ECONOMETRICS

economic ↲ ↳ measurements

def. brand of economics that uses mathematical + statistical tools to analyze economic phenomena
→ Allows us to deal w/ quantitative measurements, predict economic phenomena, + test hypothesis about economic phenomena.

Five Steps of Economic Study

Step 1 | Mathematical formulation of economic theory.

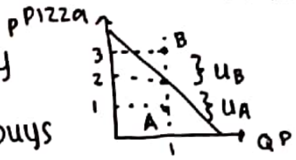
1.1. General (implicit) form of a function
↳ $y = f(x_1, x_2, x_3, \dots x_n) \Rightarrow Q_x^d f(P_x, P_y, M \dots)$ where y = dependent var.

1.2. Explicit form of a function
↳ deterministic: can determine y if x_1 or/+ x_2 is known
↳ $y = a + bx_1 + cx_2^2$
→ a (intercept): value not depending on Δ in x_1 or x_2
→ b + c (coefficients): physical properties or behavior that is constant under specific conditions
↳ responsiveness (rate of Δ) of dependent variable given the Δ in value of respective independent var.
→ ex: Demand func $Q_x^d = a + bP_x + cP_y + M$

Step 2 | Model Construction w/ view of Statistical Measurement + Testing

2.1. Econometric models are not deterministic

2.2. Random disturbance error term (u): allow for random deviations
↳ ex: $Q_x^d = a + bP_x + u$



Amy (−): saving money
Bobby (+): hungry
on ave, a consumer buys 1 pizza for $2

2.3. Reasons for Error (not predictable)
↳ sampling error: should be random + representative; no bias
↳ errors of specification: values missed or other factors (immediate var) that cannot be predicted
↳ errors of measurement: rounding, irrelevant values

2.4. Properties of Error Terms
↳ unique to each observation
↳ unpredictable
↳ can be positive or negative but zero on average

Step 3 | Data Collection

3.1. Economists deal w/ observed data (result of actual outcomes)
↳ subject to influence or bias: must remove seasonal + trend behaviors, data access is limited, other sciences can use exp.

3.2. Types of Data
↳ cross-sectional data: data that consists of observations on a variety of units (individ, household, firm, gov't) taken at a given time
↳ time series data: data that consists of observations on a var. or several var. at diff. periods of time
→ often interdependent: dependent on past series
↳ want to work w/ independent data
→ must make sure its the same frequency
↳ daily (stock prices), weekly (money supply), monthly (CPI, PPI), quarterly (GDP), annually (pop)
↳ Panel or longitudinal data: data that consists of time series for each cross-sectional member in the data set
→ useful in addressing dynamics of change

Step 4 | ~~Statistical Estimation~~

4.1. Est parameters a or b (empirical interpretation of the economy)

Step 5 | ~~Statistical Inference Relating Economic Theory to Empirical Analysis~~

5.1. can test the reliability of data or theories

Applications: business decision, forecast, public policy, etc.

Review of Statistics

→ Probability Framework for Statistical Inference

1. Probability
   - ↳ experiment: a procedure that can be repeated theoretically an infinite amt of time + has a well-defined set of outcomes
   - ↳ sample outcome: potential eventuality of an experiment
   - ↳ sample space: set of all possible outcomes of the experiment
     - → Difficulties of running an experiment
       - ↳ confounding effects (omitted factors), simultaneous causality, + correlation does not imply causation
   - ↳ population: totality of all possible outcome; think as ∞ large
     - → sample: subset of a pop. to characterize a whole
   - ↳ random variable: variable that takes on a numerical value + has an outcome determined by an experiment (uncertain)
   - ↳ probability theory: mathematical study of randomness
   - ↳ probability: measure of uncertainty / how likely an event is
     - → Properties
       1. $0 \leq P(A) \leq 1$
       2. Denote sample set $S = \{A, B, C \dots\}$ then $P(A+B+C\dots) = 1$
       3. If each outcome is equally likely then $P(A) = \#$ of A occur./other
       4. $P(A) = 1 - P(A^c)$
       5. Addition rule
          - ↳ mutually exclusive: $P(A+B+C) = P(A) + P(B) + P(C)$
          - ↳ not " ": $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
       6. multiplication rule
          - ↳ independent: $P(A \text{ and } B) = P(A)P(B)$
          - ↳ not ": $P(A \text{ and } B) = P(B)P(A|B)$
          - ↳ conditional + independent: $P(A|B) = P(A)$
   - ↳ exclusive events: If A happens, B cannot
   - ↳ independent events: outcome A has no effect on B

   } If mutually exclusive, then dependent

2. Random (Stochastic) Variables + Probability Distribution
   - ↳ discrete RV: takes on finite values
     - → probability mass function

       $f(x) \begin{cases} P(X=x) & \text{for } i = 1, \dots \\ 0 & \text{for } x \neq x_i \end{cases}$

     - → relates possible outcomes w/pr
     - → histogram
   - ↳ continuous RV: takes on infinite va
     - → probability density function
       1. $f(x) \geq 0$
       2. $\int_{-\infty}^{\infty} f(x)dx = 1$
       3. $\int_a^b f(x)dx = P(a < x < b)$
     - → tells us likelihood of random
     - → graph is in a bell curve
     - → probability of one value = 0
       - ↳ need a range

Cumulative Distribution Function

→ both continuous + discrete

→ $P(X \leq x)$ w/one RV

↳ marginal cdf

→ $f(x,y) = P(X \leq x, Y \leq y)$ w/2 or more RV

↳ joint cdf

Conditional Probability

$$Pr_{X|X,Y}(X=x|Y=y) = \frac{Pr(X=x|Y=y)}{Pr(Y=y)}$$

→ conditional expect.

$$E_{X|Y=y}(X|Y=y) = \sum_{i=1}^{z} P(X_i|y)X_i$$

* this is a number whereas $E_{X|Y}(X|Y) = RV$ b/c Y can take probabilities

* $E_Y[E_{X|Y}(X|Y)] = \#$; source of random was Y, but now weighted

3. Features of Probability Distributions

↳ expected value or mean: center of mass of distribution

→ discrete RV: weighted ave. of all possible values taking the probability of each outcome as its weight

$E(X) = x_1p_1 + x_2p_2 + \dots = \sum_{i=1}^{n} x_i p_i$

↳ population mean $= \mu_x$

↳ sample mean $= \bar{x}$

$\left.\right\} E(\bar{x}) = \mu_x$

* expected value of func. of DRV: $E(g(x)) = \sum g(x_i)(p_i)$

→ continuous RV: $E(X) = \int_{-\infty}^{\infty} x p(x) dx$

→ Expected Value Properties

1. $E(c) = c$ where c is a constan[t]
2. If you have RV X, Y, + Z then
3. If X = RV + c = constant then
4. $E(XY) = E(X)E(Y)$ if X + Y are

→ ex: $E(2x-10) = E(2x) - 10$ (proper[ty]

↳ variance: spread about the mean

→ If $\mu = E(X)$ then $Var(x) = E[(x-\mu)$ ...y

→ $sd(x) = \sigma_x = \sqrt{var(x)}$ or $var(x) =$ ...

→ Proof: $var(x) = E(X-\mu)^2 = E(x^2$ ...

$= E(x^2) - 2E(x\mu)$ ...

$= E(x^2) - 2E(xE($ ...

$= E(x^2) - E(x)^2$

Variance Examples:

1. $\mu_x = 5$ + $\mu_y = 1$

$E(x - 5Y + 10) = E(x) - 5E(Y) + E(10)$

$= E(x) - 5E(Y) + 10$

$= 5 - 5 + 10 = 10$

$Var(x - 5Y + 10) = var(x) + 25\,var(Y) + 0$

* assume x + y are independent

2. customer order $\mu_A = \$50$ + $\sigma_A = 11$ w/10,000 independent orders

E(total) = \$500,000

$sd(total) = \sqrt{var(\sum x_i)}$

$= \sqrt{\sum (11)^2}$

$= \sqrt{1210000} = 1100$

3. Bernoulli $Var = P(1-P)$ where P = probability that x = 1

→ Variance properties

1. $Var(c) = 0$ where c = constant
2. $Var(ax+b) = a^2 var(x) \Rightarrow Va$ ...
3. $Var(x+y) = var(x) + Var(y)$ ...
4. $Var(x+y) = var(x) + var(y) + 2cov(x, y)$

→ ex: weekend $E(\pi) = 3000$

weekday $E(\pi) = 2000$

sd = 150 for both

week $E(\pi) = E(2x + 5y) = 2E(x) + 5E(Y)$

$sd(w\pi) = sd(2x + 5y)$

↳ $var(2x + 5y) = 4var(x) + 25var(Y)$

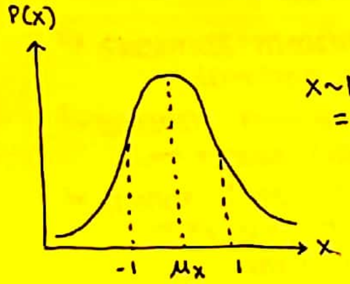
$= 4(150)^2 + 25(150)^2$

→ ex2: $R_a + R_b$ are monthly returns on stocks a + b

↳ b has a larger variance

↳ a is riskier but average return is greater

Standard Distribution



$x \sim N(\mu, \sigma^2)$

$= N(0, 1)$

mean = 0, var = 1, skewness = 0

$P(x > 0) = 50\%$

± 1 SD = 68%

$f(x) = e^{-x^2/2} / \sqrt{2\pi}$ when continuous

... about the mean

... on where likelihood is greater

...ss

...reater probability of outliers

* note kurtosis + variance are independent!

...measure of risk

...tock $\sigma_y^m = 0.1$

$\sqrt{12\sigma^2{}_y^m} = \sqrt{12}\,\sigma_y^m$

... + Y) their pop. cov is $cov(X, Y) = \sigma_{xy}$

$\dots E(Y))] = E[(X - \mu_x)(Y - \mu_y)] = \sigma_{xy}$

... population mean

...s above the mean, Y is also (on ave)

__________, Y is below the mean on ave

$\dots \mu_y)] = E(XY) - \mu_x \mu_y$

...not tell us how strongly they are related

→ value Δs depending on unit of measurement

→ covariance Properties

1. If X + Y are independent, then $Cov(X, Y) = 0$ (converse not true)

↳ $cov = E(X)E(Y) - \mu_x\mu_y = 0$

2. If $Y = V + W$ (RVs) then $Cov(X, Y) = Cov(X, V + W) = Cov(X, V) + Cov(X, W)$
3. For any constant a + RV X + Y, $Cov(aX, Y) = a\,Cov(X, Y)$
4. ______________, $Cov(X, a) = 0$

↳ correlation coefficient: $corr(X,Y) = \frac{cov(X,Y)}{sdX\,sdY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \rho_{XY}$

→ measures the strength + the direction

→ correlation coefficient properties

1. $-1 \le \rho_{XY} \le 1$
2. If $\rho_{XY} = 1$ then X + Y are perfectly positively linearly related
   ↳ we have $Y = aX + b$ where $a > 0$

$$corr(X,Y) = \frac{cov(X, aX+b)}{sd(Y)sd(X)} = \frac{cov(X,aX) + cov(X,b)}{\sigma_X \sigma(aX+b)}$$

$$= \frac{a\,cov(X,X)}{\sigma_X \sqrt{a^2\sigma_X^2}} = \frac{var(X)}{\sigma_X^2} = 1$$

   ↳ this means we can tell you Y for certain if X is known

3. If $\rho_{XY} = -1$ then X + Y are perfectly negatively linearly related
4. If $\rho_{XY} = 0$ then X + Y are not linearly related
   ↳ but may be non-linearly related!
5. $Corr(aX, bY) = Corr(X,Y)$ if $a > 0$ + $b > 0$
6. $Corr(aX, bY) = -Corr(X,Y)$ if $a < 0$ or $b < 0$ (but not both)
   ↳ $corr(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} \Rightarrow corr(aX,bY) = \frac{cov(aX,bY)}{\sqrt{var(aX)var(bY)}}$

   *values cancel out

$$= \frac{ab\,cov(X,Y)}{ab\,\sigma_X \sigma_Y}$$

→ ex: 9 stocks (S) — should you buy new stock A?

$\mu_S = 0.15$ $\mu_A = 0.2$
$\sigma_S = 0.1$ $\sigma_A = 0.12$
$\sigma_{S,A}^2 = 0.0012$

New Portfolio $= 0.9P_S + 0.1P_A$

$\mu_{NP} = 0.9(0.15) + (0.1)(0.2) = 0.155$

$\sigma_{NP} = \sqrt{var(0.9P_S + 0.1P_A)}$

$= var(0.9P_S) + var(0.1P_A) + 2cov(0.9P_S, 0.1P_A)$

$= (0.9)^2 var(P_S) + (0.1)^2 var(P_A) + 2(0.9)(0.1)cov(P_S, P_A)$

$= (0.9)^2(0.1)^2 + (0.1)^2(0.12)^2 + (2)(0.9)(0.1)(0.0012)$

$= 0.00846 = \sqrt{0.00846} \approx 0.0919$

*high return + low risk!

4. **Sampling + Estimators**

→ **estimator**: rule or func. of a data set to find its value

↳ **estimate** changes from every sample

↳ estimators of population mean (μ)

→ $\bar{X} = \frac{1}{n}\sum_{i=0} X_i$

↳ estimator of pop. variance (σ^2): $\hat{\sigma}^2 = \Sigma(x_i - \bar{x})^2\left(\frac{1}{n-1}\right)$

→ lose one degree of freedom

↳ estimator of cov: $\widehat{cov(X,Y)} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \hat{\sigma}_{xy}$

↳ estimators of corr: $\hat{\rho} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) / \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\Sigma(y_i - \bar{y})^2} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y}$

→ Desired Properties of Estimators

1. **Unbiasedness**: statistical sampling or testing error caused by systematically favoring some outcome over others
   ↳ unbiased estimator: an est. whose ave. value = pop. parameter
   ↳ **bias** $= E(\hat{\theta} - \theta)$; it is unbiased when $E(\bar{X}) = \mu$
   ↳ ex: $E\left(\frac{\Sigma(X_i - \bar{X})^2}{n-1}\right) = \sigma^2$ (better) $> E\left(\frac{\Sigma(X_i - \bar{X})^2}{n}\right) = \frac{n-1}{n}\sigma^2$ (underest.)
   ↳ θ = true parameter

   θ

   o = biased
   I = unbiased
2. **Efficiency**
   ↳ efficient estimators: unbiased est w/ var ≤ var of other unbiased est.
   → $\hat{\theta}_1$ is efficient if $var(\hat{\theta}) < var(\hat{\theta}_2)$
   → comparative concepts
   → can only compare efficiency of alt. est only if using the same information
   → If something is biased, it cannot be efficient
   ↳ ex: $\tilde{X} = \frac{1}{4}X_1 + \frac{1}{2}X_2 + \frac{1}{4}X_3$ where $X_1, X_2, + X_3$ are iid
   assume $E(X_1) = E(X_2) = E(X_3) = \mu$
   $E(\tilde{X}) = E(\frac{1}{4}X_1 + \frac{1}{2}X_2 + \frac{1}{4}X_3) = \frac{1}{4}\mu + \frac{1}{2}\mu + \frac{1}{4}\mu = \mu \Rightarrow$ unbiased

3. **Consistency**: possess a probability limit + its distribution collapses to a single point (true parameter) as sample size gets larger

$\text{Plim}\,\hat{\theta} = \theta \Rightarrow \lim_{n\to\infty} P(|\hat{\theta} - \theta| > \varepsilon) = 0$

↳ a single point is located at true value of pop. characteristics

↳ inconsistent: distribution collapses @ point other than true value

Simple Regression Model

I. Introduction

→ **regression analysis**: statistical methods that attempts to determine

1. the extent to which one variable (dependent) is dependent on one or more other variable(s) (independent)
2. expected mean value of the dependent variable for given values of the independent variable

II. Model

2.1. Population Regression

→ lose degrees of freedom when using a sample: use n-1

→ ex: study determinants of future income

x (yrs of edu)

Y in $1000	8	12	16	20
	32	35	41	45
	33	38	43	47
	34	39	44	48
	36	41	46	51
	37	42	47	53
	38	45	49	56

↳ dependent var: Y (income)

↳ independent var: X (edu)

↳ $E(y|x=8) = \frac{\Sigma(y_i|x=8)}{6} = 35$

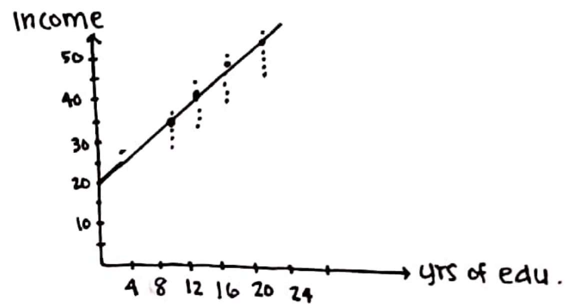
$E(y|x=12) = 40$

$E(y|x=16) = 45$

$E(y|x=20) = 50$

↳ Next steps:

1. Scatterplot
2. $E(y|x_i)$ plot
3. Population regression line: line that passes through $E(y|x_i)$
4. Define pop. regression func.



2.1.1. **The Population regression function**: the equation that describes the relationship between x_i + its respective $E(y|x_i)$

→ $E(y|x_i) = f(x_i) = \beta_0 + \beta_1 x_i$

*deterministic: $E(y)$ + x

→ **meanings of linearity**

1. Linear in variables: linear $\Rightarrow E(y) = \beta_0 + \beta_1 x_1$
   non linear $\Rightarrow E(y) = \beta_0 + \beta_1\sqrt{x_i}$

2. Linear in parameters: linear $\Rightarrow E(y) = \beta_0 + \beta_1 x_i$ or $E(y) = \beta_0 + \beta_1 \ln x_i$
   nonlinear $\Rightarrow E(y) = \beta_0 + \beta_1^2 x_i$ or $E(y) = \beta_0 + \ln(\beta_1 + x_i)$

→ **Linear regression function**: linear in parameters + either linear or nonlinear in variables (ex: $\ln(E(y)) = \beta_0 + \beta_1 \ln(x)$)

2.2.2. **Population Regression Relation**: equation that describes the relationship between the actual value of y + respective x_i (harder to predict)

→ $E(y|x_i) = f(x_i) = \beta_0 + \beta_1 x_i$

$y_i = E(y|x_i) + u_i$

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

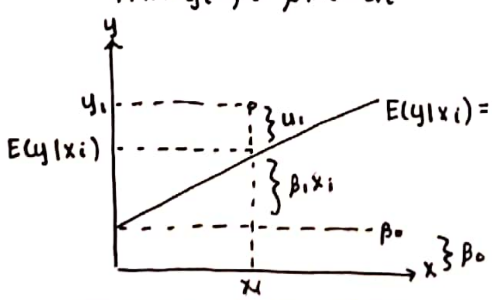
β_0 = intercept

β_1 = slope

y = dependent/explained/endog./predicted var

x = independ./explanatory/exog/predictor/regressor

u = error/disturbance that describes other factors besides x that influences y



linearity implies $\Delta y = \beta_1 \Delta x$ or $\beta_1 = \frac{\Delta y}{\Delta x}$

*run regression of Y on X

2.2. Sample Regression

2.2.1. Sample Regression Function + Line: describes relationship between x_i + its respective estimated values of β_0, β_1, + $E(y_i)$

→ $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ where $\hat{y}_i$ is an estimator of true $E(y|x_i)$ + $\hat{\beta}_0$ is an estimate of β_0, etc. * $\hat{u}_i$ = residual

III. Estimation

*SLR = simple linear regression model

3.1. Assumptions

1. SLR1: Linear in Parameters: defines population model
   → relation between x + y are linear
2. SLR2: Var X is non-random across samples
   → X are nonstochastic var. whose values are fixed (constant)
   → values of X receive is same throughout time = not realistic
3. SLR3: $Cov(x_i, u_i) = 0$ *errors + influence of x are different
   → independent influence on Y $(E(u|x) = E(u))$
   → only want β_1 to be the influence on x
4. SLR4: Error term has an expected value of 0
   → $E(u_i) = 0$ for $i = 1, 2, 3 \ldots N$
   → not realistically independent (income = $\beta_0 + \beta_1$education + u_i)
5. Homoskedasticity ("equal spread"): error terms have a constant var for all observ
   → $Var(u) = E(u^2) = \sigma_u^2$
   → if var is not constant, alt. situation is heteroskedasticity
6. SLR6: No autocorrelation: error terms are statistically independent thus $cov(u_i, u_j) = 0$ for all $i \neq j$, $i, j = 1 \ldots N$
   → diff. error terms are independent (not systematic)
   ↳ serial correlation if they are correlated
7. SLR7: Error terms are normally distributed ⇒ $u_i \sim N(0, \sigma_u^2)$

Other Assumptions

8. Model is correctly specified (mispecified: nonlinear, true model needs >1 equation)
9. Relationship being estimated is identified (has unique mathematical form)
10. The xs are measured accurately, independent of error term
11. All var are correctly aggregated
12. Number of observ > number of coefficients being estimated or > the number of var in the model
    → need more degrees of freedom to be more precise
    ↳ positive errors cancelled by neg. errors

3.2. Ordinary Least Squares

→ PRR: $y_i = \beta_0 + \beta_1 x_i + u_i$ ⇒ goal to est. $\beta_0 + \beta_1$ $\{y_i, x_i\}$ $i = 1, \ldots, n$

→ fitted value: value predicted for y when $x = x_i$ $(\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x)$

→ residual: diff. between actual value of $y_i + \hat{y}_i$ $(\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x)$

↳ n residuals ⇒ goal: minimize them

→ $\sum_{i=1}^{n} \hat{u}_i$ but on average, error terms = 0

→ $\sum_{i=1}^{n} (\hat{u}_i^2) = \hat{u}_1^2 + \hat{u}_2^2 + \cdots + \hat{u}_n^2$ (sum of squared Residuals or SSR/RSS)

↳ $SRR = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2$

3.3. Derivation of the OLS Estimators

1. Differentiate SSR w/respect to $\hat{\beta}_0 + \hat{\beta}_1$ + set them = 0

a) $\frac{dSSR}{d\hat{\beta}_0} = \frac{d(\Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2)}{d\beta_0}$

$= -2\Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$

$= \Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$

b) $\frac{dSSR}{d\hat{\beta}_1} = 2\Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i)$

$= -2x_i\Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$

$= x_i\Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$

2. Transform equations using the summation rules

a) $\sum_{i=1}^{n} y_i - \sum \hat{\beta_0} - \sum \hat{\beta_1} x_i = 0$

$\sum y_i - n\hat{\beta_0} - \hat{\beta_1}\sum x_i = 0$

b) $\sum y_i x_i - \sum \hat{\beta_0} x_i - \sum \hat{\beta_1} x_i^2 = 0$

$\sum y_i x_i - \hat{\beta_0}\sum x_i - \hat{\beta_1}\sum x_i^2 = 0$

3. Put equation into normal form (unknown on one side + β_0/β_1 on another)

a) $\sum y_i = n\hat{\beta_0} + \hat{\beta_1}\sum x_i$

b) $\sum y_i x_i = \hat{\beta_0}\sum x_i + \hat{\beta_1}\sum x_i^2$

4. Solve the normal equations for $\hat{\beta_0} + \hat{\beta_1}$

a) $\sum y_i = n\hat{\beta_0} + \hat{\beta_1}\sum x_i$

$n\hat{\beta_0} = \sum y_i - \hat{\beta_1}\sum x_i$

$\hat{\beta_0} = \frac{\sum y_i}{n} - \hat{\beta_1}\frac{\sum x_i}{n}$

$\hat{\beta_0} = \bar{y} - \hat{\beta_1}\bar{x}$

*use $\sum x_i(x_i - \bar{x}) = \sum(x_i - \bar{x})^2$

$\sum x_i(y_i - \bar{y}) = \sum(x_i - \bar{x})(y_i - \bar{y})$

b) $\sum y_i x_i = \hat{\beta_0}\sum x_i + \hat{\beta_1}\sum x_i^2$

$\sum y_i x_i = (\bar{y} - \hat{\beta_1}\bar{x})\sum x_i + \hat{\beta_1}\sum x_i^2$

$\sum y_i x_i = \bar{y}\sum x_i - \hat{\beta_1}\bar{x}\sum x_i + \hat{\beta_1}\sum x_i^2$

$\sum y_i x_i - \bar{y}\sum x_i = \hat{\beta_1}\sum x_i^2 - \hat{\beta_1}\bar{x}\sum x_i$

$= \hat{\beta_1}(\sum x_i^2 - \bar{x}\sum x_i)$

$\hat{\beta_1} = \frac{\sum y_i x_i - \bar{y}\sum x_i}{\sum x_i^2 - \bar{x}\sum x_i} = \frac{\sum x_i(y_i - \bar{y})}{\sum x_i(x_i - \bar{x})}$

$\hat{\beta_1} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$

$\hat{\beta_1} = \frac{\hat{cov}(x,y)}{\hat{var}(x)} = \frac{\hat{\sigma}^2_{xy}}{\hat{\sigma}^2_x}$

3.4. Interpretation

→ $\hat{\beta_0}$: predicted value of y when x=0

→ $\hat{\beta_1}$: predicted value of y when x Δs 1 unit ($\Delta\hat{y}/\Delta x$ or $\Delta\hat{y} = \hat{\beta_1}\Delta x$)

→ Big Picture: PRF ⇒ $E(y|x) = \beta_0 + \beta_1 x$

↳ but $\beta_0 + \beta_1$ are unknown so SRF: $\hat{y} = \hat{\beta_0} + \hat{\beta_1}x$

→ ex: You have PRF E(col GPA | HS GPA) = β_0 + β_1 (HS GPA) where n=141

↳ SRF: $\widehat{colGPA} = 1.32 + 0.482$ (HS GPA)

→ positive relationship

→ for every point incr. in GPA (HS), your colGPA is expected to increase by 0.482

→ but if you have 0 HSGPA then you get a 1.32 colGPA?

↳ may be due to lack of 0 HS GPA in sample

→ ex 2 → PRF: $GPA = \beta_0 + \beta_1 candy + u$

↳ $\hat{\beta_1} = \frac{\hat{cov}_{xy}}{\hat{\sigma}_x^2} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$

↳ $\hat{\beta_0} = \bar{y} - \hat{\beta_1}\bar{x}$

1. Find $\bar{x} + \bar{y}$

→ $\bar{x} = 5 + \bar{y} = 3$

2. Find $(x_i - \bar{x}) + (y_i - \bar{y})$

student	candy (x)	GPA (y)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	5	3	0	0	0	0
2	4	2.5	-1	-0.5	0.5	1
3	4	3.5	-1	0.5	-0.5	1
4	9	4	4	1	4	16
5	3	2	-2	-1	+2	4
					$\Sigma = 6$	$\Sigma = 22$

$\hat{\beta_1} = 6/22 \approx 0.27$

$\hat{\beta_0} = 3 - 0.27(5) \approx 1.63$

$\widehat{GPA} = 1.63 + 0.27$ candy

β_0 = GPA of student who doesn't eat candy (implausible: no 0 candy in sample)

β_1 = for add. candy, predicted GPA incr. of 0.27

if x=5 then $\widehat{GPA} = 3$

if x=1 then $\widehat{GPA} = 2$

> 4 extra candy ⇒ 1 point incr. for GPA

→ To see data analysis on excel go to:

1. File
2. Options
3. Add-Ins
4. Analysis Tool Bank
5. Data Analysis

3.5. Goodness of Fit: how well the line fits the data

→ residuals (large = poor fit; small = good fit)

↳ $y_i = \hat{\beta_0} + \hat{\beta_i} x_i + \hat{u_i}$ } $\Delta y \Rightarrow$ results from Δx or Δu_i

explained by $\hat{\beta_0}, \hat{\beta_i}, x_i, + u_i$ — explained ↗ ↖ unexplained

→ Total Sum of Squares (TSS): how spread out y is from the mean

↳ $TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$ ↳ $y_i = \hat{y_i} + \hat{u_i}$

→ Explained Sum of Squares

↳ $ESS = \sum_{i=1}^{n} (\hat{y_i} - \bar{y})^2$

→ Residual Sum of Squares

↳ $RSS = SSR = \sum (y_i - \hat{y_i})^2 = \sum_{i=1}^{n} u^2$

* used to derive OLS

} $TSS = ESS + RSS$

R^2 = coefficient of Determination

$R^2 = ESS/TSS = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$

measures the fraction of variation (total) in y that is explained by our regression model / by Δ in x

$0 \leq R^2 \leq 1$

$R^2 = 0$

$R^2 = 1$

Interpretation: $R^2 \times 100$ = % of Δ explained by the Δ in x

ex: $\widehat{CEO\ salary} = 971 + 18.62 ROE$

$R^2 = 0.0132$

} 971 is CEO salary when ROE = 0

for every 1% increase on ROE, e(incr) = \$18.62

only 1.3% Δ in equity can explain Δ in CEO salary (98.7% Δ in salary explained by other variables)

ex 2: $\widehat{vote} = \beta_0 + \beta_1$ campaign expenditure in \$1K

$\beta_0 = 43.171163$

$\beta_1 = 0.0236042 \Rightarrow$ for every \$1K spent, gain 0.023 votes

$R^2 = 0.156141 \Rightarrow$ campaign exp. explains 15.61% variation of voting r.

$\widehat{vote} = \beta_0 + \beta_1$ (share of exp)

$R^2 = 0.856145 \Rightarrow$ ↑1% of shared expend., expect to get β_1 votes

$\beta_0 = 26.81 \Rightarrow$ 26.81 votes given no expend.

IV. Properties of OLS Statistics

4.1. Algebraic Properties

1. $\sum_{i=1}^{n} \hat{u}_i = 0$ (FOC of derivation of OLS estimators)

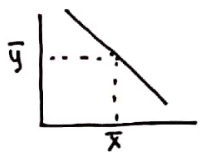
↳ $\sum (y_i - \hat{\beta_0} - \hat{\beta_i} x_i) = 0 \Rightarrow \sum \hat{u_i} = \sum (y_i - \hat{y})$

2. Covariance of regressors + OLS estimators = 0

$\sum_{i=1}^{n} x_i \hat{u_i} = 0$ (FOC (2))

3. OLS line passes through average of $(\bar{x}, \bar{y})$ →

$\bar{y} = \hat{\beta_0} + \hat{\beta_i} \bar{x}$



4.2. Statistical Properties

1. Unbiasedness of OLS: $E(\hat{\beta_i}) = \beta_1$

→ Proof: $\hat{\beta_i} = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}$ + $y_i = \beta_0 + \beta_1 x + u$ so

$$= \frac{\sum (x_i - \bar{x})(\beta_0 + \beta_1 x + u)}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})\beta_0}{\sum (x_i - \bar{x})^2} + \frac{\sum (x_i - \bar{x})\beta_1 x}{\sum (x_i - \bar{x})^2} + \frac{\sum (x_i - \bar{x}) u}{\sum (x_i - \bar{x})^2}$$

$$= \frac{\beta_0 \sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} + \frac{\beta_1 \sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} + \frac{\sum (x_i - \bar{x}) u}{\sum (x_i - \bar{x})^2}$$ + $\sum (x_i - \bar{x}) = 0$ so

$$= 0 + \beta_1 + \frac{\sum (x_i - \bar{x}) u}{\sum (x_i - \bar{x})^2}$$ when $\sum (x_i - \bar{x}) x_1 = \sum (x_i - \bar{x})^2$

Take expectation

$$E(\hat{\beta_i}) = E(\beta_1) + E\left(\frac{\sum (x_i - \bar{x}) u}{\sum (x_i - \bar{x})^2}\right)$$

$$E(\hat{\beta_i}) = \beta_1 + \frac{\sum (x_i - \bar{x}) E(u_i)}{\sum (x_i - \bar{x})^2}$$ } SLR 2: treat X as nonrandom; SLR4: $E(x_i - \bar{x}) = 0$

$E(\hat{\beta_i}) = \beta_1$

→ all SLR3: $\hat{\beta_i} = \beta_1 + \frac{\sum (x_i - \bar{x})(u_i - \bar{u})}{\sum (x_i - \bar{x})^2}$

Unbiasedness of $\hat{\beta_0}$

→ Proof: $\hat{\beta_0} = \bar{y} - \hat{\beta_1}\bar{x}$ + $\bar{y} = \beta_0 + \beta_1\bar{x} + u$ so

$$\hat{\beta_0} = (\beta_0 + \beta_1\bar{x} + u) - \hat{\beta_1}\bar{x} = \beta_0 + (\beta_1 - \hat{\beta_1})\bar{x} + \bar{u}$$

↳ Take the expected values of both sides:

$E(\hat{\beta_0}) = E[(\beta_0 + \beta_1\bar{x} + u) - \hat{\beta_1}\bar{x}]$ → alt SLR3: β

$= E(\beta_0) + E[(\beta_1 - \hat{\beta_1})\bar{x}] + E(\bar{u})$ } SLR2: treat x as nonrandom

$= \beta_0 + \bar{x}E(\beta_1 - \hat{\beta_1}) + 0$ } SLR4: $E(x_i - \bar{x}) = 0$

↳ Previously, we found $E(\beta_1) = E(\hat{\beta_1})$

$E(\hat{\beta_0}) = \beta_0$

2. **Variance of OLS**

*only works under SLR 5 (homoskedasticity); otherwise, formulas are invalid

a) $var(\hat{\beta_0}) = \dfrac{\sigma_u^2 \Sigma(\frac{x_i^2}{n})}{\Sigma(x_i - \bar{x})^2}$ b) $var(\hat{\beta_1}) = \dfrac{\sigma_u^2}{\Sigma(x_i - \bar{x})^2} = \dfrac{\sigma_u^2}{n\sigma_x^2}$

→ σ_u^2: larger quantity of dispersion, the more uncertainty (want this value to be small)

→ as sample size n↑, σ_u^2 ↓

→ more spread in x = more info on x

→ Proof for b: $Var(\hat{\beta_1}) = Var\left(\dfrac{\Sigma(x_i - \bar{x})u_i}{\Sigma(x_i - \bar{x})^2}\right)$

$$= \frac{Var(u)\Sigma(x_i - \bar{x})^2}{(\Sigma(x_i - \bar{x}))^2}$$

→ in order to use these variables, we need to know σ_u^2

↳ errors are unobserved (from PRF w/ β_0, β_1)

↳ residuals are computed from data (SRF w/ $\hat{\beta_0}$; $\hat{\beta_1}$)

$\sigma_u^2 = E(u^2)$ + we know $E(x) = \Sigma \frac{x_i}{n}$

$= \Sigma \frac{u_i^2}{n}$ ← unobserved but known est.

$= \Sigma \frac{\hat{u}_i^2}{n-2}$ ← fix 2 points ($\beta_0 + \beta_1$) so we lose 2 degrees of freedom

$\hat{u} = y_i - \hat{y}_i = y_i - \hat{\beta_0} - \hat{\beta_1}x_i$

→ **Standard Error of Regression**

↳ measures magnitude of typical regression residual in units of Y (alternative of R^2 as goodness of fit measure)

↳ $SER = \sqrt{\sigma_u^2} = \sigma_u$

Gauss - Markov Theorem:

Given the assumptions SLR1-6, the OLS estimators are the Best (most efficient) linear Unbiased Estimators [BLUE] in the sense that they have the minimum variance of all linear unbiased estimators

*does not involve SLR7

Recall the properties:

1. OLS Estimators are linear
   →easy to use
2. ______ are unbiased
   →centered around true values
3. ______ are efficient
   → more consistent to truth

Hypothesis Testing

I. Introduction

$y = \beta_0 + \beta_1 x + u$

hypothesize these values (β_0, β_1)

u: error term treated as sum of many diff. unobserved factors

we know:
$E(\hat{\beta_i}) = \beta_i,\ i = 0,1$
$var(\hat{\beta_i}) \rightarrow$ SLR5

→ need full sampling distribution
SLR7: $u_i \sim N(0, \sigma_u^2)$

Central Limit Theorem: If there is a large number of iid RVs, then w/few exceptions, the distribution of their sum tends to be a normally distribution as the number of such variables incr. indefinitely

Theorem: Assuming SLR1-SLR7, $\hat{\beta} \sim N(\beta_i, \sigma^2_{\hat{\beta_i}})$

→ If $X \sim N(-) \Rightarrow Y = a + bx \Rightarrow Y \sim N(-)$

→ therefore $\frac{\hat{\beta_i} - \beta_i}{sd(\hat{\beta_i})} \sim N(0,1)$

* if SLR7 does not hold then $var(\hat{\beta_i}) = var[(x_i - \mu_x)u_i] / n(\sigma_x^2)^2$

(1) **Significance test**: $H_0: \beta_i = 0$ → means x has no impact on y

(2) **General test**: $H_0: \beta_i = \beta^*$

II. Testing the significance of $\beta_0 + \beta_1$

→ **test statistic**: any function of observed data whose numerical value dictates whether or not the null is "accepted" or rejected

→ ex: wage $= \beta_0 + \beta_1$ education $+ u$

↳ null hypothesis: outcome you do not expect

→ $H_0: \beta_i = 0$ where $i = 0,1$

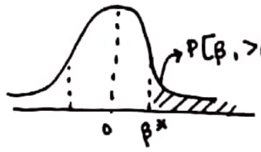
→ $H_1: \beta_i > 0$, $\beta_i < 0$ } two-sided
$\beta_i \neq 0$ } one-sided

Conclusion (1) reject
(2) fail to reject

→ P(reject H_0 | H_0 is true)

1. Specify the test statistic (rejection area)
2. Need to know its distribution
   - ↳ distribution of test statistic assuming H_0 is true
   - ↳ significance level (α): probability that the test statistic lies in the critical region (rejection region) when H_0 is true

$\hat{\beta} \sim N(\beta_i, \sigma^2_{\hat{\beta_i}})$



$P[\beta_1 > 0 \mid \beta_1 = 0] = \alpha = 0.05$

2.1. T-Test (Ratio)

$$t = \frac{\hat{\beta_i} - \beta_i^*}{se(\hat{\beta_i})} = \frac{\text{estimated} - \text{hypothesized value}}{\text{standard error}}$$

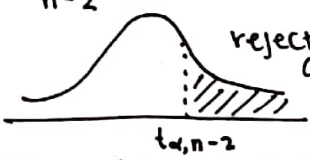
where $\hat{sd}(\hat{\beta}) = se(\hat{\beta})$

standard error ← fix for sampling uncertainty

→ If close in value, t-stat is small → fail to reject H_0

→ We know $\hat{\beta} = \frac{\hat{cov}(x,y)}{\hat{var}(x,y)}$ $se(\hat{\beta_i}) = \frac{\sigma_{\hat{u}}^2}{n\sigma_x^2}$ $\sigma_{\hat{u}}^2 = \frac{\Sigma \hat{u}^2}{n-2}$

* as n ↑, they converge



→ Proposition: Under SLR assumptions, $t = \frac{\hat{\beta_i} - \beta_i}{se(\hat{\beta_i})} \sim t_{n-2}$ (student distrib)

Null/H_0	H_1	reject H_0 if
(a) $\beta_1 = \beta_1^*$	$\beta_1 > \beta_1^*$	$t = \frac{\hat{\beta_i} - \beta_i^*}{se(\hat{\beta_i})} > t_{\alpha, n-2}$
(b) $\beta_1 = \beta_i^*$	$\beta_1 < \beta_i^*$	$t < -t_{\alpha, n-2}$
(c) $\beta_1 = \beta_i^*$	$\beta_1 \neq \beta_i^*$	$\lvert t \rvert > t_{\alpha, n-2}$

→ ex: Defensive expenditure w/other expenditure in the economy

$\text{consump}_t = 31 - 0.27\ \text{def}$, $R^2 = 0.71$
(2.9) (0.093) $n = 500$

$H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$ } do not use a hat! we know the est. parameter!

1. Assume H_0 is true
2. Find the t-statistic
3. Find critical value w/ degrees of freedom $(n-2)$, α, + one or two sided
4. Reject H_0

$$t = \frac{\hat{\beta_i} - \beta_i^*}{se(\hat{\beta_i})} = \frac{0.27-0}{0.093} = -2.9$$

$t_{\alpha,n-2} = t_{0.05,498} = 1.965$

$|-2.9| > 1.965$ (two-sided)

III. General Tests of Regression Coefficient ($H_0: \beta_i = \beta_i^*$)

ex: consumption function → $consump = \beta_0 + \beta_1 income + u$

→ β_1 = marginal propensity to consume

→ $H_0 = \beta_1 = 0.9$ (people save 10%)

$H_1: \beta_1 < 0.9$

$\widehat{consump} = 46 + 0.34\, income$
$(4.2)\ (0.102)\quad n = 800$

1. Assume the null is true
2. Find t-stat: $t = \frac{0.34-0.9}{0.102} = -5.49$
3. Critical value
   $\left.\begin{array}{l} d.f. \approx 800 \\ \alpha = 5\% \\ \text{one tail} \end{array}\right\} t_{0.05,800} = 1.645$
4. Rejection Rule: $t < -t_{\alpha,n-2}$
5. Reject H_0

ex2: (Jensen) Can hedge funds beat the market? (excess returns on market vs. excess returns on mutual funds)

→ model: $R_{jt} - R_{ft} = \alpha_j + \beta_j(R_{mt} - R_{ft}) + u_{jt}$ for $j = 1 \ldots 115$

R_{jt}: return on portfolio j at time t
R_{ft}: return on risk-free security on time t
R_{mt}: return on a market portfolio proxy

$\alpha = 0$ (not better than the market)
$\alpha > 0$ (outperform)
$\alpha < 0$ (worse than)
β: volatility of fund w/respect to market (>1 = risky)

$\left.\begin{array}{l} H_0: \alpha = 0 \\ H_1: \alpha \neq 0 \end{array}\right\}$ performance relative to the market

$\left.\begin{array}{l} H_0: \beta = 1 \\ H_1: \beta > 1 \end{array}\right\}$ risky or less risky

IV. P value

* If we choose a smaller α (critical value %), the critical value ↓

→ Q: Given observed value of t-stat, what is the smallest significance level at which the H_0 would be rejected?

↳ P value: lowest significance level at which H_0 can be rejected

↳ If we fail to reject H_0 @ 5%, then P value > 0.05

→ need to be less than the P value to be significant

* se is measured in units of y
* Data Analysis on excel: P values reported for 2 side tests

V. Confidence Intervals

→ Before, we only looked at the point est. $\hat{\beta}$ (looking @ one point)

↳ but β will change from sample to sample

↳ confidence intervals give a range of likely values of the parameter

→ Distribution

$$\left.\frac{\hat{\beta_i} - \beta_i}{sd(\hat{\beta})} \sim N(0,1) \quad + \quad \frac{\hat{\beta_i} - \beta_i}{se(\hat{\beta})} \sim t_{\alpha,n-2}\right\} \beta \in [\hat{\beta_i} \pm t_{\alpha,n-2} \cdot se(\hat{\beta})]$$

ex: $consump_t = 31 - 0.27\, def_t$, $n = 500$ @ 5% significance
$(2.9)\ (0.093)$

$(-0.27 - (1.965)(0.093), -0.27 + (1.965)(0.093)) = (-.452, .087)$

true value of β lies in this interval 95% of time

→ summary: Can test hypothesis w/

1. T-test
2. P value
3. Confidence Interval

Review:

→ SLR Assump needed to

1. Prove GM Theorem by allowing us to make est. of $\beta_0 + \beta_1$ (SLR 2-4)
2. Do hypothesis testing

↳ SLR3: when we have a regression, you want to est. effect of x on y through β_1. but if u is also affected, you cannot capture the influences of x independent of the error term

↳ SLR5: Used to find var($\hat{\beta}$) → simplify by only finding one

→ if homoskedasticity is changed, result for β_0 or β_1 is not affected because it does not involve

↳ invalidates hypothesis testing however

PART II NOTES

V. Nonlinear (in variables) Regressions Model

→ linear: every value of x has the same effect on y
→ looking at different models:

1. Log-Linear models: $\log(y) = \beta_0 + \beta_1 \log(x) + u$
2. Reciprocal models: $y = \beta_0 + \beta_1 \frac{1}{x} + u$
3. Polynomial models: $y = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_K x_i^K + u$

] linear in parameters

→ Log properties to know

$\ln(ab) = \ln(a) + \ln(b)$
$\ln(a/b) = \ln(a) - \ln(b)$
$\ln(x^a) = a\ln(x)$
$\ln(e) = 1 \Rightarrow e^0 = 1$

When β is big, we need:
$e^{\beta} - 1$ to test the effect

Numerically:
$\ln(1.01) = 0.00995 \approx 0.01$
$\ln(1.1) = 0.0953 \approx 0.1$ (sort of)

$\ln(x + \Delta x) - \ln(x) = \ln(1 + \frac{\Delta x}{x}) \approx \frac{\Delta x}{x}$

When $\beta < 0.2$, approx is good

5.1. Log-linear Regression Model

→ non-linear relationships th... ...g the ln of both sides of the equati...
→ ex: $y_i = e^{\beta_0} x^{\beta_1}$ ...
$\ln(y_i) = \ln(e^{\beta_1} x_i^{\beta_1} e^{u_i})$ ...
↳ if $y_i = \beta_0 x^{\beta_1}$ then ...
→ we have dx = absolute Δ = ...
→ relative Δ = (new − refere...
→ ex: $\ln(y) = \beta_0 + \beta_1 \ln(x) + u$
$\beta_1 = \frac{d\ln(y)}{d\ln(x)} = \frac{\text{relative } \Delta \dots}{\text{relative} \dots}$

Examples:

1. Coffee Demand
$\ln(\hat{y}_t) = -0.77 - 0.253\ln(x)$
↑ cups per day ↑ coffee price per lb
... increase of price of coffee, the quantity of coffee demanded increases by 0.253%

→ Semi-log models (log-lin or lin-log): one variable transformed when taking log or ln of one variable

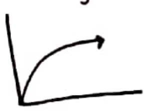
Example:

1. We have $\ln(wage) = \beta_0 + \beta_1 education + u$
↳ $wage = e^{\beta_0 + \beta_1 educ + u}$
$\beta_1 = \frac{d\ln(y)}{dx} = \frac{\text{relative } \Delta}{\text{abs. } \Delta} = \frac{\Delta y / y \cdot 100}{\Delta x \cdot 100}$
$100\beta_1 = \frac{\% \Delta y}{\Delta x}$

] If $\beta_1 = 0.082$ then every 1 yr of educ is expected to increase wage by 8.2%

*cannot infer if linear or log model has a better fit w/ R^2 because we Δ the y variable

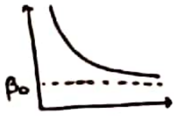
2. Lin-log Model (right transformed w/ logs)
↳ ex: GDP per capita + human development or wage + age
↳ good for decreasing returns to scale (÷100)



3. Log-lin model
↳ ex: $wage = 3 + 10\ln(age)$
$\beta_1 = \frac{dy}{d\ln(x)} = \frac{\Delta y}{\Delta x / x} \cdot \frac{100}{100} \Rightarrow \beta_1/100 = \Delta y / \% \Delta x$
↳ good for increasing returns to scale (×100)

Model	Dependent Var	Independent Var	Interpretation	
lin-lin	y	x	$\Delta y = \beta_1 \Delta x$	
lin-log	y	$\ln(x)$	$\Delta y = \beta_1/100 (\% \Delta x)$	
log-lin	$\ln y$	x	$\% \Delta y = 100\beta_1 \Delta x$	for 1 unit Δx, y predict to Δ $100\beta_1$
log-log	$\ln y$	$\ln(x)$	$\% \Delta y = \beta_1 \% \Delta x$	for every 1% Δ in x, y is expected to Δ β_1 % in y

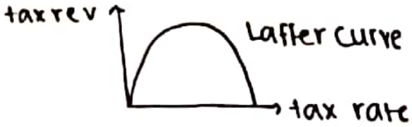
5.2. Reciprocal Models: $y = \beta_0 + \beta_1(\frac{1}{x}) + u$



→ x cannot be zero
→ β_1 cannot be interpretted the same

5.3. Polynomial Model: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k + u$

→ ex: $y = \beta_0 + \beta_1 x + \beta_2 x^2$



* quadratic w/a negative β^2
→ how to interpret the parameter:
↳ when x Δs by 1 unit
$x = 2 \Rightarrow x = 1$
then $\Delta y = [\beta_0 + \beta_1 2 + \beta_2 2^2] - [\beta_0 + (\beta_1 1) + \beta_2 1^2)$
$= \beta_1(2-1) + \beta_2(4-1)$

→ use statistical testing to choose powers
$y = \beta_0 + \beta_1 x_1 + \beta_2 x^2 + \dots + \beta_k x^k + u$
$H_0: \beta_3 = 0$
$H_1: \beta_3 \neq 0$ } if we fail to reject ⇒ $y = \beta_0 + \beta_1 x_1 + \beta_2 x^2$
$H_0: \beta_2 = 0$
$H_1: \beta_2 \neq 0$

→ Nonlinear Regression Population Regression Func.
↳ General Ideas: If relation between Y + X is nonlinear if
1. Effect on Y of a Δ in X depends on value of X
→ marginal effect of X is not constant
2. A linear regression is mis-specified: the functional form is wrong
3. Estimator of the effect on Y of X is biased
→ in general, it isn't even right on average
4. The solution is to estimate a regression function that is nonlinear in X

↳ We have $\ln(x+\Delta x) - \ln(x) = \ln(1 + \Delta x / x) \cong \Delta x / x$ * $d \ln x / dx = 1/x$
→ when Δs are large, this rule does not follow well
↳ numerically: $\ln(1.01) = 0.00995 \cong 0.01$ + $\ln(1.1) = 0.0953 \cong 0.1$ (sort of)
↳ in transformation: $\beta = \ln(Y_N / Y_R) = \Delta y / y \Rightarrow e^{\beta} = 1$
→ when $\beta < 0.2$, the approximation is good

↳ How to know which formula to use?
→ see positive/negative relationship + decr/incr. returns

→ Interpretting Estimated Regression Functions
↳ Plot Predicted Values
→ ex: $\widehat{TestScore} = 607.3 + 3.85 inc. - 0.0423 inc.^2$
$(2.9) \quad (0.27) \quad (0.0048)$
] If we move from 5K to 6K
$\Delta \hat{TS} = 607.3 + 3.85(6) - 0.0423(6^2) - 607.3 + 3.85(5) - 0.0423(5^2) = 3.4$

For a 1K incr. in income, predicted Δ in TS is ↗

→ ex 2: $\ln(wage) = \hat{\beta_0} + \hat{\beta_1} edu. + \hat{u}$
* careful in finding $\ln(\hat{y})$ → we know $E(y) = E(e^{\hat{\beta_0}} + \hat{\beta_1} x e^u)$
usually $E(u) = 0$ but $E(e^u) \neq 1$ (will be slight bias)

Multiple Regression Models

$y = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u$
(vector notation)

→ Adding more variables for more controls
→ In simple regression models, some assumptions are violated ($cov(x,u) \neq 0$)
↳ often you care about one variable + the rest are controlled

II. Assumptions

* to prove Gaus-Markhov

MLR1: Linear in Parameters — relation between x + Y is linear (definition)
→ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + u$

MLR2: X is uncorrelated with u: $cov(x_i, u_i) = 0$
→ failure results in biased estimates of coefficients of explanatory variables

MLR3: Error term has zero expected value: $E(u_i) = 0$ for $i = 1, \dots, N$
→ failure results in biased estimate of constant term

MLR4: Homoskedasticity — error term has constant variance for all observations
→ $Var(u) = E(u^2) = \sigma_u^2$
→ failure results in inefficient estimates + biased tests of hypothesis

MLR 5: No autocorrelation – error terms are statistically independent
- → $cov(u_i, u_j) = 0$ for all $i \neq j$, $i,j = 1, \dots, N$
- → failure results in efficient estimates + biased test of hypothesis

MLR 6: No perfect collinearity – there is no exact linear relationship among the independent variables; explanatory variables are not perf. correlated
- → $\rho_{x_1 x_2} \neq \pm 1$.
- → failure results in inefficient estimates, impossibility of interpreting β_1 + biased test hypothesis (OLS becomes unreliable)
- → ex: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$, $corr(x_1, x_2) = 1$
  $y = \beta_0 + \beta_1(a + b x_2) + \beta_2 x_2 + u$
  $= \beta_0 + \beta_1 a + \beta_1 b x_2 + \beta_2 x_2 + u$
  $= \underbrace{(\beta_1 + \beta_1 a)}_{A} + \underbrace{(\beta_1 b + \beta_2)}_{B} x_2 + u$ } can be shown as a 2 var relation (to fix, get rid of x_1 or x_2)
  $y = A + B x_2 + u$
  OLS cannot be used to est β + interpretation of OLS coefficient will be impossible
  - ↳ case 1: $pcar = \beta_0 + \beta_1 \text{milage} + \beta_2 \text{condition} + u$
    can be correlated but $\rho_{x_1, x_2} \neq 1$
  - ↳ case 2: $GDP_t = \beta_0 + \beta_1 M_t + \beta_2 M_{t-1} + \beta_3 \Delta M_t + u$
    - → violates assumption b/c $\Delta M_t = M_t - M_{t-1}$
  - ↳ case 3: $unemploy = \beta_0 + \beta_1 (immigr) + \beta_2 (immigr)^2 + u$
    - → assume when immigr. first come in, employ. decr. but as more come, employ. incr.
    - → immigr + $(immigr)^2$ is non-linear so $corr(immig, immigr^2) \neq 1$
      ↳ use OLS!

MLR 7: Error terms are normally distributed
- → failure invalidates use of student t-distribution in coefficient t-test

*assump 8-12 same as SLR

III. Estimations

OLS: $\hat{u} = y - \hat{y} = y - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \dots - \hat{\beta}_k x_k$

SSR: $\sum_{i=1}^{n} \hat{u}_i^2$ ↙ minimize

→ FOC: Take k+1 derivatives $\frac{d\sum \hat{u}_i^2}{d\hat{\beta}_0} \rightarrow \frac{d\sum \hat{u}_i^2}{d\hat{\beta}_1} \rightarrow \frac{d\sum \hat{u}_i^2}{d\hat{\beta}_2}$ x_{2i} (2 = variable, i = observation)

where $\frac{d\sum \hat{u}_i^2}{d\hat{\beta}_k} = \sum_{i=1}^{n} x_{ki}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki}) = 0$

→ ex: Sample Regression Model $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{u}_i$
Residual $\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}$

1. minimize ss of error terms: $\min_{\text{wrt } \hat{\beta}_0 \hat{\beta}_1 \hat{\beta}_2} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \hat{u}_i)^2$
2. FOC:

$\frac{d\sum \hat{u}_i^2}{d\hat{\beta}_0} = -2\sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - u_i) = 0$
$\sum y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum x_{1i} + \hat{\beta}_2 \sum x_{2i}$
$n\hat{\beta}_0 = \sum y_i - \hat{\beta}_1 \sum x_{1i} - \hat{\beta}_2 \sum x_{2i}$
$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} - \hat{\beta}_2 \bar{x}_2$

$\frac{d\sum \hat{u}_i^2}{d\hat{\beta}_1} = -2\sum x_{1i}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - u_i) = 0$
$\sum y_i x_{1i} = \hat{\beta}_0 \sum x_{1i} + \hat{\beta}_1 \sum x_{1i}^2 + \hat{\beta}_2 x_{1i} x_{2i}$
$\hat{\beta}_1$ = see MLR Estimation on bcourses

$\frac{d\sum \hat{u}_i^2}{d\hat{\beta}_2} = -2\sum x_{2i}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - u_i) = 0$
$\sum y_i x_{2i} = \hat{\beta}_0 \sum x_{2i} + \hat{\beta}_1 \sum x_{1i} x_{2i} + \hat{\beta}_2 x_{2i}^2$
$\hat{\beta}_2$ = see MLR Est. on bcourses

*Note: If there were > 2 independent var. then normal eq can be generalized as
$\sum y_i x_{ki} = \hat{\beta}_0 \sum x_{ki} + \hat{\beta}_1 \sum x_{1i} x_{ki} + \hat{\beta}_2 \sum x_{1i} x_{ki} + \dots + \hat{\beta}_k x_{ki}^2$

→ Interpretation of Parameter Estimates

↳ ~~Partial effect / ceteris paribus~~ (holding all else constant)

↳ each x has a partial effect on y

$\hat{\beta}_0$ = predicted value of y when $x_1 + x_2 = 0$

$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2$

$\hat{\beta}_1$ = measures Δ in predicted value of y per unit Δ in x_1, holding all other variables constant

$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1$, holding x_2 fixed

↳ examples:

PR: $wage = \beta_0 + \beta_1 ed. + \beta_2 exp + \beta_3 ten$

If we expect incr. returns, then we can use a lin-log model where

$\ln(y) = \beta_0 + \beta_1 ed. + \beta_2 + e$

SRF: $\ln(\widehat{wage}) = 0.284 + 0.092 ed. + 0.0041 exp + 0.022 ten$ $n = 526$

(0.0104) $(—)$ $(—)$ $(—)$ $R^2 = 0.316$] std form

↳ one yr of ed. causes wage to incr. by 9.2% on average holding all

↳ else constant

$\Delta\ln(wage) = 0.0041\Delta exp + 0.022 ten\Delta = 0.026$

↳ What is $\widehat{wage}$ w/ ed = 12 yr, exp = 5 yr, tenure = 2 yr?

$\ln(\widehat{wage}) = 0.284 + 0.098(12) + 0.0041(5) + 0.022(2) = 1.425$

$e^{1.425} \approx 4.27$ (predicted hrly wage)

→ Comparing Simple + Multiple Regression Estimates

$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + u$

$y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 + \tilde{\beta}_3 x_3 + u$] $\hat{\beta}_0 \stackrel{?}{=} \tilde{\beta}_0$ when we add more variables could "change transform"? ⇒ $\hat{\beta}_0 \neq \tilde{\beta}_0$

* exceptions 1) $\hat{\beta}_2 = \hat{\beta}_3 = 0$ (no effect on Y)

2) x_1 is uncorrelated w/ $x_2 + x_3$ in the sample

↳ ~~omitted variable bias~~ (OVB):

when you omit an imp. variable from the data set

1. OV is a determinant of Y
2. OV is correlated w/ X

$\hat{\beta}_1 = \beta_1 + \frac{\Sigma(x_i - \bar{x})u_i}{\Sigma(x_i - \bar{x})^2}$

$= \beta_1 + \frac{\Sigma(x_i - \bar{x})(u_i - \bar{u})}{\Sigma(x_i - \bar{x})^2}$

$= \beta_1 + \frac{cov(x,u)}{var(x)} = \hat{\beta}_1$

$E(\hat{\beta}_1) = \beta_1 \Rightarrow corr(x,u) = 0$

True Model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$

$y = \beta_0 + \beta_1 x_1 + u^*$ $u^* = u + \beta_2 x_2$

If x_2 is correlated w/ x_1, then we have OVB + $cov(x, u^*) \neq 0$

$\hat{\beta}_1 = \beta_1 + \beta_j \frac{cov(x_1, x_j)}{var(x_j)}$

→ might have ± bias depending on the relationship

→ strongly related = more bias

IV. ~~Properties of MLR Estimators~~

→ ~~Algebraic Properties~~

1. $\Sigma_{i=1}^{n} \hat{u}_i = 0$
2. $\Sigma x_{ji}\hat{u}_i = 0$ for $j = 1, \ldots, k$
3. $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} + \hat{\beta}_2 \bar{x}_2 + \ldots + \hat{\beta}_k \bar{x}_k$ (line contains the mean)

→ ~~Statistical Properties~~ ↗ OLS

1. Unbiasedness: $E(\hat{\beta}_j) = \beta_j$ for $j = 0, 1, \ldots, k$
2. Var of OLS Estimators

$var(\hat{\beta}_j) = \frac{\sigma_u^2}{\Sigma(x_{ji} - \bar{x}_j)^2(1 - \rho_{x_1,x_2})}$ $j = 1, 2, \ldots$

↳ collinearity if $\rho_{x_1 x_2} = 1$

↳ $\hat{\beta}_{OLS}$ are efficient

3. Var of Error Term

MLR 5: $var(u) = c = \sigma_u^2$

$var(u) = \Sigma u_i^2 / n$ but u is unobservable ⇒ use residual

$\hat{\sigma}_u^2 = \frac{\Sigma \hat{u}_i^2}{n-k-1}$ d.f. $n - (k+1)$ # of est. parameters

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

$\hat{\sigma}_u^2 = \frac{\Sigma \hat{u}_i^2}{n-4}$

4. Gauss-Markhov

Given MLR 1-6, OLS estimators are BLUE in the class of all linear unbiased est.

V. Goodness of Fit: $y_i = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u_i$

$R^2 = \frac{ESS}{TSS}$ — y_i: TSS; $\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$: ESS (explained var); u_i: RSS

R^2 is the fraction of total Δ in y that is explained by the model

* cross sectional $R^2 \to 0$ + time series $R^2 \to 1$

Q: What happens to R^2 when we add more explanatory variables to the model?
A: R^2 will incr. unless $\beta = 0$ in which case it would not Δ
Thus R^2 is a poor tool in determining if a var. should be added [TSS = same, ESS = ↗, RSS = ↘]

Adjusted R^2 ($\bar{R}^2$) $= 1 - \frac{RSS}{TSS} \cdot \frac{n-1}{n-k-1} = 1 - \frac{\hat{\sigma}_u^2}{\hat{\sigma}_y^2}$ where $\sigma_y^2 = \frac{n-1}{TSS}$ + $\sigma_u^2 = \frac{RSS}{n-k-1}$

↳ when you add a regressor:
1. RSS ↓ ⇒ $\bar{R}^2$ ↑
2. $\frac{n-1}{n-k-1}$ ↑ ⇒ $\bar{R}^2$ ↓

] if $\bar{R}^2$ ↑ or ↓ depends on strength of effect

↳ tells you the predictive power of the model + goodness of fit
↳ R^2 or $\bar{R}^2$ does not tell you whether:
1. An included var is statistically significant
   $\bar{R}^2$ ↑ $|t| > 1$, statistical significance $|t| > 2 \approx 1.96$
2. The regressors are the true cause of movements in the dependent var
   * implies correlation + not causation
3. You have chosen the most appropriate regressor

$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1}$ ⇒
1. $\bar{R}^2 < R^2$
2. As n incr, $\bar{R}^2$ approaches R^2
3. As n + k get closer, $\bar{R}^2 < 0$

* can also use the SE of the regression

Example: $\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 exp$.
1) $\ln(wage) = 0.21685 + 0.09794 educ + 0.01035 exp$.
   on ave, an ↑ 1 yr of educ. will cause wage to incr. by 9.794%.
2) $\ln(wage) = 0.284 + 0.092 educ + 0.0041 exp + 0.22 tenure$, $R^2 = 0.316$ + $\bar{R}^2 = 0.312$
   → why did the exp coefficient Δ so much?
   ↳ exp. + tenure is correlated but tenure is more imp. than exp.
   ↳ OVB example
   ↳ corr(exp, ten) = ↑, corr(ten, educ) = ⟷

Hypothesis Testing in a Multiple Regression Model

I. Introduction
→ Assume homoskedasticity $u_i \sim N(0, \sigma_u^2)$ ⇒ from assump. $\hat{\beta}_i \sim N(\beta_j, \hat{\sigma}^2_{\hat{\beta}_i})$
→ T test: $t = \frac{\hat{\beta} - \beta^*}{se(\hat{\beta})} \sim t_{n-k-1}$

II. Testing the Significance of Individual Regression Coefficients
ex: $\ln(wage) = 0.2 + \underset{(0.007)}{0.092} educ + \underset{(0.0017)}{0.0041} exp + \underset{(0.003)}{0.022} ten + u$, $n = 510$

1. Assume null is true
   $H_0: \beta_2 = 0$
   $H_1: \beta_2 > 0$
   ] work exp prior to current job (var includes tenure) has no effect on the % Δ in wage
2. $t = \frac{\hat{\beta}_2 - \beta_2^*}{se(\hat{\beta}_2)} = \frac{0.0041}{0.0017} = 2.4$
3. critical value: $t_{0.05, 506} = 1.648$
4. Rejection Rule: $|t| > t_{\alpha, n-k-1}$
5. Reject H_0 (2.4 > 1.648): exp is important in determining / influencing the wage, holding all variables constant

* if $\beta = 1$, a yr of exp incr. wage by 100%

III. Testing Joint Hypothesis (F-Test)

$H_0: \beta_1 = \beta_1^*, \beta_2 = \beta_2^*, \ldots, \beta_k = \beta_k^*$
H_1: H_0 is not true (one restriction does not hold)

→ **Joint Significance**

$H_0: \beta_1 = \beta_2 = \cdots = 0$
H_1: At least one is not true

| Alt
| $H_0: R^2 = 0$
| $H_1: R^2 \neq 0$

If you use a t-test:
$H_0: \beta_1 = \beta_2 = 0$ + $\alpha = 5\%$
Reject if $|t_1| < 1.96$ + $|t_2| < 1.96$
Assuming x_1 + x_2 are independent
$P(|t_1| < 1.96 \text{ and } |t_2| < 1.96)$
$= (0.95)(0.95) = 0.905$
$P(\text{reject}) = 1 - 0.905 = 0.095$
* Power of test lowers
→ more var = lose more power

F statistic: looks at how much RSS drops if H_0 is true
↳ If we drop a var from the eq/regression, then RSS incr.
↳ compares restricted (under H_0: $y = \beta_0 + u$) + unrestricted ($y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$) models by $F = \dfrac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k-1)}$ where

RSS_R: residual sum of squares from restricted model
$n-k-1$: d.f. for RSS_{UR}
q: numerator d.f. $= df_R - df_{UR}$
↳ # of excluded variables
↳ # of restrictions
* numerator always expected to be positive so $F > 0$ because $RSS_R - RSS_{UR} > 0$
* always 2-sided $F \sim F(q, n-k-1)$

reject when $F > F_{\alpha,(q, n-k-1)}$

Different Types of Hypothesis:
1. **Overall Significance**
$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ or $H_0: R^2 = 0$
2. **Partial F-Test**
$H_0: \beta_1 = \beta_2 = 0$
3. **Linear Combinations**
$H_0: \beta_1 + \beta_2 = 1$ or
$H_0: \beta_1 = \beta_2$
* cannot test $\beta_1^2 = \beta_2$ (nonlinear)

low power w/ t-test; use F-stat which compares original reg. + reg where H_0 is true (see if drop in RSS is big enough to see if var are imp)

Steps for **F-Stat Test**:
1. Run an unrestricted regression
→ estimate $\hat\beta_1 \ldots \hat\beta_k$ + find $RSS_{UR} = \sum_{i=1}^{n}(y_i - \hat y_i)^2$
2. Run a restricted regression
→ estimate model assuming H_0 is true + find $RSS_R = \sum_{i=1}^{n}(y_i - \hat y_{iR})^2$
3. Calculate the F-statistic
→ $F = \dfrac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k-1)}$
4. Find critical value w/ d.f. $= n-k-1$, q, + significance level
5. Rejection rule + conclusion: $F > F_{\alpha,(q, n-k-1)}$
→ reject H_0: variables + are jointly statistically significant

rejection region

* For **overall significance** → $H_0: \beta_1 = \cdots = \beta_k$
↳ can only run $F = \dfrac{R^2/k}{(1-R^2)/(n-k-1)} = \dfrac{ESS/k}{RSS/n-k-1}$ where k = # of var in original regression

Example: $\ln(wage) = 0.284 + 0.092 ed + 0.0041 exp + 0.022 ten$
→ $H_0: \beta_1 = \beta_2 = \beta_3 = 0$
H_1: at least one $\beta \neq 0$
→ $F = \dfrac{R^2/q}{1-R^2/n-k-1} = \dfrac{0.3161/3}{(1-0.316)/(526-3-1)} = 80.38$
$F = \dfrac{ESS/k}{RSS/n-k-1} = \dfrac{46.874/3}{101.456/522} = 80.4$
same
→ critical value: $F_{\alpha,-} = 2.62$
→ $80.38 > 2.62$ ⇒ reject H_0, var are jointly significant

* on excel: F + signif. F = P value for F-test

Example 2: $\ln(salary) = \beta_0 + \beta_1 yrs + \beta_2 gamesyr + \beta_3 bavg + \beta_4 hrunyr + \beta_5 bisyr$

MLB salary ↗ $= 11.192 + 0.0689yr + 0.0126gamesyr + 0.00098bavg + 0.0144hruns + 0.0108bisyr$

For every yr he plays in MLB, salary is expected to incr. by 6.89% holding all other variables constant

→ Partial Test: $H_0: \beta_{k-m} = \dots = \beta_k = 0$ w/m restrictions] test to see if bavg, hrun, + bis is significant jointly
H_1: H_0 is not true

1. Run UR Regression: $RSS_{UR} = 183.186$
2. Run R Regression: $RSS_R = 198.3$
   ↳ $\ln(salary) = 11.224 + 0.0713yr + 0.0202games$
3. Find the F-stat: $F = \dfrac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k-1)}$
   $= \dfrac{(198.3 - 183.186)/3}{183.186/(353-5-1)} = 9.5$
4. Critical value: $\alpha = 5\%$, $df = (3, 347)$ → $F_{0.05,(3,347)} = 2.63$
5. Rejection rule: $F > F_c$, $9.5 > 2.63$ → reject H_0 = bavg, hrun, bis are jointly stat signif.

* these 3 var are correlated which may cause var ↑ (overlap in effect) so they look insignificant separately (high p value) but testing shows they are jointly significant
↳ $var(\hat{\beta}) = \dfrac{\sigma_u^2}{(1-corr(x_1, x_2))}$

Testing Linear Combinations:

ex: $C_t = \beta_0 + \beta_1 W_t + \beta_2 P_t + u_t$
where W_t = total wage income
P_t = all other income
β_1 = mpc from wage income
β_2 = mpc from all other income

$H_0: \beta_1 = \beta_2$ → $q = 1$
$H_1: \beta_1 \neq \beta_2$

$C_t = \beta_0 + \beta_1 W_t + \beta_2 P_t + u_t$
$C_t = \beta_0 + \beta_1 W_t + \beta_1 P_t + u_t$
$C_t = \beta_1 + \beta_1 Z_t + u_t$ where $Z_t = W_t + P_t$

Confidence Intervals

* same concept $\hat{\beta} \pm t_{\alpha, n-k-1} \cdot se(\beta)$

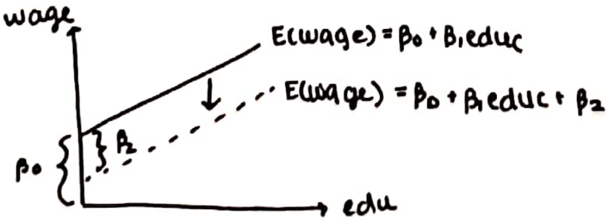
Dummy Variable Regression

1. Intro

→ dummy variable: binary variable
↳ whatever is 0 is the base for benchmark group
↳ name is category 1

* if dependent, it proximates probability

→ ex: $wage = \beta_0 + \beta_1 educ + \beta_2 female + u_i$ ⇒ β_2: difference in hourly wage between females + males given same amt of educ. + error
↳ male: $E(wage \mid female = 0) = \beta_0 + \beta_1 educ$
↳ female: $E(wage \mid female = 1) = \beta_0 + \beta_1 educ + \beta_2$



* cannot have male + female dummy var = perfect collinearity

→ If a model w/intercept β_0 has a qualitative variable w/m categories, then we can introduce up to m-1 dummy variables

$GPA = \beta_0 + \beta_1 study + \beta_2 IQ + \beta_3 Fresh + \beta_4 Soph + \beta_5 Junior + \cancel{\beta_6 Senior}$

diff. in grades b/t fresh + senior holding all other constant

dummy var trap!

II. Interactions Using Dummy Variables

→ **interaction variable**: independent variable that is the multiple of 2 or more other independent variable

↳ ex: $\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 female + \beta_3 nonwhite + \beta_4 female\,nonwhite$

→ $E(Y \mid female=0, nonwhite=0) = \beta_0$ | white male

→ $E(Y \mid \text{——} = 1, \text{——}) = \beta_0 + \beta_2$ | white female

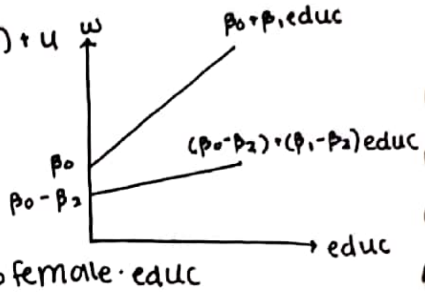
→ $E(Y \mid \text{——} = 0, \text{——} = 1) = \beta_0 + \beta_3$ | nonwhite male

→ $E(Y \mid \text{——} = 1, \text{——} = 1) = \beta_0 + \beta_2 + \beta_3 + \beta_4$

↳ effect nonwhite female relative to white male

↳ ex: $\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 female + \beta_3 (educ)(female) + u$

return to educ for men ↗ main effect ↗ ↖ interaction effect



→ $E(y \mid female=0)$: $\ln(wage) = \beta_0 + \beta_1 educ + u$

→ $E(y \mid female=1)$: $\ln(wage) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) educ$.

* gap incr. w/higher level of educ

$\ln(wage) = 0.826 + 0.077 educ - 0.36 female - 0.00006 female \cdot educ$

→ the female coefficient in this ex. + the previous is the same

↳ effect of interaction variable is small

↳ cannot use t-test on this variable b/c of interaction var

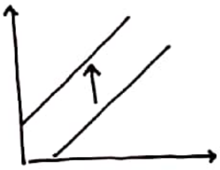
→ cannot hold other variables constant

↳ significance of female variable drop b/c se(female) incr. from high correlation between interaction variable

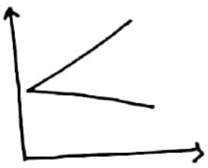
→ $se(\hat\beta) = \sqrt{\dfrac{\sigma_{\hat u}^2}{1 - corr\, x_2 x_3}}$

↳ difference of $\hat\beta$ between categories of variables / dummy

main effects
DV for Y



interactions
2 DV
DV + contin.



III. Applications

1. ~~Event Studies~~

→ ex: $Delta = \beta_0 + \beta_2 Sep11 + \beta_2 control$ where close dates after Sep 11 = 1

2. **Seasonality** (need seasonally adjusted data)

→ ex: time series w/excess return = return − R_f

↳ price of risk (return in investing in riskier assets)

* If you don't have an intercept, you can include all dummy variables

3. **Fixed Effects Model**

→ using dummy variables for panel data + see dynamics of reaction w/time

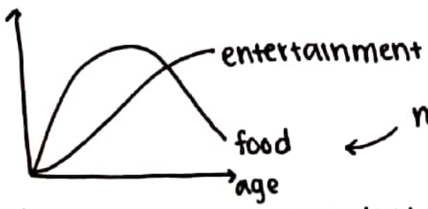
↳ $y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_k X_{kit} + u_{it}$

→ i = entity / state for cross-sectional data

→ t = time + 1…K = #

→ **Panel Data** (ex: substitution effect for entertainment + food consump)

* need to look at the same person + expenditures over time



making more food at home when older

↳ Panel Data can control for factors that

1. vary across entities but not across time
2. could cause OVB if they are omitted
3. are unobserved or unmeasured

→ cannot be included in the regression (ex: characteristics, culture)

* If omitted variable does not Δ over time, then any Δs in y over time cannot be caused by the omitted variable

→ How to run the regression?

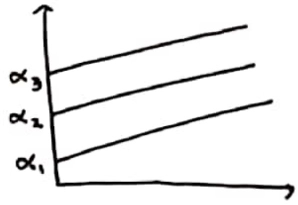
↳ Pull the data

↳ Entity fixed effects: $y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 FE_i + u_{it}$

no subscript t b/c it is fixed over time; creates a supplemental intercept

$y_{it} = \alpha_i + \beta_1 x_{it} + u_{it}$

$\alpha_i = \beta_0 + \beta_2 FE_i$



or $y_{it} = \beta_0 + \beta_1 x_{it} + \gamma_2 D2_i + \gamma_3 D3_i + \cdots + \gamma_k Dk_i + u_{it}$

→ how to test presence of fixed effects of this model?

↳ $H_0: \gamma_2 = \cdots = \gamma_k = 0$ w/ F-test

↳ if only one entity, do not need fixed effects

↳ if you fail to reject → pool the data

* every person has a different intercept

→ Estimation of Fixed Effects Regression: Estimation Methods

1. "n-1" Binary Regressor

★ $y_{it} = \beta_0 + \beta_1 x_{it} + \gamma_2 D2_i + \cdots + \gamma_u Dn_i + u_{it}$ where $D2_i = \begin{cases} 1 & \text{if } i = 2 \\ 0 & \text{otherwise} \end{cases}$

Steps: 1) Generate dummy variables $D2_i \ldots Dn_i$

2) Estimate ★ by OLS

↳ impractical when n (# of entities) is large

→ as you incr. variables, your df (n-k-1) goes down so it may make OLS unfeasible

→ occurs when there are a lot of cross sections

2. Entity Demeaned OLS Regression

★★ $\tilde{y}_{it} = \beta_1 \tilde{x}_{it} + \tilde{u}_{it}$ where $\tilde{y}_{it} = y_{it} - \frac{\sum_{t=1}^{T} y_{it}}{T}$

$(y_{it} - \bar{y}_{it}) = \underbrace{(\alpha_i - \bar{\alpha})}_{0} + \beta_1 (x_{it} - \bar{x}) + (u_{it} - \bar{u}_i)$

* subtract fixed effects by subtracting entity means

* fix person + view average consumption across time

Steps: 1) Generate entity demeaned variables: $\tilde{y}_{it}$ + $\tilde{x}_{it}$

2) Estimate ★★ by OLS

3) Calculate standard errors in a way that accounts for panel data (".xtset state year")

* standard errors: can make data no longer iid → $corr(u_{i1}, u_{i2}) \neq 0$ ($t=1$, $t=2$)

↳ violates autocorrelation (MLR5)

↳ ex: past behavior is usually correlated w/ current

3. "Before + After"/Differences Specification

↳ only works when there are two time periods

* in stata: fixed effects "xreg", use option "fe" (fixed effects) → vce(cluster-state)

→ Ex: Prison population size on crime rates (14 time periods + 50 states)

↳ state specific effects (gretl: model → panel → fixed or random)

1. $\ln(crime) = \beta_0 + \beta_1 prison + u$

→ test $H_0: \beta_1 < 0$

2. $\ln(crime) = \beta_0 + \beta_1 prison + \beta_2 (age\ 18\text{-}24) + \beta_3 metro + \beta_4 police + \beta_5 unemploy + u$

→ when adding state specific (entity) + fixed effects, the signific. of prison goes away

↳ may be due to lagged effect (person's crime may be recorded in the same time period they enter prison)

→ Time Fixed Effects

↳ entities affected by one time shift

$y_{it} = \underbrace{\lambda_t}_{\text{time fixed effect}} + \beta_1 x_{it} + u_{it}$

* must think about

→ degrees of freedom

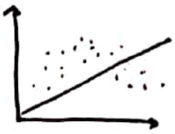
→ multicollinearity problems

IV. Violating the MLR Assumptions

→ MLR 1-6 used to prove Gaus-Markhov (BLUE estimators)
↳ we will look at consequences, detection, + solutions

1. Violation of Linearity in Variables (MLR1)

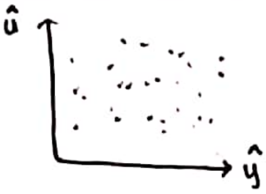
↳ consequences of functional form misspecification →



→ slope (β_2) + intercept (β_1) are biased
↳ ex: if you fit linear in nonlinear models, you will have ↑ errors
↳ wrong predictions

↳ **Detection**

→ Informal: look at plots of data



$\hat{u}$ vs. $\hat{y}$
* want random

y vs. $\hat{y}$

→ **Formal**

1. Choosing between lin-lin + log-log models $\underbrace{\text{(Makinnon, White, Davidson)}}_{\text{MWD}}$

H_0: linear model
H_1: log-log model (nonlinear)

Steps
1) Estimate linear model: $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$
   → get fitted values: $y_f = \hat{y}$
2) Estimate log-log model: $\ln(y) = \beta_0 + \beta_1 \ln(x_1) + \cdots + \beta_k \ln(x_k) + u_i$
   → get fitted values: $\ln f = \widehat{\ln}(y)$
3) Generate $z_1 = \ln(y_f) - \ln f$
   → if linear, the difference should be zero
4) Regress + run: $y = \beta_0 + \beta_1 x_1 + \beta_k x_k + \alpha_1 z_1 + u_i$
   → reject H_0 if α_1 is statistically significant
5) Generate $z_2 = \text{antilog}(\ln f) - y^f$ ← linear specification
6) Run $\ln(y) = \beta_0 + \beta_1 \ln(x_1) + \cdots + \beta_k \ln(x_k) + \alpha_2 z_2 + u$
   → test α_2 + reject nonlinearity (H_1) if α_2 is stat signif.

2. **RESET** (Regression Specification Error Test) – func. form test

Steps
1) $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \Rightarrow$ get $\hat{y}$
2) $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + u$
   → $H_0: \delta_1 = \delta_2 = 0 \Rightarrow$ model is linear
   ↳ if reject, there is some nonlinearity
3) Use the F test: $F \sim F_{2, n-k-3}$

↳ **How to fix:** apply a nonlinear transformation

2. Correlation of X w/ u (MLR2)

↳ can result from omitted variable bias

3. Expected value of the error is nonzero (MLR 3)

↳ $E(u) \neq 0 \rightarrow E(u) = \bar{u}$
$y = \beta_0 + \beta_1 x_1 + u \rightarrow E(y) = \beta_0 + \beta_1 x_1 + \bar{u}$

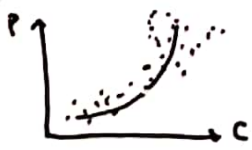
* results in biased intercepts

4. Heteroskedasticity (MLR 4)

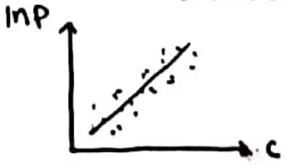
homosk
$\text{var}(u) = \sigma^2$

hetero
* usually dealt w/in cross-sectional data
* can be caused by outlier or model misspecification (OVB, or non-linearity)

ex: Diamond Price v. Carat



→ incr. returns to scale
→ log-lin



↳ **consequences** of violating homoskedasticity

→ we use homosk. assump because it is easier to est. one point

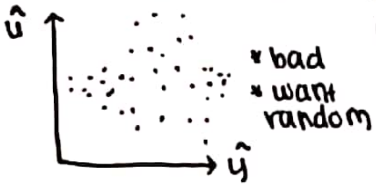
	homo	heter
x_1	σ^2	σ_1^2
x_2	σ^2	σ_2^2
x_3	σ^2	σ_3^2

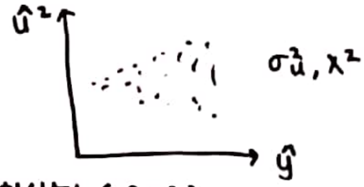
→ $var(\hat{\beta}) = \frac{\sigma_u^2}{n_x \sigma_x^2}$ → $se(\hat{\beta}) = \sqrt{var}$ → t-test, F-test, confidence interval } unreliable

→ does not affect OLS of β_0, β_1. but $\hat{\beta}$ is not efficient + tests are biased *no best in BLUE

↳ **How to detect**

→ informal: plots of ($\hat{u}$ vs. $\hat{y}$) or ($\hat{u}^2$ vs $\hat{y}$) or ($\hat{u}^2$ vs x_i)





→ formal: **Breush-Pagan Test** for Heteroskedasticity (1979)

↳ $H_0: Var(u) = E(u^2) = \sigma^2 = \text{constant}$

*tests if residuals are related to x $\quad H_1: \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$ or $R^2 = 0$] joint hypothesis for overall sign

↳ $u^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + v$ but u^2 is unobserved

$\hat{u}^2 =$ ______________________

↳ use $F = \frac{R^2_{\hat{u}^2}/k}{(1-R^2_{\hat{u}^2})/n-k-1} \sim F_{k,n-k-1}$

↳ or use: Koenker (1981) LM Statistic w/ chi sq. distribution

$LM = n \cdot R^2_{\hat{u}^2} \sim \chi^2_k$

↳ **How to fix**

1. Calculate variances differently ⇒ use robust variance
   → White's variance: $Var(\hat{\beta}_j) = \Sigma \hat{r}_{ij}^2 \hat{u}_i^2 / RSS_j$
   ↳ $\hat{r}_{ij}^2$: ith residual from regressing x_i on all other independent var
   ↳ RSS_j: RSS from ↑

   *robust variable/variance not the most reliable)
2. Change Estimation Procedures
3. Transform Data
   → **Weighted Least Squares method**: depends on whether you know what type of heteroskedastic (homosk = $var(u) = \sigma_u^2$)
   A. **Known type** of heteroskedasticity
   → $var(u) = \sigma_u^2 \cdot h(x)$ where $h(x)$ is known
   B. **Unknown type** of heteroskedasticity

A. Known Type

→ $var(u) = \sigma_u^2 h(x)$ where $h(x)$ is known

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

→ divide both sides by $\sqrt{h(x)}$

$$\frac{y}{\sqrt{h(x)}} = \frac{\beta_0}{\sqrt{h(x)}} + \beta_1 \frac{x_1}{\sqrt{h(x)}} + \beta_2 \frac{x_2}{\sqrt{h(x)}} + \frac{u}{\sqrt{h(x)}}$$

$$y^* = \beta_0^* + \beta_1 x_1^* + \beta_2 x_2^* + u^*$$

→ parameters are the same but data Δ

$$var(u^*) = var\left(\frac{u}{\sqrt{h(x)}}\right) = \frac{1}{\sqrt{h(x)}} var(u)$$

$$= \frac{\sigma_u^2 h(x)}{h(x)} = \sigma_u^2 = \text{homoskedastic}$$

* var of WLS is more efficient
* small data = robust not good

B. Heteroskedasticity of Unknown Form

→ $h(x)$ is partially known

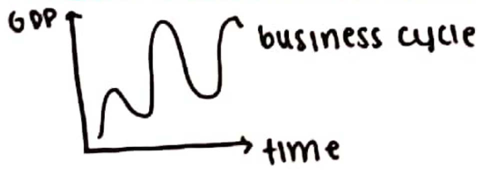
$\sigma_u^2 = var(u) = \sigma^2 e^{\delta + \delta_1 x_1 + \dots + \delta_k x_k}$

↳ always positive so good

→ Test: estimate unknown $\delta_0, \delta_1, \dots \delta_k$

→ steps
1. Regress y on x + obtain residuals $\hat{u}_i$
2. Run $\ln(\hat{u}^2) = \delta_0 + \delta_1 x_1 + \dots + \delta_k x_k + v$
   Get predicted values → $\hat{g}_i$
   ⇒ estimates of $\hat{h(x)} = e^{\hat{g}_i}$
3. Weigh data by $1/\sqrt{h(x)}$

5. **Autocorrelation / Serial correlation (MLR5)**



→ no autocorrelation: $cov(u_i, u_j) = 0$

↳ TS: correlating error term from one period to another

↳ but $corr(u_t, u_{t-j}) \neq 0$

→ usually due to time series due to momentum

↳ $corr(u_t, u_{t-j}) \neq 0$

$y_t = \beta_0 + \beta_1 x_t + u_t$, where u_t depends on past values — AR(1)

$u_t = \rho u_{t-1} + \varepsilon_t$, where ρ = autocorrelation coefficient — ↳ lag 1 → u_{t-1}

whitenoise ↗ $-1 < \rho < 1$ but we want $\rho = 0$

→ why does it occur?

↳ OVB, time series, or mispecification
↳ lack of lags create autocorrelation

] $consump = \beta_0 + \beta_1 income + \beta_2 consump_{t-1}$
*include lagged consump due to habit formation

→ **consequences**

↳ take "Best" out of BLUE (inefficient estimates + biased tests)

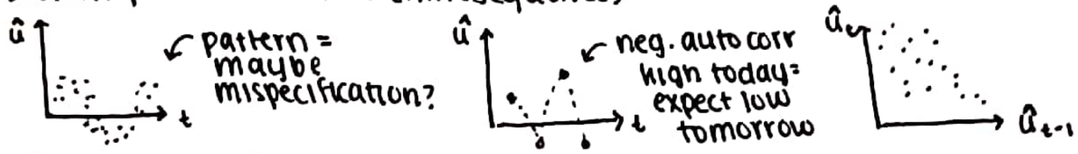
→ no efficiency due to problems w/SE

$var(\hat\beta) = var(\beta) + \underline{\quad \nearrow 0}$ when $cov(u_i, u_j) = 0$

1. $\hat\sigma_u^2$ is likely to underestimate true σ_u^2
2. t, F, CI are unreliable + invalid
3. Autocorrelation might make t-stat look higher

→ **How to detect?**

1. Look at plots: $\hat u$ vs. time (time sequence) or $\hat u_t$ vs. $\hat u_{t-1}$



↳ autocorrelation plot

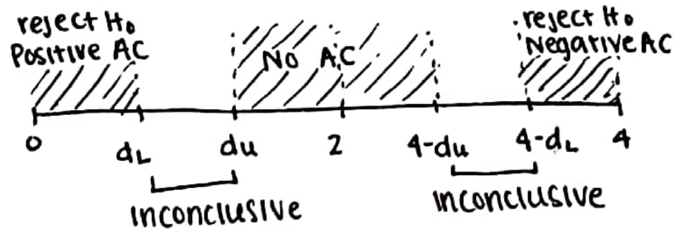
2. **Durbin Watson Test**

↳ test specific residual autocorrelation of time period 1

↳ Assumptions: 1) $u \sim AR(1)$ – error term today depends on error term yesterday
2) no intercept
3) β_0
4) original regression has no lagged terms (no y_{t-i})

$y = \beta_0 + \beta_1 x + u$

1. Run the regression + get the residual ($\hat u_i$) ↙ lagged AR(1)
2. Durbin Watson: $d = \frac{\Sigma(\hat u_t - \hat u_{t-1})^2}{\Sigma(\hat u_t)^2} \approx 2(1-\rho)$ + $0 < d < 4$
   ↳ no auto corr: $d \approx 2$
   ↳ different table
3. Breusch-Pagan: $u_t = \varphi_1 \hat u_{t-1} + \varphi_2 \hat u_{t-2} + \cdots + \varphi_p \hat u_{t-p} + \varepsilon_t$
   $(T-P)R^2_{\hat u^2} \sim \chi^2_p$ (chi sq distribution)



→ **How to fix?**

1. HAC robust variance
2. **Generalized Least Squares**

$y_t = \beta_0 + \beta_1 x_{1t} + u_t$ (★) → $u_t = \rho u_{t-1} + \varepsilon_t$ - autocorr

Difference term:

1) $y_{t-1} = \beta_0 + \beta_1 x_{1t-1} + u_{t-1}$
2) multiply by ρ: $\rho y_{t-1} = \rho\beta_0 + \rho\beta_1 x_{1t-1} + \rho u_{t-1}$ (★★)
3) subtract: $\underbrace{y_t - \rho y_{t-1}}_{y^*} = \underbrace{\beta_0(1-\rho)}_{\beta_0^*} + \underbrace{\beta_1(x_{1t} - x_{1t-1}\rho)}_{\beta_1 x_1^*} + \underbrace{u_t - \rho u_{t-1}}_{u^*}$ (★-★★)
4) Estimate $y^* = \beta_0^* + \beta_1 x_1^* + u^*$ using OLS

6. Perfect Collinearity (MLR6)

→ consequence: inefficient estimators, impossibility of interpretting β_1, + biased hypothesis testing

→ Detect: collinearity: $x_1 = a + bx_2$ ← cannot detect w/cross correl.

multicollinearity: $x_1 = a + bx_2 + bx_3$ → high R^2 but low t-stat

→ Fix: drop or transform (x_1/x_2)

V. Instrumental variables

MLR2: $Cov(x_i, u_i) = 0$ ⇒ biased estimates

ex: job training on wage overestimated b/c of OVB (motivation?)

Causes of $cov(x,u) \neq 0$

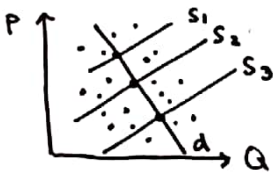
1. OVB: (1) determinant of y + corr w/explanatory variable
2. Simultaneous causality bias ($x \to y, y \to x$)

$y = \beta_0 + \beta_1 x + u$ and $x = \alpha_0 + \alpha_1 y + \tilde{u}$ * IV allows us to infer causality

$u > 0 \Rightarrow y\uparrow \Rightarrow x\uparrow \Rightarrow cov(x,u) \neq 0$

3. measurement errors in x: $y = \beta_0 + \beta_1(x + \varepsilon_1) + u$ $\quad \frac{dy}{dx} = \beta_1 + \frac{du}{dx}] \neq 0$

Wright: instrumental variable



D: $Q^d = a + bP + u^d$ ← $corr(P, z) \neq 0$

S: $Q^s = c + dP + u^s$

factors that shift supply are instrumental var.

→ affect demand indirectly

ex: $wage = \beta_0 + \beta_1 educ + u_i$ ov = ability

$= \beta_0 + \beta_1 educ + u(x)$ ↳ creates $cov(x,u) \neq 0$

$corr(educ, ability) \neq 0$

$\frac{dy}{dx} = \beta_0 + \frac{du}{dx}$ * want to separate educ → $cov(educ, u) = 0$ / → $cov(educ, u) \neq 0$

↑ ↑ increasing

z_i is the instrument

Two conditions for valid instruments

1. Relevance: $corr(z_i, x_i) \neq 0$
   → only channel instrument can impact y (no direct effect)
2. Exogeneity: $corr(z_i, u_i) = 0$

[Endogenous: variables corr w/u

Exogenous: " uncorr w/u

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

↑ endo ↑ exo ↑ exo ↑ endo]

* Δs in instrument z are unrelated to Δs in u but related to movements in x

↳ have to break the link [$x \to y$] ($z \nearrow$, $\searrow u \nearrow$)

Two Stage Least Square Estimator for SLR

Stage 1 | Find effect of z_i on x_i

→ isolate part of x that is corr w/u by regressing x on z using OLS

$x_i = \underbrace{\pi_0 + \pi_2 z_i}_{\text{uncorrelated w/error term}} + \underbrace{v_i}_{\text{correlated w/error term}}$ ← error term

keep ⇐ uncorrelated w/error term; correlated w/error term ⇒ problematic

→ create fitted values / calculate variation in x_i explained by z_i

$\hat{x_i} = \hat{\pi_0} + \hat{\pi_1} z_i$ { $H_0: \pi_1 = 0$; reject = condition is satisfied

Stage 2 | Estimate causal effect of x_i on y_i * isolates part of variation in x uncorr w/u

→ Replace x_i by $\hat{x}_i$ in the regression of interest using OLS

$y_i = \beta_0 + \beta_1 \hat{x_i} + u_i$ (regress y on $\hat{x}$ by OLS)

→ resulting estimator: $\hat{\beta_1}^{2SLS} = \hat{\beta_1}^{IV} = \hat{\beta_1}^{TLS}$ * R^2 not useful any more

Formula for IV Estimator

→ ex: $\Delta z = 1 \to \Delta x = 0.5 \to \Delta y = \$500$

one unit Δ in z ⇒ ↑ $\Delta x = 0.5$ ⇒ incr. in wage \$500

$\Delta x = 1 \Rightarrow \Delta y = \$1000$

We have $\frac{dx}{dz} = 0.5$, $\frac{dy}{dz} = 500$, $\beta_1 = \frac{dy/dz}{dx/dz}$

Scanned by CamScanner

→ Direct Algebraic Derivation : $y = \beta_0 + \beta_1 x_i + u_i$

↳ want to find $cov(y, z) = cov(\beta_0 + \beta_1 x_i + u_i, z)$

$= cov(\beta_0, z) + cov(\beta_1 x_1, z) + cov(u_i, z)$

$= 0 + \beta_1 cov(x_1, z) + 0$ ⟶ by assumption

↳ we have $\beta_1 = \frac{cov(y,z)}{cov(x,z)}$ + if $x = z$ then $\beta_1 = \frac{cov(y,x)}{var(x)}$

Properties

1. $\hat{\beta}_1^{IV} = \frac{\hat{cov}(y,z)}{\hat{cov}(x,z)} \xrightarrow{plim} \beta_1^{IV}$ ⇒ consistent ($plim_{n\to\infty} \hat{\beta}^{2SLS} = \beta$)
2. $\hat{\beta}^{IV} \sim N(\beta_1, \sigma^2_{\hat{\beta}^{IV}})$ ⇒ normal distribution for hypothesis testing
3. $Avar(\hat{\beta}_1^{IV}) = \frac{\sigma_u^2}{n\sigma_x^2 S^2_{x,z}}$ ⇒ asymptotic variance

Asymptotically normally distributed when n gets large then approx $\hat{\beta}^{TSLS} \sim N(\beta, var(\hat{\beta}^{TSLS}))$

↳ $S^2_{x,z} < 1 \Rightarrow var(\hat{\beta}^{IV}) > var(\hat{\beta}^{OLS})$

→ if $x = z$ then it will be the same OLS as before

→ bad, but OLS will give you biased estimators

want to be smaller to have more precision

* all assumes that the instruments are valid

Example 1: $ln(wage) = \beta_0 + \beta_1 educ + u$ ← ability $cov(educ, ab) \neq 0$

→ candidates

1. Card (1995): proximity to college
2. Angrist + Krueger (1991): month of birth / quarter
3. Number of siblings

↳ Stage 1: Run $educ = \pi_0 + \pi_1 sibs + v$

→ p value of sibs to see $corr(z, educ) \neq 0$

→ $\hat{Educ} = 14.128 - 0.227 sibs$

↳ Stage 2: $ln(wage) = \beta_0 + \beta_1 \hat{educ} + u$

$ln(wage) = 5.1 + 0.122\, \hat{educ}$ ← double from original regression

* use a program instead of in stages to get more accurate SE (like vce (cluster) for panel)

Example 2: Prison pop size on crime rate (Levitt, 1996)

$ln(crim) = \beta_0 + \beta_1 pris + u$

→ issue of simultaneity (diff to isolate causal effect)

→ instrument: status of state prison overcrowding litigation

1. $corr(z, pris) \neq 0$: ligitation usually ↓ growth of prison pop
2. $cov(z, u) = 0$

Review

1. Find instrument, z_i
   - → instrument relevance: $cov(x_i, z_i) \neq 0$
   - → instrument exogeneity: $cov(z_i, u_i) = 0$
2. First Stage: effect of z_i on x_i
   - → OLS: $x_i = \pi_0 + \pi_1 z_i + v_i$
   - → calculate variation in x_i explained by z_i ($\hat{x}_i$)
3. Second Stage: estimate causal effect of x_i on y_i
   - → OLS: $y_i = \beta_0 + \beta_1 \hat{x}_i + u_i$
   - → $\hat{\beta}_0 + \hat{\beta}_1$ are consistent + asymptotically normal

→ Property: variance

$avar(\hat{\beta}^{IV}) = \frac{\sigma^2}{n\sigma_x^2 corr(x,z)^2} > var(\hat{\beta}^{OLS}) = \frac{\sigma^2}{n\sigma_x^2}$

↳ larger variance but better test

if $corr(x,z) = 0.01$ then $var(\hat{\beta}^{IV}) = 10000 var(\hat{\beta}^{OLS})$

——— 0.1 ——— $100 var(\hat{\beta}^{OLS})$

——— 0.5 ——— $4 var(\hat{\beta}^{OLS})$

*want a strongly correlated instrument

Formula for IV Estimator

$\beta_1^{IV} = \frac{cov(Y, Z)}{cox(X, Z)}$

ex: $wage = \beta_0 + \beta_1 educ + u$

# of siblings → educ → wage
(z) (x) (y)

effect of sib # on wage = effect of sib # on educ × effect of educ on wage

effect of educ on wage = β_1^{IV} = $\frac{\text{effect of sib \# on wage}}{\text{on educ}}$

$= \frac{cov(Y,Z)}{cov(X,Z)}$

→ R^2: If x is corr w/ u then $var(y) \neq \beta var(x) + var(u)$

↳ non-natural interpretation

↳ most packages compute $R^2 = 1 - \frac{\Sigma \hat{u}^2}{\Sigma(y_i - \bar{y})^2}$ but $RSS > TSS$ so $R^2 < 0$

↳ cannot interpret

TSLS for a MLR: $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 W_1 + \cdots + \beta_k W_k + u_i$

→ assume $cov(W_j, u) = 0$ where $j = 1, \cdots k$ + $cov(X_1, u) \neq 0$

→ $W_1, \cdots W_k$ are exogenous + X_1 is potentially endogenous

Stage 1: Run reduced form regression: $X_1 = \pi_0 + \pi_1 Z_1 + \underbrace{\pi_2 W_1 + \cdots + \pi_k W_k}_{\text{exo explained by } Z + W} + \overset{\text{endo}}{u}$

Get $\hat{X}_1 = \hat{\pi}_0 + \hat{\pi}_1 Z_1 + \hat{\pi}_2 W_1 + \cdots + \hat{\pi}_k W_k$

Stage 2: Run original regression w/ $\hat{X}_1$ replacing X_1

$Y = \beta_0 + \beta_1 \hat{X}_1 + \beta_2 W_1 + \cdots + \beta_k W_k + u_i$ * SE incorrect!

TSLS w/a Single Endogenous Regressor: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 W_{1i} + \cdots + \beta_{1+r} W_{ki} + u_i$

* m instruments

Stage 1: Regress X_1 on all exogenous regressors (use OLS to reg X_1 on $W_1, \cdots W_k, Z_1 \cdots Z_m$)

Compute predicted values $\hat{X}_{1i}$ $i = 1, \cdots, n$

Stage 2: Regress Y on $\hat{X}_{1i}$, $W_1 \cdots W_k$, + an intercept by OLS

TSLS w/ Multiple Endo Reg: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_m X_{mi} + \beta_{m+1} W_{1i} + \cdots + \beta_{k+m} W_{ki} + u_i$

Stage 1: Reduced Form Reg

$X_1 = \pi_0 + \pi_1 Z_1 + \cdots + \pi_{\tilde{m}} Z_{\tilde{m}} + \pi_{\tilde{m}+1} W_1 + \cdots + \pi_{\tilde{m}+k} W_k + u$

$X_m =$ ______

Get $\hat{X}_j = \hat{\pi}_0 + \hat{\pi}_1 Z_1 + \cdots$

Stage 2: Run original reg w/ $\hat{X}_1, \hat{X}_2, \cdots \hat{X}_m$

Evaluating Instrument

We need 1. Relevance: $Corr(Z, X) \neq 0$ → easy to test

2. Exogeneity: $corr(Z_{mi}, u_i) = 0$ → hard

] If not satisfied, then IV estimator has asymptotic bias > OLS

→ Checking Condition 1: Instrument Relevance

We have $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 W_{1i} + \cdots + \beta_{k-1} W_{ki} + u_i$

First Stage Reg $X_i = \pi_0 + \pi_1 Z_{1i} + \cdots + \pi_m Z_{mi} + \pi_{m+1} W_{1i} + \cdots + \pi_{m+k} W_{ki} + u_i$

↳ instruments are relevant if at least one π_m are non zero

→ perform partial F-test; if F > 10 then instrument is strong

→ weak instruments explain little variation in X beyond that explained by Ws

↳ consequences: estimator is biased

stat inferences can be misleading

sampling distribution of TSLS + t-stat are not normal even w/ large n

$\hat{\beta}_1^{IV} = \frac{\hat{cov}(Y, Z)}{\hat{cov}(X, Z)}$ nearly zero, so $\beta \uparrow$ (biased)

Ex: Accmoglu et al — institutional quality on economic growth

→ simultaneous causality ⇒ instrument: mortality rate

↳ ↓mr = create better institutions

Ex 2: $GPA = \beta_0 + \beta_1 alc$; instrument: price of alcohol

→ Checking Condition 2: Instrument Exogeneity

↳ if $Corr(Z_{mi}, u_i) \neq 0$ then first stage 2LSLS cannot isolate part of X uncorr w/error term so $\hat{X}$ is corr w/u + 2SLS is inconsistent

↳ when instruments > endogenous regressors (over identified), it is possible to partially test for this condition

We have $Y_i = \beta_0 + \beta_1 X_i + u_i$ w/2 instruments Z_{1i}, Z_{2i}

Compute two separate 2LSLS est.

→ If 2 est are very different, one or both instruments is invalid

Make this comparison w/ J-test

→ statistically precise

J-Test: Anderson-Rubin Test using TSLS estimator instead of hypothesized value $\beta_{1,0}$; the steps are

* only if #Z's > #X's

$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_m X_{mi} + \beta_{m+k} W_{ki} + u_i$

1. Est. equation using TSLS + all r instruments: $\hat{Y}_i = \hat{\beta}_0^{IV} + \hat{\beta}_1^{IV} + \hat{\beta}_2^{IV} W$
   - → compute predicted values $\hat{Y}_i$ using actual X's (not $\hat{X}$)
2. Compute residuals $\hat{u}_i = Y_i - \hat{Y}_i$
   - → exogenous Zs will be uncorr w/u
3. Regress against $Z_{1i}, \dots Z_{ri}, W_{1i} \dots W_{mi}$
4. Compute F-stat testing hypothesis that coefficients on $Z_{1i} \dots Z_{ri}$ are all 0
5. J-stat is $J = r * F$
   - → F = f-stat testing coeffients on Z_{1i}, Z_{ri} in a regression of the TSLS residuals against $Z_{1i}, Z_{ri}, W_{1i} \cdots W_{ki}$
   - → Distribution: has chi-sq. distribution w/r-m d.f (χ^2_{r-m}) under null that all instruments are exogenous
     - ↳ If r=m, J=0
     - ↳ If some instr. are laro/exogenous + others endo, J-stat will be large + null will be rejected

Final Review

NONLINEAR REGRESSION MODELS

I. Log-Linear Regression Models
1. Log-lin: $\ln(y) = \beta_0 + \beta_1 x_1 + u$
2. Lin-log: $y = \beta_0 + \beta_1 \ln(x_1) + u$
3. Log-log: $\ln(y) = \beta_0 + \beta_1 \ln(x_1) + u$

* Log properties
  - $\ln(ab) = \ln(a) + \ln(b)$
  - $\ln(a/b) = \ln(a) - \ln(b)$
  - $\ln(x^a) = a\ln(x)$
  - $\ln(e) = 1 \rightarrow e^0 = 1$

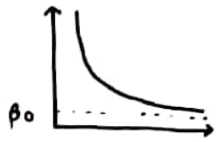
when $\beta > 0.2$, we need $e^{\beta} - 1$ to test effect
but otherwise it is a good approx:

$$\ln(x+\Delta x) - \ln(x) = \ln\left(1 + \frac{\Delta x}{x}\right) \approx \frac{\Delta x}{x}$$

$$\Delta y = f(x_1 + \Delta x_1, x_2 \ldots x_k) - f(x_1, x_2, \ldots x_k)$$

II. Polynomial Regression Models
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \ldots + \beta_k x^k + u$

III. Reciprocal Model: $y = \beta_0 + \beta_1 \frac{1}{x} + u$



* note that x cannot be 0

MULTIPLE REGRESSION MODEL

Elasticity of $E(y|x) = \frac{dE(y|x)}{dx} \cdot \frac{x}{E(y|x)} = \frac{d\ln E(y|x)}{d\ln x}$

I. MLR Assumptions
1. Linear in Parameters: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$
2. X is uncorrelated w/u: $cov(x_i, u_i) = 0$
3. Conditional Mean Zero: $E(u_i | x_{1i}, x_{2i}, \cdots) = 0$
4. Homoskedasticity: $var(u) = \sigma_u^2$] needed for efficiency
5. No autocorrelation: $cov(u_i, u_j) = 0$ for all $i \neq j$ + $ij = 1, \ldots, N$
6. No perfect collinearity: $\rho_{x_1 x_2} \neq \pm 1$
7. $(x_{1i}, \cdots x_{ki}, y_i)$ are iid
8. Finite fourth moments

] w/MLR3 needed for unbiasedness + consistency

II. Interpretation of Estimates
→ Partial effect (ceteris paribus): β_1 measures Δ in predicted value of y per unit Δ in x_1, holding all other variables constant, on average

III. Omitted Variable Bias
$y = \hat{\beta_0} + \hat{\beta_1} x_1 + u$
$y = \tilde{\beta_0} + \tilde{\beta_1} x_1 + \tilde{\beta_2} x_2 + \tilde{\beta_3} x_3 + u$

] $\hat{\beta_1} = \tilde{\beta_1}$ when 1) $\tilde{\beta_2} = \tilde{\beta_3} = 0$ (no effect on Y)
2) x_1 is uncorr w/ x_2 + x_3

Omitted variables are
1. Determinant of Y
2. Correlated with X

$\hat{\beta_1} = \beta_1 + \frac{cov(x,u)}{var(x)}$ *biased + inconsistent

$E(\hat{\beta_1}) = \beta_1$ when $cov(x,u) = 0$

IV. Properties
→ Algebraic Properties
1. $\sum_{i=1}^{n} \hat{u}_i = 0$
2. $\sum x_{ji} \hat{u}_i = 0$ for $j = 1, \cdots k$
3. $\bar{y} = \hat{\beta_0} + \hat{\beta_1} \bar{x} + \hat{\beta_2} \bar{x}_2 + \ldots + \hat{\beta_k} \bar{x}_k$ (line contains the mean)

→ Statistical Properties
1. Unbiasedness: $E(\hat{\beta_j}) = \beta_j$ for $j = 1, 2, \cdots$
2. Variance of OLS Estimators
$var(\hat{\beta}_j) = \frac{\sigma_u^2}{\sum (x_{ji} - \bar{x}_j)^2 (1 - \rho_{x_1 x_2})}$ for $j = 1, 2, \cdots$ *efficient
3. Variance of Error Term
$\sigma_u^2 = \sum \hat{u}_i^2 / n-k-1$ (using residuals)
4. Gauss-Markhov
→ given MLR 1-6, estimators are BLUE

V. Goodness of Fit

1. Adjusted R^2: $\bar{R}^2 = 1 - \frac{SSR}{TSS} \cdot \frac{n-1}{n-k-1}$ * b/c RSS always decr. when a regressor is added $n-1/n-k-1$ offsets it

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

2. Standard Error of the Regression: $\hat{\sigma}_u = \sqrt{\frac{SSR}{n-k-1}}$

HYPOTHESIS TESTING

I. T-Test

→ reject: $t = \frac{\hat{\beta}_j - \beta_{oj}}{SE(\hat{\beta}_j)} > t_{\alpha, n-k-1}$ where $H_0: \hat{\beta}_j = \beta_{jo}$

II. F-Test

→ Partial F-Test: $F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/n-k-1} = \frac{(R^2_{UR} - R^2_R)/q}{(1-R^2_{UR})/n-k-1}$ where $F \sim F_{\alpha,(q, n-k-1)}$

* when $q=1$ then $F = t^2$
* this test is two-sided + views distance betweem β

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

III. Confidence Intervals

→ $\hat{\beta} \pm t_{\alpha, n-k-1} * se(\hat{\beta})$ *ellipsis $(\hat{\beta}_1 \Delta X_1 \pm 1.96 SE(\hat{\beta}_1)\Delta X_1)$ $SE(\Delta\hat{Y}) = SE($____$)$

DUMMY VARIABLES

→ binary + can include m-1 dummies in a regression

I. Interactions

1. Between 2 Binary Variables: $Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} * D_{2i}) + u_i$

where $E(Y_i | D_{1i}=1, D_{2i}=0) = \beta_0 + \beta_1$

$E(Y_i | D_{1i}=1, D_{2i}=1) = \beta_0 + \beta_1 + \beta_2 + \beta_3$] $E(Y_i | D_{1i}=1, D_{2i}=1) - E(Y_i | D_{1i}=1, D_{2i}=0) = \beta_2 + \beta_3$

2. Between Continuous + Binary

→ Diff intercept, same slope: $Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i$

→ Diff intercept, diff slope: $Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i * D_i) + u_i$

→ Same intercept, diff slope: $Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i * D_i) + u_i$

3. Between 2 Continuous Variables: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} * X_{2i}) + u_i$

* effect of Δ in X_1 on Y depends on X_2 + vice versa

II. Applications

1. Event Studies: control for before + after significant event
2. Seasonality: include dummies at approp. time intervals
3. Fixed Effects Model: $Y_{it} = \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_K X_{Kit} + \alpha_i + \lambda_t + u_i$

→ Panel data can control for factors that are unobserved or unmeasured, vary across entities but not time, + can cause OVB if omitted

↳ these effects are constant over time

↳ use F test to see presence of fixed effects

→ Estimation of Entity Fixed Effects

1. "n-1" Binary Regressors

↳ may reduce degrees of freedom, making OLS impossible

2. Entity Demeaned OLS Regression

$(y_{it} - \bar{y}_{it}) = (\alpha_i - \bar{\alpha}) + \beta_1 (x_{it} - \bar{x}_{it}) + (u_{it} - \bar{u}_i)$

$\tilde{y}_{it} = \beta_1 \tilde{x}_{it} + \tilde{u}_{it}$ where $\bar{y}_{it} = \sum_{t=1}^{T} y_{it} / T$

* fix person + view average consumption across time

3. Before + After Differences Specification

↳ only for 2 time periods

→ Time Fixed Effects

↳ be wary about d.o.f. + multicollinearity problems

VIOLATING MLR ASSUMPTIONS

1. Linear in Parameters: functional form misspecification
   - → consequences: slope + intercept are biased (poor predictions), incr. errors, OVB
   - → detection: plot $\hat{u}$ vs. $\hat{y}$ (want random) or y vs. $\hat{y}$ (want fitted) *not BLUE
     - ↳ Makinnon, White, Davidson (MWD) Test
       1. Estimate $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u_i$
       2. Get fitted values: $y_L = \hat{y}$
       3. Estimate $\ln(y) = \beta_0 + \beta_1 \ln(x_1) + \cdots \beta_k \ln(x_k) + u_i$
       4. Get fitted values: $\ln(f) = \hat{\ln}(y)$
       5. Generate $z_1 = \ln(y_f) - \ln(f)$ * if linear, difference should be zero
       6. Regress + run $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \alpha_1 z_1 + u_i$
          H_0: linear model → reject if α_1 is statistically significant
       7. Generate $z_2 = \text{antilog}(\ln F) - y_f$
       8. Run $\ln(y) = \beta_0 + \beta_1 \ln(x_1) + \cdots + \beta_k \ln(x_k) + \alpha_2 z_2 + u$
          H_0: log-log model → reject if α_2 is stat. sign.
     - ↳ Regression Specification Error Test (RESET)
       1. Get $\hat{y}$ from $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u_i$
       2. Run $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + u$
          → H_0: model is linear ($\delta_1 = \delta_2 = 0$) ⇒ use $F_{2,n-k-3}$
   - → Fix: nonlinear transformations
2. X is correlated w/ u: $cov(x_i, u_i) \neq 0$
   - → consequences: biased estimates, OVB
   - → detection: qualitative
   - → fix: instrumental variables
3. Conditional mean zero: $E(u|x) \neq 0$
   - → consequences: biased intercepts / constant
4. Heteroskedasticity
   - → consequences: inefficient estimates, unreliable/biased hypothesis testing (t, F, CI), not <u>BLUE</u> (does not affect OLS)
   - → detection: plot ($\hat{u}$ vs. y) or ($\hat{u}^2$ vs y) or ($\hat{u}^2$ vs x_i) * residuals over time
     - ↳ Breusch-Pagan Test
       1. Estimate $y_i = \beta_0 + \beta_1 x_i + \cdots + \beta_k x_{ki} + u_i$
       2. Generate $u_i^2 = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_k x_k + v$ ⇒ $\hat{u}_i^2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \cdots + \hat{\alpha}_k x_k$
          H_0: constant variance $E(u^2) = \sigma^2$
          $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$
       3. Run F Test w/ $\dfrac{R^2_{\hat{u}^2}/k}{(1-R^2_{\hat{u}^2})/n-k-1} \sim F_{\alpha,(k,n-k-1)}$
       4. or Koenker LM Statistic: $LM = nR^2_{\hat{u}^2} \sim \chi^2_k$
   - → Fix
     1. Robust variance (White's Variance)
        $var(\hat{\beta}_j) = \sum \hat{r}_{ij}^2 \hat{u}_i^2 / RSS_j$
        ↳ $\hat{r}_{ij}^2$: i^{th} residual from regressing x_i on all other indep. var.
     2. Weighted Least Squares
        - ↳ Known: $var(u) = \sigma_u^2 h(x)$
          1. Divide regression by $\sqrt{h(x)}$
             $y^* = \beta_0^* + \beta_1 x_1^* + \beta_2 x_2^* + u^*$
             $var(u^*) = var(u/\sqrt{h(x)})$
             $\frac{1}{h(x)} var(u) = \frac{1}{h(x)} \sigma_u^2 (h(x))$
             $= \sigma_u^2$
        - ↳ unknown: $\sigma_u^2 = var(u) = \sigma^2 e^{\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k}$
          1. Regress y on x + obtain residuals $\hat{u}_i$
          2. Run $\ln(\hat{u}_i^2) = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + v$
          3. Get predicted values: $\hat{g}_i$
          4. Estimate $\hat{h}(x) = e^{\hat{g}_i}$
          5. Weigh data by $1/\sqrt{\hat{h}(x)}$

          *variance of WLS is more efficient but not good w/ small data

5. Autocorrelation / Serial correlation : $E(u_i, u_j) \neq 0$

→ consequences: inefficient est. + biased hyp tests. due to problems w/SE ; $\hat{\sigma}^2_u$ is likely to underestimate true σ^2_u, t-stat may look bigger

→ detect: plot u vs. time (usually in time series due to momentum) or $\hat{u}_t$ vs $\hat{u}_{t-1}$ (want no pattern)

↳ Durbin - Watson Test

→ Assume 1) $u \sim AR(1)$: error term today depends on error yesterday

2) no intercept

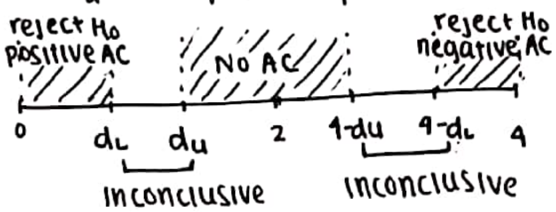
3) original regression has no lagged terms

1. Run OLS + obtain residual $\hat{u}_i$

2. Calculate $d = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1} (\hat{u}_t)^2} \approx 2(1-\rho)$ so $0 \leq d \leq 4$

$u_t = \varphi_1 \hat{u}_{t-1} + \varphi_2 \hat{u}_{t-2} + \dots + \varphi_p \hat{u}_{t-p} + \varepsilon_t$

$(T-P) R^2_{\hat{u}^2} \sim \chi^2_p$ (chi-square distribution)



↳ Regress u_{it} on u_{it-1} w/robust SE to test if coefficients reject zero

→ Fix:

1. HAC Robust variance
2. Generalized Least Squares
    1. Estimate $y_{t-1} = \beta_0 + \beta_1 x_{1t-1} + u_{t-1}$
    2. Multiply all by ρ
    3. Subtract from original

    $y_t - \rho y_{t-1} = \beta_0 (1-\rho) + \beta_1 (x_{1t} - \rho x_{1t-1}) + (u_t - u_{t-1}\rho)$

    4. Estimate $y^* = \beta_0^* + \beta_1 x_1^* + u^*$ using OLS

6. Perfect multicollinearity : $\rho_{x_1, x_2} = \pm 1$

→ consequences: inefficient est, impossible to interpret β, biased hypothesis testing, unreliable OLS

→ detect: correlation matrix, regress multicollinear variable on each other (Klein's Rule of Thumb: high R^2 but low t-test = problematic), large Δs when dropping variables

→ Fix: drop or transform variable; collect more data

INSTRUMENTAL VARIABLES

1. MLR 2 Violation: $Cov(x_i, u_i) \neq 0$

→ consequences: biased estimates, inconsistent estimators

→ causes: 1) Omitted Variable Bias

2) Simultaneous causality

3) measurement errors

→ Two Conditions for valid instruments

1. Instrument Relevance: $corr(z_i, x_i) \neq 0$
2. Instrument Endogeneity: $corr(z_i, u_i) = 0$

→ endogenous: variables correlated w/u ("X")

→ exogenous: variables uncorrelated w/u ("W")

→ detect: empirically, economic intuition

→ fix: instrumental variables!

II. Two Stage Least Squares Estimator

1. Stage 1: Find effect of z_i on x_i (reduced form) — part correlated w/ the error term

   $x_i = \pi_0 + \pi_1 z_1 + \dots + \pi_m z_m + \pi_{m+1} W_1 + \dots + \pi_{m+k} W_k + v$

   * Do this regression for all endogenous var (x_m)

   → Get $\hat{x}_j = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \dots$

2. Stage 2: Estimate Causal Effect of x_i on Y_i

   $Y_i = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2 + \dots \beta_M \hat{x}_M + \beta_{M+1} W_1 + \dots + \beta_{M+k} W_k + u$

   * R^2 may not be useful anymore

→ Formula for IV Estimator: $\beta_1 = \dfrac{cov(y,z)}{cov(x,z)}$

→ Properties for TSLS Estimators for SLR

1. Consistent: $\hat{\beta}_1^{IV} = \hat{cov}(y,z)/\hat{cov}(x,z) \xrightarrow{plim} \beta_1^{IV}$ ($plim_{n\to\infty} \hat{\beta}^{2SLS} = \beta$)
2. Normally Distributed for Hypothesis Testing: $\hat{\beta}^{IV} \sim N(\beta_1, \sigma^2_{\hat{\beta}_1^{IV}})$
   ↳ asymptotically normally distributed when n gets large
3. Asymptotic Variance: $Avar(\hat{\beta}_1^{IV}) = \dfrac{\sigma_u^2}{n\sigma_x^2 S^2_{x,z}} = \dfrac{\sigma^2}{n\sigma_x^2 corr(x,z)^2}$
   ↳ want to be as small as possible
   ↳ if $corr(x,z) < 1 \Rightarrow var(\hat{\beta}^{IV}) > var(\hat{\beta}_{OLS})$
   ↳ higher variance but unbiased estimates

III. Evaluating the Instrument

1. Condition 1: Instrument Relevance
   → Perform partial F-test w/ $H_0: \pi_1 = \dots = \pi_m = 0$
   ↳ strong if $F > 10$
   ↳ consequences of weak instruments
   → estimator is biased: $\hat{\beta}_1^{IV} = \hat{cov}(Y,z)/\hat{cov}(x,z)$ ← close to zero, $\beta\uparrow$
   → stat inferences are misleading
   → sample distribution of TSLS + t stat are not normal even in large n

2. Condition 2: Instrument Exogeneity
   → consequences of violation: 2SLS is inconsistent
   → Use overidentified (instruments > endo.), partially test
   $Y_i = \beta_0 + \beta_1 X_1 + u$ w/ z_{1i}, z_{2i}
   ↳ compute 2 separate 2SLS est + compare w/ J-Test
   → J-Test: Anderson-Rubin Test
   1. Est. equation w/ TSLS
      ↳ compute predicted values $\hat{Y}_i$ using actual X_i
   2. compute residuals $\hat{u}_i = Y_i - \hat{Y}_i$
   3. Regress against $z_{1i}, \dots z_{ri}, W_{1i}, \dots, W_{mi}$
   4. Compute F-stat that all coeff on z_{mi} are 0
   5. J-stat $= r \times F$
      ↳ H_0: all instruments are exogenous
      ↳ chi-sq distribution w/ r-m degrees: χ^2_{r-m}
      → so if $r=m$ then $J=0$
      ↳ reject: some instruments are endogenous