbrace{\alpha\, x_2 x_3}_{\text{interaction term}} \dots + e_i$
  - don't want param to be statist signif → if yes? missed explan var
  - want □s to pick up pattern; need @ least those
  - $H_0: \alpha_2 = \alpha_3 = \alpha_4 \dots = \alpha_k = 0$ } jointly = to 0
    H_A: @ least one not 0

F test:

- Recall: TSS: $\Sigma(Y_i - \bar{Y})^2 \rightarrow n-1$ df
  ESS: $\Sigma(\hat{Y}_i - \bar{Y})^2 \rightarrow$ ~~k-1 df~~ $k-1$ df
  RSS: $\Sigma(Y_i - \hat{Y}_i)^2 \rightarrow n-k$ df
  - Bigger RSS then less more var is unexplained

$(n-1)(k-1)(n+k) \rightarrow n-1$

- TSS: $n y_i - 1\bar{y} \rightarrow n-1$
- ESS: $k\hat{y}_i - 1\bar{y} \rightarrow k-1$
- RSS: $n y_i - k\hat{y}_i = n-k$
  → # B to calc

df → # of indiv RV to calc

Recall

- F test: used to eval H_0 that all model param (w/ exception of intercept) are 0
  - for ~~regre~~ residual analy: want to <u>Not</u> Reject
  - Means none of the var shown have affect on y
  - if H_0 not reject Model as a whole cannot be said to be useful to explain y
  - H_A: AT LEAST <u>ONE</u> x not equal to 0

$F^* = (ESS/k-1) / (RSS/n-k)$

- Under H_0: w/ 4 OLS ~~pam~~ assump + u's normal; F^* follows F distrib w/ $k-1$ & $n-k$ df
- compare F^* stat calc to critical F table

F distrib

$F^* > F$ reject H_0

Oct 3 Quiz

Other types of F test:

- joint test on several coef → subset
- test involve linear function of the regression coef
- test involve equality of coef of dif regression equations

Joint test

- eval subgroup of x
  - ↳ like category
- unrestricted Model: $Y = B_1 + B_2X_2 + B_3X_3 + \ldots + B_kX_k + u$
- want to test if subset of q of regres. coef jointly = to 0
- Rewrite:

$$Y = B_1 + B_2X_2 + \ldots + B_{k-q}X_{k-q} \vdots B_{k-q+1}X_{k-q+1} + \ldots + B_kX_k + u$$

· last q coef all = 0 so:

$$Y = B_1 + B_2X_2 + \ldots + B_{k-q}X_{k-q} + u$$

- H_0: $B_{k-q+1} = B_{k-q+2} = \ldots = B_k = 0$
- H_A: @ least ONE not 0
- Need to est Rest. + unRest model
- $RSS_r \geq RSS_{UR}$
- if H_0 is correct; drop q var have little effect on explan model
  - ↳ RSS_r slight higher than RSS_{ur}

$$F^* = \frac{\frac{RSS_r - RSS_{ur}}{q}}{\frac{RSS_{ur}}{n-k}}$$

- Not equivalent to indiv t test
- B/c multicolinearity possible all t test are insig but joint F shows rejection H_0
  - b/c F is impervious to multicolin

12: Oct4 - Oct6

Test of Linear functions of Regression coef

- Recall in cobb douglas production function sum of exponential coef indicates Returns to scale of production → ex: $Y_i = e^{\hat{\beta}_1} X_{i2}^{\hat{\beta}_2} X_{i3}^{\hat{\beta}_3}$

  $B_2 + B_3 = 1 = CRTS$

- Production economists often want to test if CRTS so:

  $H_0: B_2 + B_3 = 1$ OR $B_2 = 1 - B_3$

  get: $\ln y = \hat{B}_1 + (1 - \hat{B}_3) \ln x_2 + \hat{B}_3 \ln x_3$

  OR

  $-\ln x_{i2} + \ln y = \hat{B}_1 + \hat{B}_3 (\ln x_{i3} - \ln x_{i2})$

- another example is $B_2 - B_3 = 0$ OR $B_3 = B_2$

  UR: $y = B_1 + B_2 x + B_3 x \ldots$

  R: $y = B_1 + B_2 (x_2 + \ldots$

  OR .. $B_2 (x_2 - x_3)$ which is $B_3 = -B_2$ ✱ look @ signs

- Fstat is same as before:

  $F^* = (RSS_R - RSS_{UR} / q) / (RSS_{UR} / (n-k))$

  q = # restrictions

- can also have linear functions of more than 2 coef but each Function = restriction not each parameter

F test for equality of coef of dif. Regression eq:

- want to test if same model work for 2 data sets or if need new model
- "chow" test
- can have dif # observations; closely related dep. var + same indep var (but can take dif var)
- Ho: two models identical
- Ha: @ least 1 par of coef not the same
- $RSS_{UR} = RSS_1 + RSS_2$
- df: $n_1 + n_2$
- RSS_R = merged model
- $F^* = ((RSS_R - RSS_{UR}) / k) / (RSS_{UR} / (n+m-2k))$
- $df(RSS_R) - df(RSS_{UR}) = (n+m-k) - ((n-k) + (m-k)) = n+m-k-n-m+2k = k$ = # rest imposed
- if Ho rejected data cant be pooled
- dummy vars are soln

Quiz oct 4-6

Use of dummy var:

- Some indep var are qualitative → values w/ no natural/logical ordering
- modeled w/ dummy var
- set of dv created for each categor var → # dv = # categories
  Make sure to exclude one category → avoid multicolinearity
- B_1 = intercept for excluded cat; param = dif b/w B_1 categ + that categor
- use t test to see if B_i is statistically dif
- to see if cotton yeilds statist. dif. under varieties 1+2; est 2nd version of model w/ dif exclud cat.
- w/ just intercept var → parallel shift "intercept shifters"
- slope shifts: $B_3 D_{3i} x_{4i}$ → $B_4 + B_5$ is slope x_4 when ver't 2 plant

Multi colinearity

- when 2 or more indep var in a regression model are highly correlated to each other it is difficult to determine if var have indep effect + the magnitude of said effect
- Math: occurs b/c, everything cp, se assoc w/ given ols param est will be higher if corresp IV is more highly corel w/ other IV
- $\hat{V}[\hat{B}_2] = \hat{\sigma}^2 / [(1-r_{23}^2)\Sigma(x_{2i}-\bar{x}_2)^2]$
  ↳ Perfect cor = 1 have to divide by 0 ... Not possible
  $(x'x)^{-1}$ doesn't work
- higher correl = ↑var → var double if .5; se effect = $\sqrt{\text{Veffect}}$
- could be reason why key IV are not statist. signif w/ t-test
- Not a mistake or violation of OLS → just undesirable
- more common in time series data

Perfect Multi colinearity:

- occurs when perfect linear correl between two IV
  ↳ dummy var
  ↳ IV = const value
- if present: OLS cannot produce param. or stand est: THIS IS A MISTAKE
  ↳ Redundant info?

Severe multi colin:

- multicol that correl is high + interfers w/ est. of parameters @ desired Statist. level

Quiz: Oct 4 - Oct 6

Symptoms of multicol:

- Key indep var not statistic signif according to t test
- High R^2, high signif F test; few/no stat signif t test
- Param est Δ [BIG] when exclud some indep var

Detecting Multicolin:

- conduct "artificial" regression b/w each indep var + remaining indep var
- Variance Inflation Factors (VIF)

$VIF_j = 1/(1-R_j^2) \rightarrow R_j^2$ is artific regres R^2 w/ j i.e.

→ calculates inflation on $V[\hat{B}_j]$ caused my multicol.; if none VIF=1

- VIF=2 = $V[\hat{B}_j]$ twice if no multicol. · what % ΔX_2 is already accounted for in others
- VIF>10 is clear evidence → 3-4 is expected
- also condition # of X matrix
  - ↳ if IV are incr correl. ratio of highest + lowest eigen value of X'X gets larger
  - ↳ condition # is square root of this
  - if >100 = problem
  - good: gives overall read
  - Bad: not immune to scale Δ
- others: Theil's multi colin. effect; determinant of correl matrix of IV (1=none; 0=perfect

Addressing multicol:

- be aware of presence → help justify why you left something in w/ "not signif" t
- new sample/more observ will ↓ multicol
- exclude IV that cause problem (but might be of interest... effy)

***- even if not the cause: recall omitting rel. indepvar = OLS est BIASED

- modify model: · use real not nominal
  - · reciprocal not polynomial
  - · use non linear models

· will be some multicol: just not too much "as long as statistical inferences not dissapointing"

Quiz oct. 11→13 [Heteroskedasticity]

Heteroskedasticity:

- Occurs when error term (~~therefore~~ y) don't have const variance across observations
- OLS parameter est still unbias but OLS se. are bias → $var\ v \neq \sigma^2 I$
  - ↳ can't do $var\ B = \sigma^2 (x'x)^{-1}$
  - ↳ get $var\ B = (x'x)^{-1} x' E[UU'] x (x'x)^{-1}$
- means any test that uses SE ~~or ci~~ or ci is wrong on avg.
  - ↳ will skew t ~~test~~
- OLS parameter est no longer most efficient (min variance even if error is normal)
  - ↳ Gauss-Marcov does not apply → another linear est. that is better "more efficient"

Test for Heteroskedasticity:

- [White test] is most common
  - Process: Regress the squared OLS residuals as the depend var. vs. explanitory var. their squares and cross products as explan var
  - * this Aux. Regression must include an intercept + exclude any redund Rt hand side variables
  - Under Ho: Homoskedascity (good): $n \cdot R^2$ of Aux Regression is distrib as a χ^2 w/ P degrees freedom → P = # parameters in aux regression − 1.
    - ↳ excludes intercept
  - want a lower # [Fail to reject]
  - careful w/ dummy var → dont need the square
- ** Key point: dont want to fail to reject incorrectly (Think you're good but you're not)
  - ↳ what kind of α? → high α; will ↓ critical value
- Disadvantage: can be other reasons we reject null w/ white test
- if misspesification of random component; incorrect functional form; correl b/w indep var + u → will trigger rejection of Ho
  - Assump H2
- only use white test if confident all OLS Assump hold ***
- [Breusch-Pagan]: older legrange multiplier test
  - specific Ha: $\sigma_t^2 = h(z, \gamma)$ where h is any possible nonlinear function of z_t, γ and z_t is (1×s) vector of suspect var + intercept; γ is vector of parameters
    - → can have weird combo functional form that is not accounted for here
  - Since the intercept (γ_1) is included in z_t the null hypothesis of homosked. is equiv to $H_0: \gamma^* = 0$ where γ^* includes all other param in γ
    - all except for intercept are 0

Quiz → oct 11→13 Pg 2

Tests for heteroskedacity:

[Breusch-Payan test]

- $Z_i = [1 \quad z_2 \quad z_3 \quad z_5]$ ← subset of suspects $S=4$
- $\gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix}$ new parameters (x var on U)
- $\gamma_1 + \gamma_2 z_2 + \gamma_3 z_3 + \gamma_4 z_5$ linear function of suspect variables

- if error term is normally distributed; under Ho; $BPTS = ESS_{AR}/2$ follows $\chi^2_{(S-1)}$ where ESS_{AR} is Expl. sum of squares of regression w/ $\hat{u}_i^2/\tilde{\sigma}^2$ as dep var and suspect variables in Z_i are explan var
  - $\tilde{\sigma}^2 = \sum \hat{u}_i^2/n$
  - * RSS/n * → $\hat{\sigma}^2 = \frac{RSS}{n-k}$
- compare computed statistic w/ table value of $\chi^2_{(S-1)}$ at desired α
  - ** error has to be Normal or χ^2 is not exact *
- Ho = Homosked.

- Advantage: ~~Most~~ more targeted to heterosked (not as easily tricked)
- disadvantage: Ha more restricted → requires $u \sim$ Normal

Dealing w/ heterosked:

- if single suspect variable → can resolve w/ weighted least squares estimator
  - Before OLS all data multiplied by set of weights; where the weights are a set of n values that when multiplied by error term make it have a constant variance $var[v_i] = z_i \sigma^2$ $w_i = \frac{1}{\sqrt{z_i}}$
  - ex: $\frac{1}{\sqrt{z_i}} Y = B_1 \frac{1}{\sqrt{z_i}} + B_2 \frac{x_i}{\sqrt{z_i}} + \frac{v_i}{\sqrt{z_i}}$
- if only one variable try to use dif. functions of it? (ie. $w_i = z_i^2; z_i^{1/2}; \frac{1}{z_i}$ etc)

- must retest residuals of WLS to see if problem is fixed
  - → use white test to see if model is better
  - → R^2 not valid: y is transformed but B are still comparable
- More than one var: use predictions from Aux regression to ~~be~~ obtain est for ## error term variances for each of the obs → if some are (-) just add lowest to = 0 to all
  - → since dep var of AR is $\hat{u}_i^2$; Predictions are proportional est for σ_i^2
  - → predictions might need to be transformed in order to obtain appropriate wts to correct Heterosked.

Av cross has const $\hat{u}_i$

wts are parameters

Dealing w/ Heterosked

- iid: $var(v) = \begin{bmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ 0 & & \sigma^2 \end{bmatrix} = \sigma^2 I$

Not iid:

$var(v) = \Omega = \begin{bmatrix} var\,u_1 & & cov. \\ & var\,u_2 \ddots & \\ cov. & & var\,u_n \end{bmatrix} = \sigma^2 \begin{bmatrix} d_1 & & cov \\ & d_2 & \\ cor & d_3 & d_n \end{bmatrix}$

Heterosked.

$d \neq d$ → heterosked

* factor out const. → "variance"
→ diagonal is not the same or 1
"correl. matrix"

Generilized least squares:

- general method to est regression model w/ error term not iid
  ↳ $var[v] = \sigma^2 \psi \neq \sigma^2 I_n$
  corr/cov matrix

- if: ψ is positive definite symmetric matrix there is another matrix P such that $P\psi P' = I_n$

- therefore if $var[v] = \sigma^2 \psi$, $var[Pv] = P\,var[v]\,P' = \sigma^2 P\psi P' = \sigma^2 I$
- Also $\psi = P^{-1}(P')^{-1}$ → $\psi^{-1} = P'P$

- $P\psi P' = I_n$
- $P^{-1} P \psi P'(P')^{-1} \Rightarrow P^{-1} I_n (P')^{-1}$
- $\psi = P^{-1}(P')^{-1}$
- $\psi^{-1} = P'P$

know

P gives wted matrix
↳ transformation matrix

$P\psi P' = I_n$
$\psi = P^{-1}(P')^{-1}$
$\psi^{-1} = P'P$
$P^{-1} I_n (P')^{-1} = P^{-1} P \psi P'(P')^{-1}$

- End up w: $PY = PX\beta + PV$ or $Y^* = X^*\beta + U^*$ → U^* is iid
- if ols used on transform model → do get efficient pram est & unbias se est.
- ols to this model: GLS estimator
  $\tilde{\beta} = (x'P'Px)^{-1} x'P'Py = \cancel{(x'\psi x)} \; (x'\psi^{-1}x)^{-1} x' \psi^{-1} y$ → I_n
  $V[\tilde{\beta}] = \sigma^2 (x'P'Px)^{-1} = \sigma(x'\psi^{-1}x)^{-1}$

* - WLS is an application of GLS

$P = \begin{bmatrix} \frac{1}{\sqrt{z_1}} & & \\ & \frac{1}{\sqrt{z_2}} & \\ & & \frac{1}{\sqrt{z_n}} \end{bmatrix} \rightarrow P'P = \begin{bmatrix} \frac{1}{z_1} & & \\ & \frac{1}{z_2} & \\ & & \frac{1}{z_n} \end{bmatrix} \leftarrow \text{inverse} \rightarrow \begin{bmatrix} z_1 & & \\ & z_2 & \\ & & z_n \end{bmatrix} \cdot \sigma^2 = \Omega = \sigma^2 \psi$

why: $\psi^{-1} = P'P$ + $(P'P)^{-1} = \psi$

Dealing w/ Heterosked

- if inefficiency of OLS param. est. not a concern; simple way of dealing w/ Heterosked
- is to est Heterosked-consist se:

  $HCSE = (x'x)^{-1} x' \Omega x (x'x)^{-1}$; Ω = matrix of ~~squared~~ square ols resid on diagonal & 0s elsewhere

- yeilds consistant estimates for se; can be considered correct & confidently used for hypo testing & building ci if sample size is large (>100)

Topic 4 → Normal Error Model

- Normality results from OLS estimator being a linear function of V + U is Normaly distributed
  - only holds asymptotically (large sample) if xs are RANDOM (in small samples $\hat{B}$ is aprox. normal)

$$\left.\begin{array}{l} \hat{B} = \underbrace{(x'x)^{-1}x'}_{M} y \\ \hat{B} = B - \underbrace{(x'x)^{-1}x'}_{M} U \end{array}\right\} \text{linear function of } U + Y + \text{linear function of Normal var are normal}$$

- Means: if multipl samples taken + OLS obtains multipul est $(\hat{B}_j)$ then one can think $\hat{B}_j$ comes from Normal distrib w/ mean B_j + var → $Var[\hat{B}_j]$
  - $1\ se[\hat{B}_j]$ → 68% prob
  - $2\ SE[\hat{B}_j]$ → 95% prob
  - $3\ se[\hat{B}_j]$ → .99.8% prob

Confidence interval

- Since $\hat{B}_j \sim N(B_j, var[\hat{B}_j])$
  - → $Z = (\hat{B}_j - B_j) / se[\hat{B}_j] \sim N(0,1)$ where $se = \sqrt{Var[\hat{B}_j]}$
- Recall: σ^2 for $SE[\hat{B}_j]$ is unknown → est by $\hat{\sigma}^2$ → means one only has to est for $se[\hat{B}_j]$ → $\hat{se}[\hat{B}_j]$
- shown that: $t = (\hat{B}_j - B_j) / \hat{se}[\hat{B}_j] \sim t_{(n-k)}$
  - t distrib w/ $n-k$ degree of freedom
    k → # parameters.
- from previous: $Pr(-t_{(\alpha/2, n-k)} < t < t_{(\alpha/2, n-k)}) = 1-\alpha$

$$t \to (\hat{B}_j - B_j) / \hat{se}[\hat{B}_j]$$

- Multiply all elements by $\hat{se}[\hat{B}_j]$, subtract $\hat{B}_j$ + multiply by -1
  - ↳ $Pr(\hat{B}_j - t_{(\alpha/2, n-k)}(\hat{se}[\hat{B}_j]) < B_j < \hat{B}_j + t_{(\alpha/2, n-k)}(\hat{se}[\hat{B}_j])) = 1-\alpha$

Test one:

[Topic one: The basics]

Intro: Regression analysis used when interested in how a particular var is affected by other variables

[What is a Regression Model]

- start by conceptualizing a behavioral relation based on economic theories/knowlage or just reason
  - ↳ includes a depent var (y) + 1 or more indep var (x)
  - ↳ makes mathmatical model:

$$Y = B_1 + B_2 x_2 + B_3 x_3 + B_4 x_4 + U$$

  - ↳ B are const coef → true parameters; estimated using data on y+x
  - ↳ U → random error term recognizes relaton b/n x+y is nt exact

- U: account for other var on y → unemp explan var
  - → Random error in y Measurement
  - → pure chance

* Models seek to capture essentials; not to be a perfect representation of the process

Main uses → est magnitude of explanatory var on Y [est. parameters]
→ obtain predictions for dep. var

Population v. est:

- Popolaton B unknown
- X (observ) (which var)
- Assume populaton model generates Y_i for observe x_i

[Topic 2: OLS Estimation]

- OLS: ordinary least squares
- method: minimize $RSS = \min \sum_{i=1} \hat{u}_i^2$

$$= \min \sum (Y_i - \hat{Y}_i)^2$$
$$= \min (Y_i - \hat{B}_1 - \hat{B}_2 x_i)^2$$

- Min RSS; squared to penilize large U
- take derivitives to find B values

$$\frac{dRSS}{\partial \hat{B}_1} = \partial\left(\sum (Y_i - \hat{B}_1 - \hat{B}_2 x_2)^2\right) / \partial \hat{B}_1 = 0$$
$$= \sum (Y_i - \hat{B}_1 (n) - \hat{B}_2 (\sum x_i)) = 0$$

- end up w/ normal equations:
  - $n\hat{B}_1 + (\sum x_i)\hat{B}_2 = \sum Y_i \rightarrow \hat{B}_1 = \bar{Y} - \hat{B}_2 \bar{X}$
  - $(\sum x_i)\hat{B}_1 + (\sum x_i^2)\hat{B}_2 = \sum x_i Y_i \rightarrow \hat{B}_2 = \dfrac{\sum (x_i - \bar{x}) Y_i}{\sum (x_i - \bar{x})^2}$

[test one]

[topic 2] → OLS

Mathmatical properties of OLS:

- $\sum \hat{u}_i = 0 \rightarrow \sum(\hat{u}_i) = \sum(y_i - \hat{y}_i) = \sum(y_i) - \sum(\hat{y}_i) = 0$
- $\sum \hat{y}_i = \sum y_i \rightarrow -2\sum_{i=1}^{n}(y_i - \hat{B}_1 - \hat{B}_2 x_i) = 0 = \sum y_i - \sum \underbrace{(\hat{B}_1 - \hat{B}_2 x_i)}_{\hat{y}_i} = 0 = \sum(y_i) = \sum(\hat{y}_i)$
  - always passes through pt $(\bar{x}, \bar{y})$ → we est model through mean data
  - residuals are not correlated w/ values of xs or $\hat{y}_i$

[OLS Requirements]

- OLS is only used to est models that are linear in PARAMETERS
- # of observations > # parameters [n×4 is rule of thumb]
- Some level of variation in x
- can't be perfect multi colinearity → if is $(x'x)^{-1}$ doesn't exist → can't est $\hat{B}$

[OLS Assumptions]

- values taken by x are fixed or random + uncorrel w/ error term
  - can be corr if: proxy measured w/ error
    dep var has feedback

1 + 2 Hold?
B is unbias + consistient

- $E[U] = 0$
  - No key rel x left out
  - Functional form correctly specified (linear)

$\hat{B} = (x'x)^{-1}x'y$

$= (x'x)^{-1}x'(xB + U)$

$= \underbrace{(x'x)^{-1}x'x}_{I}B + (x'x)^{-1}x'U$

$= B + (x'x)^{-1}x'U$

$E[\hat{B}] = B + (x'x)^{-1}x'\cancel{E[U]}^{0}$

$E[\hat{B}] = B$ ✓ [unbias]

[consistent]

$\lim_{n\to\infty} \text{Prob}(|B - \hat{B}| > \delta) = 0$

- No auto correlation → u is indep distributed
  ↳ error terms are uncorr w/ eachother
- No Heteroskedasticity → $\text{var}[u_i] = \sigma^2$
  - error term variances are same regardless of xs

[Covariance Matrix]

- $\text{var}[U] = E[(U - E[U])(U - E[U])']$
- Matrix gives n×n matrix w/ var on diagonal + covar on offdiag.
- No auto correlation: all covar = 0 — independently
- No Heteroskedacity: all var are same — + identically distributed

} iid

Test one

[Topic 2: OLS]

OLS standard error:

- $\hat{B}$ is random var b/c is is a function of v & v is random var

covariance Matrix of $\hat{B}$ → symmetric

- $Var[\hat{B}] = E[(\hat{B}-E[\hat{B}])(\hat{B}-E[\hat{B}])']$

$= E[((x'x)x'v)((x'x)^{-1}x'v)']$

$= E[(x'x)^{-1}x'vv'x(x'x)^{-1}]$

$= \cancel{x'x^{-1}} (x'x)^{-1}x'E[vv'](x'x)^{-1}x'$ * [Stop unless v is iid]

$= (x'x)^{-1}x'\sigma^2 I(x'x)^{-1}x'$

$= \sigma^2(x'x)^{-1}\underbrace{xx'(x'x)^{-1}}_{I}$

$= \sigma^2(x'x)^{-1}$

to calc σ^2

$\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n}(\hat{v}_i - \bar{\hat{v}}_i)^2}{n-k}$ → 0

$= \dfrac{\sum_{i=1}^{n}(\hat{v}_i)^2}{n-k}$

$= \dfrac{RSS}{n-k}$

$= \dfrac{\hat{v}'\hat{v}}{n-k}$

se: $\sqrt{\hat{var}(\hat{B}_i)}$

- Provide a degree of percision measure
- Rule of thumb → true parameter w/in ± 2 se of ols est
- Note on SE:

$\hat{V}[\hat{B}_2] = \hat{\sigma}^2 / [(1-r_{23}^2)\sum(x_{2i}-\bar{x}_2)^2]$

→ □ of correl b/w x_2 & x_3

→ sum of □ deviation of x_2 & $\bar{x}_2$

- $\hat{\sigma}^2 \uparrow$; se↑
- $r_{23}\uparrow$; ↑se
- $\sum(x_{ij}-\bar{x}_j)^2 \uparrow$; se↓

[Properties of OLS estimators]

- unbiased; on average est expected to equal values of unknown pop. model
  - Requires Asump 1 & 2
  - if 1-4 hold estimator has min. variance among all possible UNBIAS estimators that are linear
  - OLS most likely yeild of est for B
  - if v is Normal → ols is min var of all ~~unbiast~~ unbias estimators
- BLUE: Best linear unbias estimator

Test One

[Topic 3] → goodness of fit

- OLS doesn't measure how well can this best fitting model predict y behavior
- use R^2

R^2:

- each y_i can be decomposed into explained + residual variation
    - Total: $(y_i - \bar{y})^2$
    - explained: $(\hat{y}_i - \bar{y})^2$
    - residual: $(y_i - \hat{y}_i)^2$

Σ in front = TSS, RSS, ESS

TSS = ESS + RSS

$$\Sigma (y_i - \bar{y})^2 = \Sigma [(\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)]^2$$

$$= \Sigma (\hat{y}_i - \bar{y})^2 + \Sigma (y_i - \hat{y}_i)^2 + \underbrace{\Sigma 2 (\hat{y} - \bar{y})(y_i - \hat{y}_i)}_{=0}$$

$$0 = 2\Sigma (\hat{y}_i - \bar{y})(y_i - \hat{y}_i)$$

$$= \Sigma (\hat{y}_i y_i) - \Sigma (\hat{y}_i^2) - \Sigma (\bar{y} y_i) + \Sigma (\bar{y} \hat{y}_i)$$

$$= \Sigma (\hat{y}_i (\hat{y}_i + \hat{u}_i)) - \Sigma \hat{y}_i^2 - \bar{y} \Sigma (y_i) + \bar{y} \Sigma (\hat{y}_i)$$

$$= \cancel{\Sigma \hat{y}_i^2} + \Sigma \hat{y}_i \hat{u}_i - \cancel{\Sigma \hat{y}_i^2} - \bar{y} \Sigma (\cancel{y_i - \hat{y}_i})$$

$$\emptyset = \Sigma \hat{y}_i \hat{u}_i \checkmark \text{ if uncorrel}$$

R^2 calculated as:

- $R^2 = \Sigma (\hat{y}_i - \bar{y})^2 / \Sigma (y_i - \bar{y})^2 = ESS/TSS$ → proportion of total variation explained by model

OR

- $R^2 = 1 - (\Sigma \hat{u}_i^2) / \Sigma (y_i - \bar{y})^2 = 1 - RSS/TSS$ → proportion not explained by model

So R^2 measures proportion of var explained by model

- same regardless of units

- only take values b/w $\emptyset$ + 1
    - $R^2 = 0$ = no fit
    - $R^2 = 1$ = perfect fit
    - $R^2 \uparrow$ = better fit
    - $R^2 = (r_{yx})^2$ → square of correl coef b/w y + x
- ~~Evaluate~~ valuate R^2 wrt type of data
    - time series? ~~$R^2 > .8$~~ $R^2 \geq .8$
    - cross sectional? $R^2 \geq .5$
- low R^2 still have value → precise parameter est
- ONLY VALID IF MODEL HAS INTERCEPT
- Bad: can ↑ if more x added

Test one

[Topic 3] → measure of best fit

[Adj R^2]

- assess if add x will ↑ predictability of y
- $\bar{R}^2 = 1-\left[\frac{RSS}{TSS} \cdot \frac{n-1}{n-k}\right]$
  - always $< R^2$
  - not same interp as R^2; can be neg
  - $\bar{R}^2$ better for model w/ lots var + few obser

Final Notes:

- Any var believed to have direct notable effect on y → include
  - having extra var prefer to leaving out
    ↳ $E[U] \neq 0$ ⁞

- indirect effect? leave out

[Topic 4] → Normal error model + Ci / hypot test.

[Normal-Error model]

- Assumptions:
  - $cov[u, x_j] = 0$
  - $E[u_i] = 0$
  - $var[u_i] = \sigma^2$
  - $cov[u_i, u_j] = 0$ for all $i \neq j$
  - error term is assumed Normal

if

- $cov[u, x_j] = 0$ } all ols asump
- $E[u] = 0$
- if in addition u is iid

} $\hat{B} \sim N(B, \sigma^2(x'x)^{-1})$
$\hat{B}_j \sim N(B_j, var[\hat{B}_j])$

- B/c → OLS estimator is linear function of U
  - so B_e w/in $B_j \pm se\,\hat{B}_j$ → 68%
    - 2 se → 95
    - 3 se → 99

[Confidence intervals]

- $t = \hat{B}_j - B_j / \hat{se}[\hat{B}_j]$
  - *vs. Z model $B_j - B / se[\hat{B}_j]$
- $Pr[-t_{(\alpha/2, n-k)} < t < t_{(\alpha/2, n-k)}] = 1-\alpha$
- $Pr\left[\hat{B}_j - t_{(\alpha/2, n-k)}\hat{se}[\hat{B}_j] < B_j < \hat{B}_j + t_{(\alpha/2 \ldots)}\,\hat{se}[\hat{B}_j]\right] = 1-\alpha$

Test one

[Topic 4/5] → hypotest

[Hypotest]

$H_0: B_j = c$ vs $H_a: B_j \neq c$

↳ compute $t = (\hat{B}_j - c)/\hat{se}[\hat{B}_j]$

· How likely that particular t val is to have been pulled from $t_{(n-k)}$

- select α
- Find table t^*
- if $|t| > t^*$ reject H_0 in favor H_a; $B_j \neq$ w/ α > possibility of being mistaken
- if $|t| < t^*$ one cant reject H_0; prob H_0 is correct is unknown

Statistical Hypothesis testing

· t test only valid if ols assump hold / $u \sim N$

· if sample size large u doesnt have to be ~N

· use t ~~test~~ stat to test if B is statistically diff from 0 → x affects y

· $H_0: B_j = 0$ → x no effect

· $H_a: B_j \neq 0$ → x has effect

· two tail to see B (−) or (+)

· t test reduce to $\hat{B}_j / \hat{se}[\hat{B}_j]$

↳ Rule of thumb $\hat{B} / se[\hat{B}_j]$ → 2 reject

· Pvalue: exact | lowest α @ which $H_0: B_j = 0$ can be rejected in favor of 2 tail

· opt for 1 tail → t val not alt but critical value Δ → fill all α in one tail

[CI for y_0]

· $\hat{y}_0 \pm t\,\hat{se}[\hat{v}_0] = y_0$

· $\hat{y}_0 \pm t\,\hat{se}[\hat{y}_0] = E[y_0]$ → narrower

$U'U = RSS$

$(y - x\hat{B})(y - x\hat{B})'$

$y'y - 2y'x\hat{B} - \hat{B}'x'x\hat{B}$

Remember*

* $\frac{dRSS}{d\hat{B}} = 0 - 2x'y + (x'x + (x'x)')x\hat{B}$

$= -2x'y + 2(x'x)\hat{B}$

$= 2(x'x)\hat{B} = 2x'y$

$\underbrace{(x'x)^{-1}(x'x)}_{I}\hat{B} = x'y\,(x'x)^{-1}$

$\boxed{\hat{B} = (x'x)^{-1}x'y}$

Test TWO → oct 2 +

chapt 8 → F test

other F test

↳ F test w/ linear Function of Regression coef

- $F^* = ((RSS_R - RSS_{UR})/q) / ((RSS_{UR})/(n-k))$
- can be extended to restrictions involv. linear functions.
- each linear function implies 1 restriction regardless of # of coef involved

↳ F test for equality of coef of dif Regressions

- see if model applies to same / somewhat dif dataset
- "chow" test
- can have dif # obs; but dep var are dif but closely related; + x var the same
- Ho: identical $B_1 = \alpha_1$ $B_2 = \alpha_2$; Ha → one pair (@ least) not identical
- UR → $RSS_1 + RSS_2$ → df sum of df from the two models → $(n-k) + (m-k)$ $m+n-2k$
- R → RSS merge
- $F^* = ((RSS_R - RSS_{UR})/k) / (RSS_{UR}/(n+m-2k))$
- Ho rejected? can't pool data → or use dummy var to bridge.

chapt 9 → Dummy Var → Regression

- qualitative variables → no natural order
  ↳ can be modeled w/ dummy var
- set of dummies for each categorical IV
- leave out one dum. onset → avoid multicol
- intercept ~~slope~~ dum → just parameter and variable: $D_{21} + B_2$
- intercept is var for excluded cat.
- slope dum → param; dum IV + reg IV.
  ↳ param on reg IV is when left out cat is used

Chapt 10 → Multicollinearity

Multicollinearity:

- when two or more indep variables in a regression model are highly correl. to eachother → difficult to tell ~~which~~ if each indiv affect y + quantify effect
- occurs b/c (everything CP) → se assoc w/ given OLS parameter est will be higher if IV is corr (highly) w/ other indep models
  ↳ larger error ~~changes~~ se → correl .5 ? double var; $\sqrt{\text{effect on var}}$ = se effect
  ↳ leads to key var w/ not signif t test
- Not mistake or violator; just undesirable
- More common in time series data

5 of 90

Test two (Oct 2?) 10/10

Chapt 4 → Regression Stuff.

Regression through the origin:

→ used when theory dictates intercept = ∅ → when all explan var are ∅ y ~~sst~~ should = ∅.

· estimated easily by exclud 1st col of 1s.

→ changes parameter estimates

→ R^2 not approp: No longer bounded b/w 0+1; TSS ≠ RSS + ESS

→ can use "Raw R^2" → not the same as R^2

↳ $(\text{corr}[y, \hat{y}]^2)$

→ ONLY recommended if STRONG evidence that true pop val of intercept is zero

Scaling + units of measurement:

· a change in scaling of any x var by fixed factor (w_x) only Δ parameters + se for that x by factor of $\frac{1}{w_x}$

· a change in scale of y → by w_y: Δ value for all model parameters + se by factor of w_y

Standard Regression:

→ Regression w/ all var (x+y) standardized to mean ∅ + variance of one

- for each var: subtract mean + divide ~~by~~ stand. deviation

- Parameters of this are "Beta coefficients"

- intercept would be ∅ here

- β coef interp: y + x measured in sd → so l Δ in y if x Δ by one sd

→ elim unit of measure → used for rel import of x on explaining y BUT doesnt say anything about RATE of flux. in y

$W_x = \frac{1}{s_y}$ $\frac{1}{W_x} = s_x$

$y_i^* = \frac{y_i}{s_y} - \frac{\bar{y}}{s_y}$

$x_i^* = \frac{x_i}{s_x} - \frac{\bar{x}}{s_x}$

Lagged variables:

· $Y_t = B_1 + B_2 X_{t-1} + U_t$ (t=2)

lagged

→ can also be more than one lag.

lose one ~~var~~ obs for every lag

Chapt 6:

first differences variable:

→ change from $t \to t+1$

↳ $y = B_1 + B_2(x_t - x_{t-1}) + U_t$ $(t=2)$ *do this before running ols

or

$Y = B_1 + B_2 \Delta x_t + U_t$

- need to belive Δ in x affects value taken by y in linear fashion

Alt model specifications

- Regression models so far assume relation b/w Y & indep var is linear
  *ols req for linear in param; but not nces. in expl var./dep. var.

- log-linear:
  - $\hat{Y}_i = e^{\hat{B}_1} x_2^{\hat{B}_2} x_3^{\hat{B}_3} \dots x_k^{\hat{B}_k} \to \ln(y_i) = \cancel{\ln} B_1 \ln(e) + B_2 \ln x_2 + B_3 \ln x_3 \dots + B_k \ln(x_k) + u_i \ln(e)$
  - incr @ incr rate
  - incr @ decr rate *Take lns before ols
  - decr @ decr rate
  - $\hat{B}_j < 0$; $x_j \uparrow$; $\hat{y} \downarrow$ @ decr rate; $\hat{B}_j = 1 \to x_j \uparrow$ $\hat{y} \uparrow$ @ const rate; $0 < \hat{B}_j < 1 \to x_j \uparrow$ $\hat{y} \uparrow$ @ decr rate
    $\hat{B}_j > 1 \to x_j \uparrow$; $\hat{y} \uparrow$ @ incr rate
  - disadvantage: all y + x_j relations assumed to be non linear
    - all x + y must be positive; can have ø (not advised
  - $\hat{B}_j$ directly measures elast of Y WRT x_j → %ΔY when xΔ1%
  - R^2 not right → not bounded → use Kno R^2 or square of corrl. coef. b/w observed any $\hat{y}$

Semi log: log ~~lin~~ lin
- log dep var; all rest are normal
- $\hat{B}_j$ → multiplied by 100 → %Δy when xΔ 1 unit.
- $B_j > 0$ x↑; y↑ @ incr rate
- $B_j < 0$; x↑; y↓ @ decr rate
- Y can't be neg; x can

Semi log model: lin log
- log of explan var is taken; y left alone
- B_j → unit ΔY for % Δ X → divide by 100
- $B_j > 0$; x↑; y↑ @ decr rate
- $B_j < 0$; x↑; y↓ @ decr rate

Test two: oct27 3 of 10

Chapt 6:

Alt model specifications:

Reciprocal: $\frac{1}{x_i}$

- $B_j > 0$; $x\uparrow$; $y\downarrow$ @ decr rate
- $B_j < 0$; $x\uparrow$ $y\uparrow$ @ decr rate
- has asymtope
- slope depend on x

* do all model manipulation before OLS

Log Reciprocal:

- $\ln y$ & $\frac{1}{x}$
- get S curve or decr@ decr rate

Polynomial:

- include x_i^2
- as $x\uparrow$: y can $\uparrow$ or $\uparrow$ @ incr or decr rate (very flexible)
  - ↳ might lead to reversal of relation
- can combine models & easily test for nonlinearity
  - ↳ ascan others that alter x NOT Y
- if $H_0: B_1 = 0$ can't be rejected → linearity can't be disproved
- can create multi colin problem

Choice of functional form:

- underlying theory will help decide
- Also consider: · sign of parameter WRT theory
  - statistical signif of param in alt model
  - R^2 or adj R^2 as long as dep var same across choices
- Residual analysis! :

Chapt 8 → F test:

F test:

- Recall $TSS = \Sigma(y_i - \bar{y})^2 \rightarrow n-1$ df
  - $ESS = \Sigma(\hat{y}_i - \bar{y})^2 \rightarrow k-1$ df
  - $RSS = \Sigma(y_i - \hat{y}_i)^2 \rightarrow n-k$ df

* df has something to do w/ # of indiv rand var ~~used~~ involved in calculation

- greater the RSS to ESS → more unexplained
- F test used to evaluate the null hypothesis: all model parameters w/ exception of intercept are zero
  - ↳ $H_0: B_2 = B_3 = \ldots = B_k = 0$
  - ↳ means none of the indep var chosen have any affect on y

Test Two → Oct 27
Chapt 8 → Ftest 4 of 10

Ftest:

- If Ho is not rejected → model as a whole cannot be said to be useful explaining y
- Ha: at least one indep var affects y
- use vas depvar? → don't want to reject
- $F^* = (ESS/k-1)/(RSS/n-k)$
- 4 OLS Assump hold + u is $\sim N$ (small samples); F^* follows F distrib w/ $(k-1)$ + $(n-k)$ df
- F^{**} → pval for F ; exact probab w/ which we allow Ho reject

Other Ftests:

↳ Joint test w/ several coef:

- eval subgroup as whole
- UR (unrestricted) Model: $Y = B_1 + B_2X_2 + B_3X_3 \dots B_kX_k + U$
- want to test subset q of regression coef if = (jointly) to zero
- Rewrite Model:

$$\underbrace{Y = B_1 + B_2X_2 + \dots B_{k-q}X_{k-q}}_{\text{Restricted Model}} + B_{k-q+1}X_{k-q+1} + B_kX_k + U$$

- $H_0: B_{k-q+1} = B_{k-q+2} = \dots = B_k = 0$
  - ↳ q = # restrictions
- estimate R + UR: $RSS_R \geq RSS_{UR}$
- if Ho correct; drop var have little affect on RSS → RSS_R slightly higher RSS_{UR}
- $F^* = ((RSS_R - RSS_{UR})/q) / ((RSS_{ur})/(n-k))$
  - ↳ Ho correct: follow F w/ q df (num) + $n-k$ (denom)
- not the same as indiv t test **
- Need OLS Assump. + $U \sim N$
- can get insignif t test + still have Ftest that rejects Ho (impervious to multicoll.)

↳ Ftest w/ linear function of regression coef

→ in cobb douglas → sum of expon = RTS
→ test CRTC

UR: $\ln \hat{y}_i = \hat{B}_1 + B_2 \ln x_2 + B_3 \ln x_3$

Restriction: $B_2 + B_3 = 1$ or ~~B_2~~ $B_2 = 1 - B_3$

R: $\ln y = B_1 + (1 - B_3)\ln x_2 + B_3 \ln x_3$ → Pay attention to sign → $B_2 + B_3 = 0$ vs. $B_2 + B_3 = 1$

$\ln y - \ln x_2 = B_1 + B_3(\ln x_3 - \ln x_2)$