

analyzing economic data/relations

- Econometrics also involves many other more advanced statistical methods that are suitable for

discipline called Econometrics

- In economics, regression analysis is the basis of a for analyzing a wide variety of problems and issues many scientific and management fields, suitable

It is a primary statistical tool for data analysis in other variables

- Regression analysis is used when one is interested on a particular variable and how it is affected by

Introduction

3rd

outcomes in market
Data +

TA hours - let me know if you need help!

- Weekend to realize your full potential in this class

Keep up with the material: 5-6 hours of studying per

material is key to being successful in this class

- The slides and book are not enough: Taking detailed

notes and understanding is well as memorizing the

- Attend class and make sure to focus, follow and understand all of the material being presented

learning all of the material discussed in class

- Might need to read the textbook beforehand

undergraduate-level courses: You are responsible for

- Graduate-level classes are much more demanding than

Some Unsolicited Advice

The Population vs. the Estimated Model

What is a Regression Model?

What is Econometrics?

Introductory Remarks

Agricultural Economics

Quantitative Techniques in

AEC 6610

out in part

(u)

you can never be perfect...

- $u = \text{expected to be } \phi$
- $u \rightarrow \text{will continually fluctuate}$
- $u \text{ can only handle small fluctuations}$

[Predictive model cannot handle all types of variation]+ do best to handle all types of variation.factors - i.e. variables] **variables** \rightarrow u .Pure chance [an outcome of Y (but not in outcome that can not be explained by any known factors)]Random error in the measurement of variablesRelatively unimportant explanatory variablessuch as (constant) dependent variable Y,factors that affect the dependent variable Y,the error term takes into account otherWhat is a Regression Model?out of controlout of controlout of controlsome \rightarrow this is goodwhen all is one of [some things doesn't work] $B_1 = \text{in intercept} \rightarrow \text{value when only model shows}$ id \downarrow goes upper goods (subst) is pos

$$B_2 = -1.5 = 14 \downarrow \uparrow = 15 \text{ increase/yr}$$

 $B_3 = -1.5$ (neg b/c price has negative effect on costs up)

$$X_2 = 8.84 (4/15)$$

$$Y = \text{Boycott} \cdot \text{costs}^{\alpha} (16/15)^{\beta}$$

Unit D What is a $\Delta Y - B_2 \rightarrow$ instantaneous effect of ΔX_2 step warwith explain w/ exact resultants.over time X_4 is not exactrecognizes that the relation between Y and X_2 , X_3 and X_4 u is a random error or disturbance term, which B_1 , B_2 , B_3 and B_4 are estimated using data on Y and X_2 , X_3 and X_4 and the regression analysis technique B_1 , B_2 , B_3 and B_4 are true parameters, i.e., constantcoefficients that quantify the relations between Y and X_2 , X_3 and X_4 derive generic interpretation and givean example) B_1 , B_2 , B_3 and B_4 are true parameters, respectivelyindependent or explanatory variables, respectively Y and X_2 , X_3 and X_4 are the dependent and thein that model (continue with demand example):What is a Regression Model?out of control/layer)prod: inputs \rightarrow resulted good! p of good, y_d/GDP to a type accounts for Δ (factor)thus + performance \rightarrow firm record allows $X \rightarrow$ var that y depends on $\rightarrow Y = B_1 + B_2X_2 + B_3X_3 + B_4X_4 + u$ a mathematical equation or model:in regression analysis, this relation is stated asdemand and production functions)believed to influence Y (economic examples)independent variables (X 's), which aredependent variable (Y) and one or morethis behavioral relation includes onebehavioral relation based on economictheories/knowledge or plain reasoning• Regression analysis starts by conceptualizingWhat is a Regression Model?

IS UNKNOWN \rightarrow PREDICT WITH $y_i | x_i$ DATA
CONTINGENT ON DATA \rightarrow MORE DATA IS BETTER
NOT THE POPULATION PARAMETER NOT BETA

PREDICTOR ESTIMATE: \hat{y}_i

- The Estimated Regression Model
- If data on y_i and $x_{i1}, x_{i2}, \dots, x_{ik}$ ($i=1, \dots, n$) are available, the unknown values of B_1, \dots, B_k can be estimated (show how to do this in Excel)
- The estimated values of B_1, \dots, B_k will be denoted by $\hat{B}_1, \dots, \hat{B}_k$ and called parameter estimates
- In the estimated model, y_i again can be decomposed into a systematic and a random component in the form $y_i = \hat{B}_0 + \hat{B}_1 x_{i1} + \dots + \hat{B}_k x_{ik} + u_i$
- The random component (u_i) is called the residual
- It is also known as the SRF or predicted value of y
- The systematic component ($y_i = B_0 + B_1 x_{i1} + \dots + B_k x_{ik}$) is also known as the population component
- In the estimated model, y_i can be decomposed into a systematic and a random component in the form $y_i = B_0 + B_1 x_{i1} + \dots + B_k x_{ik} + u_i$

$$\text{Population: } y_i = B_0 + B_1 x_{i1} + B_2 x_{i2} + \dots + B_k x_{ik} + u_i$$

↓
observed

$$E[y_i | x_i] \rightarrow \text{giving me } y_i$$

↓
predicted

So # of parameters = k - 1 \rightarrow fewer (less)

absurdness $\# \leftarrow k - 1$ \rightarrow few var

! subscript $\rightarrow y_i + x_i \leftarrow$ denotes

- Population vs Estimated Model
- Consequently, it is very important to differentiate between the population vs the estimated model
- The population model is denoted by $y_i = B_0 + B_1 x_{i1} + \dots + B_k x_{ik} + u_i$:
- where y_i is the i th observation on the dependent variables, x_{i1}, \dots, x_{ik} is the i th observation on the independent variables, B_0, B_1, \dots, B_k are unknown population parameters, and u_i is the i th error term in this model, each value of y_i can be decomposed into a systematic, predictable component ($E[y_i] = B_0 + B_1 x_{i1} + \dots + B_k x_{ik}$) also known as the PRF, and a random, unpredictable component (u_i)
- In this model, the error term (u_i) is denoted by $y_i - E[y_i]$

MANY FUTURES.

judgment call

- Other Preliminary Considerations...
- A model seeks to capture the essentials of the economic process under analysis; it is not meant to be a perfect representation of the process
- To estimate the magnitudes of the effects of the explanatory variables on the dependent variable (i.e. To obtain predictions for the dependent variable (i.e. To estimate the parameters in the regression model)
- Two main uses of a regression model are:
- Two to explainatory variables on the dependent variable (i.e. To estimate the magnitude of the effects of the explanatory variables on the dependent variable)
- Estimating models of behavioral relations, such as demand or production functions, is key to conducting economic analysis

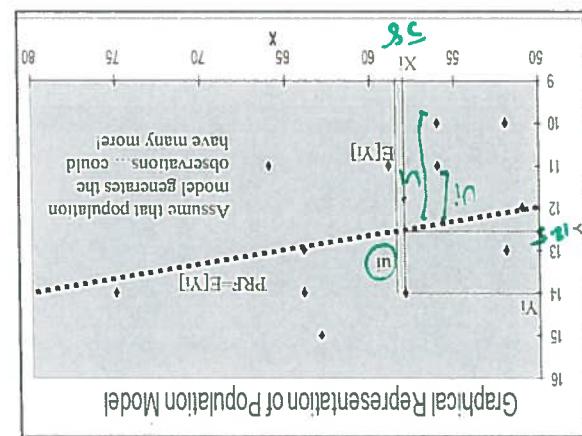
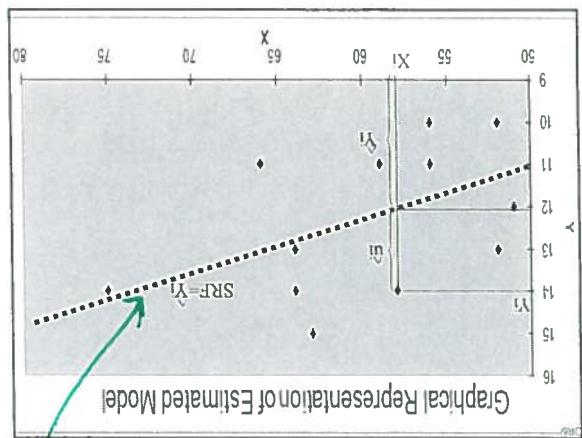
Y_i = $B_0 + B_1 X_i + U_i$

where U_i is the error term

$B_1 = \text{slope}$

$B_0 = \text{intercept}$

Assume no process that has generated your data



- A key concept in econometrics is that the population model is assumed to be the correct data-generating process
- In other words, for any given set of values taken by the explanatory variables (X_i), it is assumed that the population model generates the Y_i data by adding a random error to $E[Y_i]$
- For instance, in the case of a single-variable model, $E[Y_i] = B_0 + B_1 X_i$, which graphically is a straight line
- Assume that population observations could have many more observations than the model generates the model generates the data

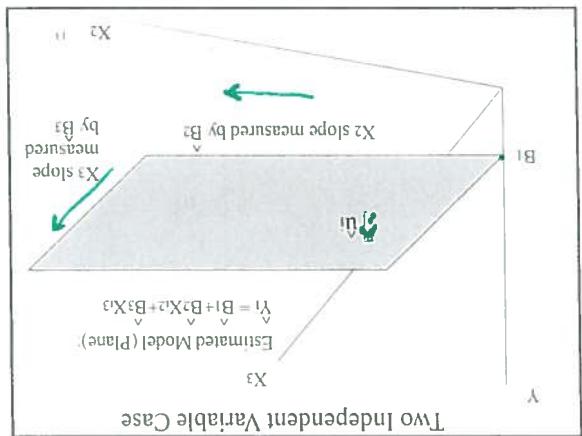
The Population Model

OR6 $B1h=5, B2h=0.12.$

Octavio Ramírez, 8/20/2013
 $Xi=58, Yih=12$

MLR 3 \rightarrow plurinominal

$$y_i = B_0 + B_1 x_{i1} + B_2 x_{i2} + B_3 x_{i3} + \epsilon_i$$



- Class \leftarrow j, fresh, soft, /, nofsr \leftarrow

double counting
 $x_{12} + x_{13} + x_{14} + x_{15} - x_{16} = 0$ \therefore sum full runs

overheads \leftarrow sum full (sum, min, full, spans)

etc.

1 x is perfect linear function to affine

$\therefore C = \text{perf multicut}$

d = fixed costs

$\oplus d_1 + d_2 x_{12}, \dots, d_k x_{1k} \neq 0$ for all!

④ Net perfect multicolumn

dummy var \rightarrow make sum rather \sim 50/50
explain var needs variability (e.g. $\frac{1}{6}$) \leftarrow xs

③ sum var it will be sum of exp var

but over 3

② least 4 obs per parameter, a little less okay
* only for obs to work \leftarrow obs will be build set
k = # parameters (e.g. 20)

n = # obs

② # obs > # parameters

Y_i = B₁ + B₂ * x_{i2} + U_i . can't see b/c non linear in parameter

. can estimate \rightarrow use x² etc exp var

Y_i = B₁ + B₂ * x_{i2} + U_i . non linear in x \leftarrow x²

- sum be non linear in var*

① only for linear in parameters

- did not mult count do obs

- non mult very rare

else \leftarrow q of 19

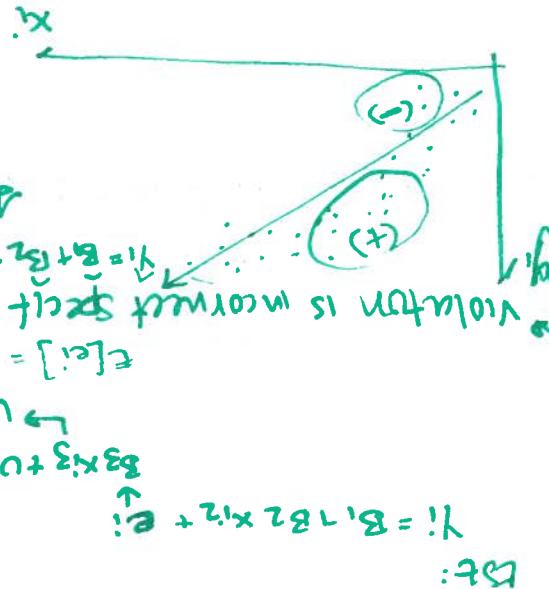
Aug 25

$$(+)=0 \text{ or } (-)$$

same outcome when $U = (-)$

higher than lower

lower curvature \rightarrow shift gives linear



$$E(U) = C$$

$$y_i = B_1 + B_2 x_i + B_3 x_3 + U_i$$

parallel:

funct + shift as curve shift specific

No key result appears excluded from systematic conf.

expected $U_i = C$ in practice requires:

$Z_1 - Z_2$ — least squares

V_i or Z_{12}
fix:

$$[Mkt \leftarrow P \leftarrow Q] \leftarrow \dots$$

deer \rightarrow food availability

unif x \leftarrow y

- random: y count effect \times back

- fixed \rightarrow external

"joint determination"

- drop var \rightarrow feed back on effect \times exp var

Aug 25 cont

Aug 25

if either one doesn't hold? No consistency ~~are~~ unbalanced.

if first is unbalanced is consistent but does it make sense?

main sum of Δ residuals
main resid surface \Rightarrow
area: the Δ distance to line squared
and residual \rightarrow weight of the point
possible
 \rightarrow min of sum \rightarrow sum and
and distance and pos \leftarrow current
 $\oplus (\text{resid.})^2$ or \leq resid.
either which difference
guide w.t.

obj: lower up w/ B1, not at close
obj: B1

How you can do with it: uses: quite

八

$$y_i + x_i = \text{dwarf} \rightarrow \text{giant} : \text{true or unknown}$$

$y_i + x_i = \text{dwarf} \rightarrow \text{giant} : \text{true or unknown}$

constant

$\hat{y}_i = \hat{B}_1 + \hat{B}_2 x_i + u_i \quad \therefore \quad U_i = y_i - \hat{y}_i = (y_i - \hat{B}_1 - \hat{B}_2 x_i)$

- The Ordinary Least Squares (OLS) Method
 - In the OLS method, the criterion for estimating the model parameters is to minimize the residual sum of squares (RSS) – briefly discuss rationale in the case of a single-variable model this is:
 - $$\text{Minimize RSS} = \text{minimize } \sum_{i=1}^n (y_i - \hat{y}_i)^2$$
 - $$= \text{minimize } \sum_{i=1}^n (y_i - B_0 - B_1 x_i)^2$$
 - $$= \text{minimize } \sum_{i=1}^n (y_i - \bar{y} - B_1 x_i)^2$$
 - The RSS is a function like any other, in which the y_i 's and x_i 's are known constants (the data) and B_0 and B_1 are to be viewed as the variables or unknowns which can take on different values and have to be "solved for".

UBLISHING THIS IS (415)

- It is obviously desirable that the estimated regression model is as close as possible to the true population regression model; i.e. that the estimated parameter values (B_1, B_2, \dots, B_k) are as close as possible to the true population parameter values (B_1, B_2, \dots, B_k).
 - The Ordinary Least Squares (OLS) method to estimate B_1, B_2, \dots, B_k (or a variation of it) is in many cases best at accomplishing this objective.

Estimation of the Model Coefficients

Topic 11

AAEC 6610 Quantitative Techniques in Agricultural Economics

The Ordinary Least Squares (OLS) Method

- To find the values of those variables (B_1 and B_2) that minimize the RSS ($\sum (Y_i - B_1 - B_2 X_i)^2$) the partial derivatives of the RSS with respect to those variables have to be taken and set equal to zero:
- $\frac{\partial \text{RSS}}{\partial B_1} = 0 \quad [(\sum (Y_i - B_1 - B_2 X_i)^2)] / \partial B_1 = 0$
- $\frac{\partial \text{RSS}}{\partial B_2} = 0 \quad [(\sum (Y_i - B_1 - B_2 X_i)^2)] / \partial B_2 = 0$
- The former partial derivatives are (chain rule):
- $-2\sum (Y_i - B_1 - B_2 X_i) = 0 \quad \text{--- } (1)$
- $-2\sum (Y_i - B_1 - B_2 X_i) \cdot B_1 = 0 \quad \text{--- } (2)$
- $-2\sum (Y_i - B_1 - B_2 X_i) \cdot B_2 = 0 \quad \text{--- } (3)$
- Notice that the two equations (in the boxes) resulting from setting the two partial derivatives equal to zero make up the following system of linear equations with two unknowns:
- (a) $B_1 + (B_2 X_i) = Y_i \quad \text{--- } (see \text{ part } 1)$
- (b) $B_2 = \frac{\sum X_i Y_i - \sum X_i^2 Y_i}{\sum X_i^2 - (\sum X_i)^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$
- These are the formulas to calculate the OLS estimates (B_1 and B_2) for B_1 and B_2 to find the values of those coefficients which would minimize the RSS in the case of any particular dataset
- Briefly discuss the multivariable case

Two give you the sum of squared error

the sum of residuals (RSS)

get first)

The Ordinary Least Squares (OLS) Method

- This can be easily solved for B_1 and B_2 to find the formulas that have to be used to compute the values of those coefficients which would minimize the RSS in the case of any particular dataset
- The formulas are (discuss summation notation):
- $$B_1 = \frac{\sum Y_i - B_2 \sum X_i}{n \sum X_i^2 - (\sum X_i)^2}$$
- $$B_2 = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$
- These are the formulas to calculate the OLS estimates (B_1 and B_2) for B_1 and B_2 to find the values of those coefficients which would minimize the RSS in the case of any particular dataset

new

multiple

sums

squares of squares

The Ordinary Least Squares (OLS) Method

- To find the values of those variables (B_1 and B_2) that minimize the RSS ($\sum (Y_i - B_1 - B_2 X_i)^2$) the partial derivatives of the RSS with respect to those variables have to be taken and set equal to zero:
- $\frac{\partial \text{RSS}}{\partial B_1} = 0 \quad [(\sum (Y_i - B_1 - B_2 X_i)^2)] / \partial B_1 = 0$
- $\frac{\partial \text{RSS}}{\partial B_2} = 0 \quad [(\sum (Y_i - B_1 - B_2 X_i)^2)] / \partial B_2 = 0$
- The former partial derivatives are (chain rule):
- $-2\sum (Y_i - B_1 - B_2 X_i) = 0 \quad \text{--- } (1)$
- $-2\sum (Y_i - B_1 - B_2 X_i) \cdot B_1 = 0 \quad \text{--- } (2)$
- $-2\sum (Y_i - B_1 - B_2 X_i) \cdot B_2 = 0 \quad \text{--- } (3)$
- The former partial derivatives are (chain rule):
- $-2\sum (Y_i - B_1 - B_2 X_i) \cdot B_1 = 0 \quad \text{--- } (2)$
- $-2\sum (Y_i - B_1 - B_2 X_i) \cdot B_2 = 0 \quad \text{--- } (3)$
- The former partial derivatives are (chain rule):
- $-2\sum (Y_i - B_1 - B_2 X_i) = 0 \quad \text{--- } (1)$
- $-2\sum (Y_i - B_1 - B_2 X_i) \cdot B_1 = 0 \quad \text{--- } (2)$
- $-2\sum (Y_i - B_1 - B_2 X_i) \cdot B_2 = 0 \quad \text{--- } (3)$
- To find the values of those variables (B_1 and B_2) that minimize the RSS ($\sum (Y_i - B_1 - B_2 X_i)^2$) the partial derivatives of the RSS with respect to those variables have to be taken and set equal to zero:
- $\frac{\partial \text{RSS}}{\partial B_1} = 0 \quad [(\sum (Y_i - B_1 - B_2 X_i)^2)] / \partial B_1 = 0$
- $\frac{\partial \text{RSS}}{\partial B_2} = 0 \quad [(\sum (Y_i - B_1 - B_2 X_i)^2)] / \partial B_2 = 0$

* Chain Rule

ignorant summation to do partial derivative

$\frac{\partial \text{RSS}}{\partial B_2} = (Y_i - B_1 - B_2 X_i)^2 = 2$

Why? $\text{def } [x_i] = 0 \rightarrow [x_i x_i] = \text{does not exist}$

* always no x can we function of
runtill crop yield example seasonal + unival

* dummy var 50/50

* @ least 4 parameter

$$y_i = B_1 + B_2 x_{i2} + u_i \quad (x)$$

$$* y_i = B_1 + B_2 x_{i2} + u_i \quad (\checkmark) \text{ nonlin}$$

$$-2(x_i - (y_i - B_1 - B_2 x_{i2})) \leftarrow \sum (y_i - \hat{y}_i)$$

+ ls+:

+ sum of output

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

* sum of predicted - sum of actual

$$\hat{y}_i = (x_i)^T \beta$$

- OLS can only be used to estimate models that have to be satisfied in order to even be able to obtain parameter estimates by OLS:
- These are minimum "technical" requirements that in the parameters, although the models can be non-linear in some level of variability in the sample number of observations must be greater than the number of parameters to be estimated (rule of thumb)
- More generally, there can't be perfect multicollinearity (discuss in some detail, give example)
- There must be some level of variability in the sample values of the explanatory variables (the more the better)
- OLS can only be used to estimate models that are linear in variables (examples)
- In the parameters, although the models can be non-linear in some level of variability in the sample number of observations must be greater than the number of parameters to be estimated (rule of thumb)

OLS Requirements

- The residuals (\hat{u}_i) are not correlated with the values of the independent variables (X_{ij} for $j=2, \dots, k$) or with the dependent variable (y_i), i.e., the residuals pass through the point (\bar{x}, \bar{y}) - why?
- $\sum y_i u_i = 0 \quad (i=1, \dots, n)$ (easier to show with matrices)
- $\sum x_{ij} u_i = 0 \quad (j=2, \dots, k)$, also from normal equations
- The residuals (\hat{u}_i) are not correlated with the values of the independent variables (X_{ij} for $j=2, \dots, k$) or with the dependent variable (y_i), i.e., the residuals pass through the point (\bar{x}, \bar{y}) - why?
- $\sum y_i = \sum \hat{y}_i$ (follows from the above result)
- $\sum u_i = 0$ (follows from the first normal equation)
- A regression model estimated by OLS method has the following key mathematical properties that always hold (show all in Gauss example):

The Ordinary Least Squares (OLS) Method

- Show example in Gauss
- Derive matrix algebra formula for OLS estimator notation
- Discuss matrix representation of population and estimated model (make sure to use X_{ij} and \hat{y}_i)

The Ordinary Least Squares (OLS) Method

effect u_i in y

$\text{df due to } x \rightarrow \text{effect } u$

$= \sigma^2 \text{ sometimes variance}$

$\text{Var of } u_i = \text{Var } u_1 \dots u_3 \dots u_n$

$\text{as } u_i \text{ independent}$

$\text{outcome should for any combination}$

$\text{should hold for any combination}$

$\text{cov}[u_i u_j] = 0$

\leftarrow meets B \leftarrow meets U non \leftarrow meets B \leftarrow meets U non \leftarrow etc.

matrix of degrees \rightarrow small etc.

x_i^T could be correlated with u_i \leftarrow

\leftarrow does not correlate with u_i \leftarrow use new $u_i + u_i$.

1. There is no autocorrelation ($\text{Cov}[u_i u_j] = 0$)
2. The error terms are independently distributed, i.e. the error terms associated with each other (focus on concept for now, will discuss in detail later)
3. There is no heteroskedasticity ($\text{Var}[u_i] = \sigma^2$, same regardless of the values taken by the explanatory variables (focus on concept for now, will discuss in detail later))
4. There is no multicollinearity ($\text{Var}[u_i] = \sigma^2$, i.e. the error terms are totally uncorrelated with each other (focus on concept for now, will discuss in detail later))

OLS Assumptions

- ✓ What is an unbiased estimator, what is a consistent estimator? (briefly, more later)
- ✓ Show that, if assumptions 1 and 2 hold, OLS yields an unbiased or at least a consistent estimator for B . *
- ✓ The functional form is correctly specified, i.e. that the true relationship between y and x is linear.
- ✓ No key explanatory variables have been excluded from the systematic component of the model (show).
- ✓ The expected value of the error term is zero ($E[u] = 0$)

OLS Assumptions

1. The values taken by the explanatory variables are fixed in repeated sampling, if random, they are not correlated with the error term ($\text{Cov}[u_j, u_i] = 0$, $i=2, \dots, k$). In practice, correlation with the error term might occur if:
- The explanatory variables are proxy variables or if they are measured with error (show why, can use IV)
 - The dependent variable has a "feedback" effect on the error term (show why, can use IV)
 - Noy are relevant explanatory variables (can test for endogeneity problem and determine which variables are exogenous)
 - The functional form is incorrect (can test for functional form specification, i.e. they are not fully specified)
2. Assumptions are "things" that have to hold or be true in order for OLS to work best: *
- "Jointly determined \rightarrow $w_{it} (p+q)$ "
- "uncorrelated \rightarrow $z_{it} \perp w_{it}$ "
- "independent \rightarrow $z_{it} \perp v_{it}$ "
- "exogenous \rightarrow $z_{it} \perp u_{it}$ "
- "no correlation between u_{it} and v_{it} "

* if not \Rightarrow not best/wrong

- Answers coming*
- Go back and discuss example in Gauss and Excel corresponding OLS estimate
 - A rule of thumb is that the true parameter value would be within \pm two standard errors of the population regression model (i.e. the B_s) - why?
 - The standard errors provide a measure of the degree of precision with which the OLS
 - Each parameter estimate has a standard error (SE) associated with it
 - Each standard error has a standard error (SE)

The OLS Standard Errors

- Answers coming*
- and the SE estimates
 - Emphasize difference between the true SE matrix (do it in Gauss)
 - Estimates from this estimated covariance matrix show how to obtain standard error
 - Explain how to obtain term $\sigma_{ij} = \text{Var}[U_j]$ - i.e. the error term variance for covariate matrix using matrix algebra
 - Derive the OLS estimator for this
 - Explain what the covariance matrix of \hat{B} is
 - Discuss why \hat{B} is a random variable

Calculating the OLS Standard Errors

- Answers coming*
- Show that if error-term is iid, $\text{Var}[U] = \sigma^2 I$ where I is an $n \times n$ identity matrix
 - Define an iid error
 - $\text{Var}[U] = E[(U - E[U])(U - E[U])^T]$ (expand)
 - Discusses error-term covariance matrix:
 - Explain the concepts of covariance and correlation
 - And discusses how to estimate item
 - Define the variance (of X_1) and the covariance and correlation (X_1 and X_2)

Covariance Matrix

$\bar{x} = \text{Sample mean}$

$$\text{Var}(x_i) = E[(x_i - \bar{x})^2] = \sigma_i^2$$

Some measure

Two measure

$$\text{Var}(x_i) = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_i - \bar{x})^2 \quad \text{Output variance}$$

$$\text{Covar}(x_1, x_2) = E[(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)]$$

Deviation of x_1 from mean

$$\text{Covar}(x_1, x_2) = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$$

$$[\text{Cov}(v/w \text{ var } d \text{ testf} = \text{variance}] = \sigma^2$$

(n-1)

$$\text{est}: \quad \sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) E[\sigma^2] = \sigma^2$$

$B = \frac{m_1 m_2}{m_1 + m_2}$ \rightarrow m_1, m_2 \rightarrow m_1, m_2

Final outcome: if error term is normal distribution of \hat{y}_i will be normal

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$

precise est; Bias \propto difference between \hat{y}_i and y_i

UNBIAS estimate \rightarrow possible bias more

Caveat: among all possible

$g_{\text{goal}} = \text{precise} \rightarrow \text{most precise}$

*MIN Var = lowest pass. SEE

could be non linear estimator
but it's more efficient

Properties of the OLS Estimators

- The Gauss-Markow Theorem also states that:
 - If OLS assumptions 1-4 hold, the OLS method is unbiased.
 - Combining with unbiasedness, this means that, if OLS is a linear estimator
 - estimator also has the minimum variance among all possible unbiased estimators that have linear functions of the dependent variable y (discuss why)
 - If OLS assumptions 1 through 4 hold, the OLS estimate is likely to be closer to the true parameter value of the model than other estimates.
- Discusses Unbiasedness-Efficiency Tradeoff

our real samples \rightarrow multiple dfls

dfn properties of \hat{B}

Properties of the OLS Estimators

- The Gauss-Markow Theorem states the properties of the OLS estimator given the sample data:
 - If $E[\hat{B}] = B$, this means that, on average, the estimates are expected to be equal to the values of the unknown coefficients of the population regression model.
 - In other words, if a large number of different samples of size n are taken, and a similarly large number of estimates for B are calculated using the OLS estimator, the averages of those estimates would be the true parameter vector B .
 - Unbiasedness requires OLS assumptions 1 and 2

Show effects. \rightarrow $\text{Var} = \text{undif Var}$

$\text{Var} = \text{SEE} \times \text{df}$

sum of \square deviation of $x_{i1} + x_{i2}$ \rightarrow $\text{df} = \text{Number of observations} - 2$

$\text{df} = \text{Number of observations} - 2$

\rightarrow error term var \neq SEE (without df)

A Note on the Standard Errors

From the formulas for computing the standard error estimates in the two independent variable case:

$$\hat{\sigma}_{\hat{B}_3} = \hat{\sigma}^2 / ((1 - r_{13}^2) \sum_{i=1}^n (X_{i1} - \bar{X}_{i1})^2)$$

$$\hat{\sigma}_{\hat{B}_2} = \hat{\sigma}^2 / ((1 - r_{23}^2) \sum_{i=1}^n (X_{i2} - \bar{X}_{i2})^2)$$

$$\hat{\sigma}_{\hat{B}_1} = \hat{\sigma}^2 / ((1 - r_{12}^2) \sum_{i=1}^n (X_{i1} - \bar{X}_{i1})^2)$$

where r_{ij} is the correlation coefficient between X_1 and X_3 (recall correlation coefficient). Note that:

The larger r_{ij} , the larger the standard errors unexplained variation), the larger the standard errors used to calculate \hat{B} given the sample data:

If $E[\hat{B}] = B$; this means that, on average, the estimates are expected to be equal to the values of the unknown coefficients of the population regression model.

In other words, if a large number of different samples of size n are taken, and a similarly large number of estimates for B are calculated using the OLS estimator, the averages of those estimates would be the true parameter vector B .

Unbiasedness requires OLS assumptions 1 and 2

\square of correlation \neq df / n

lower \rightarrow higher $\hat{\sigma}_{\hat{B}_1}$

lower \rightarrow higher $\hat{\sigma}_{\hat{B}_2}$

lower \rightarrow higher $\hat{\sigma}_{\hat{B}_3}$

lower \rightarrow higher SEE

lower \rightarrow higher var

Best in most efficient
 Properties of the OLS Estimators
 Gauss-Markow theorem precisely states
 that, if all four major OLS assumptions hold,
 the OLS estimator is BLUE, i.e. the Best Linear
 Unbiased Estimator.
 Best in the sense that it has minimum variance,
 i.e. it is the most precise
 Linear Unbiased estimator that is only best among all
 possible unbiased estimators that are linear
 functions of y .
 However, if error term (ϵ) has a normal
 distribution, the OLS estimator has the
 minimum variance (i.e. it is the most precise)

- Properties of the OLS Estimators
 - The Gauss-Markow theorem precisely states that, if all four major OLS assumptions hold, the OLS estimator is BLUE, i.e. the Best Linear Unbiased Estimator.
 - Best in the sense that it has minimum variance, i.e. it is the most precise Linear Unbiased estimator that is only best among all possible unbiased estimators that are linear functions of y .
 - However, if error term (ϵ) has a normal distribution, the OLS estimator has the minimum variance (i.e. it is the most precise)

Special properties

- If σ is unbiased: OLS estm. expected to equal val. of unknown co-eff
- \hat{B} is unbiased: estm. expected to equal val. of unknown co-eff
- If $\alpha - \beta$ H0 is true: OLS estm. also has min variance among all possible numbers that are LINEAR function of y
 \rightarrow More precise
- If $\alpha + \beta$ H0 is true: OLS is most likely to yield est for E close to actual E
- If μ is normal distib → OLS is min var of all numbers (linearity)

$$\text{No Var} = \text{Vndfn}$$

$$\downarrow \text{Var} = \uparrow \text{SE}$$

$$\left\{ \begin{array}{l} \text{sum of dev from } \\ \text{mean dev} \end{array} \right.$$

$$\frac{\sum (x_i - \bar{x})^2}{n} = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{n} \sum (x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \frac{1}{n} \sum x_i^2 - \frac{2}{n} \bar{x} \sum x_i + \bar{x}^2 = \frac{1}{n} \sum x_i^2 - \bar{x}^2$$

and other terms

→ SE \hat{B}_2

error term: SEE \rightarrow [minimum]

of a least possible SEE

square of result w/ n $x_2 + x_3$

\checkmark $\hat{B}_2 = \frac{1}{n} \sum (y_i - \hat{y}_i)^2$

→ SE \hat{B}_2

Unexplained Variation

- Residual variation is measured as the square of the difference between the observed and the predicted value of Y , i.e. by $(Y_i - \hat{Y}_i)^2$
- Explained variation is measured as the square of the difference between the observed value of Y and the mean of Y , i.e. by $(Y_i - \bar{Y})^2$
- Total variation is traditionally measured as the square of the difference between the observed value of Y and the mean of Y , i.e. by $(Y_i - \bar{Y})^2$
- The R^2 concept is based on the notion that each Y observation (Y_i) can be decomposed into total, explained, and residual variation (wrt \bar{Y})

The R^2

R^2

- The main statistical use to quantify how well the estimated regression model "fits" the Y data is unanswerable
- However, the question of how well can this best fitting model predict the behavior of Y remains and, under certain conditions, it's BLUE
- Using the OLS method to estimate B guarantees that the estimated regression is the best fitting model (for the data at hand) in the sense that it has the smallest possible residual sum of squares and, under certain conditions, its BLUE
- Measures of Goodness of Fit

[Measures of Goodness of Fit]

Topic III

Quantitative Techniques
in Agricultural Economics

AEC 6610

Sept 6 start.

Chap 3

$$- E(Y_i - \hat{Y}_i)^2 = E(Y_i^2 - 2Y_i\hat{Y}_i + \hat{Y}_i^2)$$

\hat{Y}_i

\rightarrow all y_i in straight line

HOLDS TRUE $\Rightarrow R^2 \text{ undefined}$

$$0 < R^2 < 1$$

SO R^2 has no units. No Δ units.

Both sum of Δ and scale units.

$$\text{explaned ss} = \text{regression ss}$$

$$TSS = ESS + RSS$$

number for var y

$$so \rightarrow E(Y_i - \hat{Y}_i)^2 = E(Y_i^2) - E(\hat{Y}_i^2)$$

$$= Y^2 - \hat{Y}^2$$

$$= Y^2 - \hat{Y}^2$$

- The higher the R^2 , the better the fit
- In a simple regression the value of the R^2 equals $(\text{regressor}^2 / \text{total}^2)$ (i.e. the square of the correlation coefficient between Y and X)
- A perfect linear relationship between X and Y (no error)
- $R^2 = 1$ ($\text{ESS} = \text{TSS}$, $\text{ESS} = 0$) indicates a perfect fit (there is no error)
- $R^2 = 0$ ($\text{RSS} = \text{TSS}$, $\text{ESS} = 0$) indicates no fit at all (why?)
- It can only take values between 0 and 1 (why?)
- The R^2 is a ratio, it has no units, its value is the same regardless of the units in which Y is measured
- Variation in Y that is explained by the model (i.e. X)
- Thus, the R^2 measures the proportion of the total variation that is explained by the model (i.e. X)

The R^2

$$TSS = ESS + RSS$$

Plug into

- In the second case, since (RSS/TSS) measures the proportion explained by the model - why equal?
- The proportion that variation that is being explained by the model is also the proportion that is being explained by the dependent variable that is being measured
- In the first case, note that ESS/TSS measures the proportion of the total amount of variation (TSS) that is not being explained by the model
- $\Rightarrow R^2 = 1 - (\frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}) = 1 - (\text{RSS}/\text{TSS})$ or $\Rightarrow R^2 = \frac{\sum(Y_i - \bar{Y})^2 / \sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2} = \text{ESS}/\text{TSS}$
- Therefore, the R^2 is calculated as:

The R^2

- Mathematically shows that $TSS = ESS + RSS$
- Thus RSS/TSS measures $\frac{\text{RSS}}{\text{TSS}}$
- The residual sum of squares ($\text{RSS} = \sum(Y_i - \hat{Y}_i)^2$) measures how much of the total variation (TSS) is not being explained by the estimated model
- $\sum(Y_i - \bar{Y})^2$ measures how much of the total variation (TSS) is being explained by the regression model (i.e. of the TSS)
- The explained (regression) sum of squares ($\text{ESS} = \sum(Y_i - \hat{Y}_i)^2$) measures how much of that total variation of Y is being explained by the dependent variable ($\text{numerator of variance}$)
- The total sum of squares ($TSS = \sum(Y_i - \bar{Y})^2$) measures the total amount of variation in the observed values of the dependent variable ($\text{denominator of variance}$)
- Specifically:

The R^2

$$1 + \frac{1}{n} = \frac{n}{n+1} = \frac{1}{R^2}$$

survival culture
survival culture
TWS NO
TWS exp. WLR
TWS sum is user priority

useful
survives
 ≥ 1
put lots of time in
all way to account

- The Adjusted R^2
- The adjusted R^2 denoted by R^2 is better to assess whether the addition of an independent variable likely increases the ability of the model to predict Y :
 - $R^2 = 1 - \frac{\text{Var}(u)}{\text{Var}(Y)} = 1 - \frac{(\text{RSS}/\text{TSS})}{(n-1)/(n-k)}$
 - Notice that the R^2 is always less than the R^2 (why?) unless $k=1$ or $R^2=1$
 - Unfortunate, the R^2 does not have the same straightforward interpretation as the R^2 ; under unusual circumstances, it can even be negative
 - Some argue that the R^2 provides for a better measure of goodness of fit when the model is estimated with lots of variables and few observations.

$$\frac{\text{RSS}}{n-k} = 0$$

$$\frac{\text{RSS}}{n-1} > 0$$

$$\frac{\text{RSS}}{n-1} < 0$$

most likely $\rightarrow R^2$
add exp var: no matter what it is
poor R^2 still useful model?

- The R^2
- A disadvantage of the R^2 as a measure of a model's goodness of fit is that it always increases in value as independent variables are added into the model, even if those variables don't have a real effect on Y .
 - This happens because, when estimating the model's coefficients by OLS, any added independent variable would allow for a smaller (RSS) (or at least the same) RSS .
 - An increase in the R^2 as a result of adding an independent variable to the model does not mean that the "expanded" model is better, or that the independent variable really affects Y in the population.

$$\Delta R^2$$

survivors
same time
for one point of time: different
cross section: yields: data
survived over time
time series: data on past history
 $\text{COV} L^2 = R^2$

- The R^2
- The R^2 is only a measure of the model's capacity to predict Y , so models with very low R^2 but precise parameter estimates serve other useful purposes... "valid" if the model includes an intercept, "otherwise $ZY \neq ZY$! (why?) and $\text{TSS}/\text{RSS} + \text{ESS}$
 - A final important note is that the R^2 is only useful if the model's estimated using cross-sectional data a good R^2 (i.e. a good fit) is perhaps 0.30 or higher
 - In models estimated using time-series data a good R^2 (i.e. a good fit) is perhaps 0.80 or higher
 - In the type of data being analyzed the value of the R^2 must be assessed in light of

$$\Delta R^2$$

$$\text{surv} = 1/\text{var}$$

use proxy \rightarrow use indirect to est.

use it
Want to put in if believe
 \rightarrow if tiny effect & don't put in u

- A Final Note on Model Specification
- Any variable that is believed to directly and notably affect Y , and did not hold a constant value in the sample, should be included in the model
- Excluding such a variable would likely cause the estimates of the remaining parameters to be biased and inconsistent because the expected value of the error term would not be zero ($E[U] \neq 0$)
- The potential consequence of including irrelevant variables in the model (increased multicollinearity) is less serious; if in doubt, this is preferred
- If a variable only affects Y indirectly, through another independent variable in the model, it does not have to be included in the model

Parameter Estimates (estimated from data)

Observation (year)	B1hat 141.5516	B2hat -0.1469	B3hat 0.1771	B4hat 0.0331	B5hat 0.000175	TIME (lbs/pers-year) (year)	PRED. VALUE (lbs/pers-year)	RESIDUAL Ui
	Yi 1000	Xi2 (cents/lb)	BEEF PRICE (cents/lb)	CHICKEN PRICE (cents/lb)	PORK PRICE (cents/lb)	Xi5 (\$/pers-year)	Xi6 (\$/pers-year)	Uihat 4.64 3.70
1	118.80	431.10	179.30	381.40	13760	1	123.44	-4.64
2	127.50	390.20	159.80	359.00	14242	2	123.80	3.70
3	124.00	366.90	151.10	315.20	14691	3	122.78	1.22
4	117.90	417.90	155.40	335.50	15349	4	115.29	2.61
5	105.50	466.90	142.10	302.50	15547	5	103.14	2.36
6	103.40	432.00	131.10	257.90	15186	6	103.24	0.16
7	104.20	393.50	122.70	255.50	15099	7	105.78	-1.58
8	104.30	376.50	112.70	277.00	15076	8	105.67	-1.37
9	106.40	358.20	110.90	259.80	15299	9	105.97	0.43
10	106.20	345.40	118.80	237.70	15954	10	107.10	-0.90
11	106.90	323.80	108.10	229.50	16427	11	106.65	0.25
12	108.00	315.40	116.10	248.10	17090	12	108.49	-0.49
13	104.00	319.80	105.30	252.80	17507	13	104.62	-0.62
14	103.10	322.40	110.00	236.30	17804	14	103.04	0.06
15	98.40	326.60	113.90	224.70	18193	15	101.26	-2.86
16	96.10	327.70	104.80	247.90	17667	16	98.62	-2.52
17	95.40	322.60	98.50	237.10	17225	17	96.28	-0.88
18	95.00	309.10	94.40	215.10	16985	18	95.23	-0.23
19	93.00	309.40	93.90	208.40	17524	19	93.43	-0.43
20	96.40	290.90	93.50	203.70	17929	20	94.46	1.94
21	97.10	284.30	91.70	194.80	18143	21	93.31	3.79

*Dataset is a sample from the population

RESIDUAL OUTPUT

Observation Observed Predicted Residuals

	t Stat	P-Value	Lower 95%	Upper 95%	Power 95.0% per 95.09
Intercept	141.5516	7.176	0.000	99.508	183.595
B2	-0.1469	0.022	-6.779	0.0000062	-0.101
B3	0.1771	0.088	2.021	0.062	0.101
B4	0.0331	0.037	0.892	0.386	0.112
B5	0.0002	0.001	0.162	0.002	0.002
B6	-1.5391	-4.300	0.001	-2.302	-0.776

SUMMARY OUTPUT

Regression Statistics
Multiple R 0.975576
R Square 0.951748
Adjusted R Square 0.935664
Standard Error 2.434128
Observations 21

ANOVA
df SS MS F Significance F
Residual 15 1753.011 350.6022 59.17356 2.44E-09
Regression 5 88.874713 5.924981 Total 20 1841.8857

Coeficients Standard Errc t Stat P-Value Lower 95% Upper 95%

	t Stat	P-Value	Lower 95%	Upper 95%	Power 95.0% per 95.09
Intercept	141.5516	7.176	0.000	99.508	183.595
B2	-0.1469	0.022	-6.779	0.0000062	-0.101
B3	0.1771	0.088	2.021	0.062	0.101
B4	0.0331	0.037	0.892	0.386	0.112
B5	0.0002	0.001	0.162	0.002	0.002
B6	-1.5391	-4.300	0.001	-2.302	-0.776

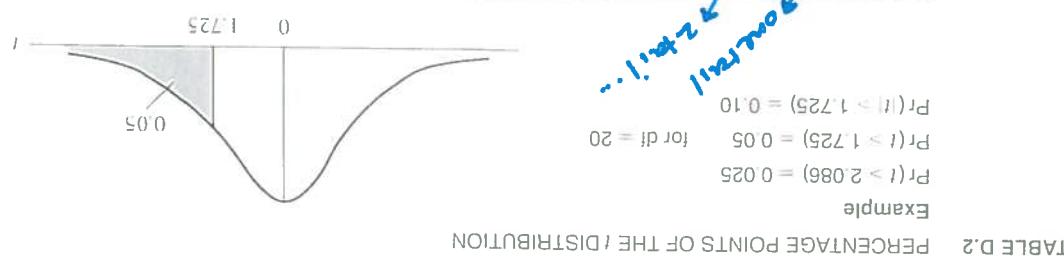
RESIDUAL OUTPUT

	97.10	93.310	3.790
20	96.40	94.455	1.945
19	93.00	93.431	-0.431
18	95.00	95.231	-0.231
17	95.40	98.622	-2.522
16	96.10	101.258	-2.858
15	98.40	104.623	-0.623
14	103.10	103.040	0.060
13	104.00	104.623	-0.623
12	108.00	108.492	-0.492
11	106.90	106.649	0.251
10	106.20	107.097	-0.897
9	106.40	105.974	0.426
8	104.30	105.673	-1.373
7	104.20	105.777	-1.577
6	103.40	103.240	0.160
5	105.50	103.138	2.362
4	117.90	115.290	2.610
3	124.00	122.775	1.225
2	127.50	123.802	3.698
1	118.80	123.441	-4.641

Cambridge University Press New York, 1966. Reproduced by permission of the editors and trustees of Biometrika
Source: From E S Pearson and H O Hartley eds, *Biometrika Tables for Statisticians* vol 1, 3rd ed, table 12.

is the area in both tails
Note: The smaller probability shown at the head of each column is the area in one tail, the larger probability

$\Pr(t > t_0)$	0.25	0.50	0.10	0.05	0.025	0.01	0.005	0.001	0.0002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31	22.327	2.086
2	0.741	1.533	1.76	2.132	2.776	3.747	4.604	7.173	4.144
3	0.765	1.638	2.920	4.303	6.965	9.925	10.214	10.214	4.025
4	0.741	1.533	1.76	2.132	2.776	3.747	4.541	5.841	4.501
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893	5.208	4.785
6	0.718	1.440	1.943	2.447	3.143	3.707	3.499	2.998	2.365
7	0.711	1.415	1.937	2.365	3.106	3.796	3.355	2.896	2.106
8	0.706	1.397	1.860	2.306	2.833	3.499	3.055	2.681	1.771
9	0.703	1.383	1.833	2.262	2.821	3.250	3.602	3.012	2.160
10	0.700	1.372	1.812	2.228	2.764	3.131	2.602	2.947	3.733
11	0.697	1.363	1.796	2.201	2.718	3.169	2.602	2.921	3.686
12	0.695	1.356	1.782	2.179	2.718	3.106	2.550	2.852	3.930
13	0.694	1.350	1.771	2.179	2.718	3.106	2.550	2.878	3.646
14	0.692	1.345	1.761	2.145	2.624	3.141	2.539	2.861	3.579
15	0.691	1.341	1.753	2.131	2.602	3.131	2.528	2.845	3.552
20	0.687	1.325	1.725	2.086	2.528	2.845	2.527	2.831	3.527
21	0.686	1.323	1.721	2.080	2.518	2.819	2.508	2.819	3.505
22	0.686	1.323	1.721	2.074	2.518	2.819	2.508	2.819	3.505
23	0.685	1.321	1.717	2.074	2.518	2.819	2.508	2.807	3.485
24	0.685	1.318	1.714	2.069	2.518	2.819	2.508	2.807	3.467
25	0.684	1.316	1.708	2.060	2.518	2.819	2.508	2.807	3.450
26	0.684	1.315	1.706	2.056	2.516	2.819	2.508	2.807	3.435
27	0.684	1.314	1.703	2.052	2.514	2.819	2.508	2.807	3.421
28	0.683	1.313	1.701	2.048	2.513	2.819	2.508	2.807	3.408
29	0.683	1.311	1.701	2.045	2.511	2.819	2.508	2.807	3.396
30	0.683	1.310	1.697	2.042	2.509	2.819	2.508	2.807	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	2.390	3.307	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	2.296	3.232	3.160
120	0.677	1.289	1.658	1.980	2.358	2.617	2.167	3.090	3.090



Sample Number	OLS Estimate	Alternative A	Alternative B	Standard Dev.
1	0.95	0.82	0.93	0.068
2	1.07	0.68	0.99	0.95
3	0.89	0.98	0.90	0.94
4	0.94	1	0.92	0.92
5	1.17	1.28	1.04	0.76
6	0.76	0.72	0.83	0.11
7	1.11	0.76	1.01	1.11
8	1.15	1.48	1.03	0.88
9	1.05	1.28	0.98	0.82
10	0.88	1.62	0.89	1.03
11	1.03	1.1	0.97	0.87
12	0.82	0.58	0.86	0.97
13	0.97	0.78	0.94	0.90
14	0.9	0.66	0.90	1.31
15	0.92	0.92	1.11	0.87
16	0.89	1.1	0.89	1.12
17	1.12	1.2	1.01	1.07
18	1.02	0.86	0.96	0.85
19	0.88	1.08	0.89	0.98
20	1.07	1.18	0.99	1.05
21	1.05	1.06	0.98	1.05
22	0.79	0.64	0.85	0.79
23	0.89	0.94	0.90	0.83
24	0.83	0.8	0.87	0.96
25	0.96	1.62	0.93	1.05
26	1.05	0.74	0.98	1.14
27	1.1	1.24	1.00	1.05
28	0.93	1.04	0.92	1.04
29	1.04	0.76	0.97	1.04
30	1.09	1.14	1.00	1.09
31	0.91	0.9	0.91	1.14
32	0.84	1.14	0.87	1.14
33	0.99	0.78	0.95	1.05
34	1	0.88	0.95	1.14
35	1.14	1.34	1.02	1.14
36	0.86	0.52	0.88	1.24
37	0.88	1.22	0.89	1.24
38	1.24	1.3	1.07	1.24
39	1.14	1.1	1.02	1.14
40	1.31	0.76	1.11	1.11
Average:	1	1	1	0.95
	0.137	< 0.274	0.068	

True σ

get se for each parameter estimate \rightarrow Max

unbiased but less efficient

outlier but less efficient

Bias

outlier

outlier & / Bias

$$Y = 10 + 1X_i + U_i \quad (\text{B} \text{ No+} \text{B})$$

True Population Model is $Y_i = B_1 + B_2 X_i + U_i = 10 + 1.0 X_i + U_i$

Based on 40 Repeated Samples Containing 25 observations each

Table 1: Illustration of Unbiasedness and Minimum Variance Properties:

Outlier
Influence

Explains an
R2

Table 1 illustrates the unbiasedness and minimum variance (i.e., efficiency) properties of OLS and the conditions of the Gauss-Markow theorem are satisfied (i.e., the five OLS assumptions), OLS is an unbiased and most efficient (i.e., minimum variance) estimator for any (and all) of the parameters of a regression model (let's say B_j).

Table 1 assumes that all those conditions are satisfied and OLS is used to obtain 40 estimates for B_j (which means that 40 different samples of size n would have to be taken in order to obtain the 40 needed data sets)*. For illustrative purposes, it also assumes that the true population value of B_j is known to be $B_j=1$ (which would not be the case in practice).

Note that the average of the 40 OLS estimates for B_j is equal to B_j , which means that OLS is an unbiased estimator for B_j (in reality, depending on the sample size, more than 40 samples might be needed for the average to be exactly equal to B_j).

Alternative A illustrates another estimator (i.e. formula or procedure) to compute estimates for B_1 on the basis of sample data. Note that, although this alternative estimator is unbiased (average of 40 estimates also equals 1), the estimates are more dispersed around their mean value of 1 than the OLS estimates: while the OLS estimates range from 0.76 to 1.31, the estimates under alternative A range from 0.52 to 1.62. Also the standard deviation of the 40 OLS estimates is 0.137 versus 0.274 in the case of alternative A. This means that the OLS estimator has a lower variance (i.e. it is more efficient) than alternative A.

- *Note that in practice only one sample of size n and, therefore only one estimate for B_j is actually obtained.
- As a general rule, unbiasedness is a more essential property than minimum variance, i.e. an unbiased estimator would be preferred to a biased estimator with a lower variance because (unlike in our hypothetical Table 1 example) the actual magnitude of the bias and the exact variances of the estimators are difficult to ascertain in practice.
- Alternative B illustrates another alternative estimator which is biased but has a lower variance than OLS. From Table 1, can you tell why alternative estimator B is biased? Can you tell that it has a lower variance than OLS? Which one would you prefer? Why?
- Alternative B illustrates another alternative estimator which is biased but has a lower minimum variance. OLS is added to the Gauss-Markow conditions without qualification, i.e. there would be no other unbiased estimator with a minimum variance that is lower than OLS's. The combination of these two attributes (unbiasedness and minimum variance) makes OLS the preferred method to estimating regression model parameters when the Gauss-Markow conditions plus the error-term normality assumption are satisfied. Other wise, there are alternative estimators that would be preferred to OLS.

Why

OLS

- Recall that under OLS assumptions 1 and 2:
 - ▷ $E[U] = 0$ and $Cov[U, X_i] = 0$
 - ▷ $E[B] = B$ and $Plim[B] = B$
- If, in addition, the error term is i.i.d. (i.e., OLS assumptions 3 and 4 hold), $Var[U] = \sigma^2$ and:
 - ▷ $Var[B] = \sigma^2(X'X)^{-1}$

The Normal-Error Model

- Central Limit Theorem (discuss briefly)
- This assumption is supposedly supported by the distribution: **assumption 2**
 - ▷ $U \sim N(0, \sigma^2)$ (briefly discuss what this means)
 - The error term is also assumed to be normally distributed:
 - ▷ $Cov[U_i, U_j] = 0$ for all i, j
 - ▷ $Var[U_i] = \sigma^2$ for all i , and
 - ▷ $E[U_i] = 0$ for all i ,
 - ▷ $Cov[U, X_i] = 0$ for all i ,
 - ▷ $E[U, X_i] = 0$ for all i ,

The Normal-Error Model

assume OLS assumptions
normally

Confidence Intervals and Hypothesis Testing
The Normal-Error Model
Chapters 4 and 5

Quantitative Techniques
in Agricultural Economics
AAEC 6610

8/26

1st exam answers
Why

$\text{why } \alpha \leftarrow \sim 100 \rightarrow \text{Type I error}$
 $\Rightarrow \text{design: further tails}$

$\nu - k - \text{df}$

$\text{d.f.} \leftarrow$

- Fortunately, it can be shown that: $t_{\text{from } N(0,1)} = \frac{(B_j - \bar{B}_j)/\text{Se}[B_j] - t_{\alpha/2}}{\sqrt{\text{Var}[B_j]}}$ (where k is the number of degrees of freedom, where \bar{B}_j is the mean of parameters in the model) briefly discusses Z versus t distribution
- Also note that since $B_j \sim N(\bar{B}_j, \text{Var}[B_j])$, we have $\text{Z} = \frac{(B_j - \bar{B}_j)/\text{Se}[B_j] - N(0,1)}{\sqrt{\text{Var}[B_j]}}$
- However, recall that $\text{Z} \sim N(0,1)$, which is needed to compute $\text{Se}[B_j]$, is unknown and has to be estimated by $\hat{\sigma}_e^2$, which means that one only has access to an estimate for $\text{Se}[B_j]$ (denoted by $\text{Se}[B_j]$)
- Furthermore, it can be shown that: $t_{\text{from } N(0,1)} = \frac{(B_j - \bar{B}_j)/\text{Se}[B_j] - t_{\alpha/2}}{\sqrt{\text{Var}[B_j]}}$

Confidence Intervals
Student Normal

* why draw curve? basis of CI
for parameter + hypothesis test

normality sampling distribution: wider for larger samples

linear function of normal r.v.
are also N

$$\hat{B} = \bar{B} + (X_i - \bar{X}) \cdot \hat{\beta}_i$$

$$\hat{B} = \bar{B} + \underbrace{(X_i - \bar{X})}_{M_i} \cdot \hat{\beta}_i$$

- The former means that if multiple samples are taken from that distribution (back to b/handout)
- As such, any particular estimate will be expected to be within $\bar{B}_j \pm 3\text{Se}[B_j]$ with 99.8% probability
- Be within $\bar{B}_j \pm 2\text{Se}[B_j]$ with 95% probability
- Be within $\bar{B}_j \pm \text{Se}[B_j]$ with 68% probability

The Normal-Error Model

- If all four OLS assumptions hold and, in addition, the error term is normally distributed:
- $\hat{\beta}_j$ diagonal element of $(X'X)^{-1}$
- $\hat{\sigma}_e^2 = \text{Var}[e_j] = \text{Var}[B_j - \bar{B}_j - (X_i - \bar{X}) \cdot \hat{\beta}_i]$, thus: **unbiased**
- $\hat{\sigma}_e^2 = \frac{1}{n-k} \sum_{j=1}^n e_j^2$ (where $\text{Var}[B_j]$ is the sample variance of B_j)
- $\text{Normality results from the fact that the OLS estimator is a linear function of } U, \text{ and } U \text{ is asymptotically distributed (this result only holds normally distributed)} \hat{\beta}_j \text{ if the } X_i \text{'s are random, in small samples the distribution of } \hat{\beta}_j \text{ would be approximately normal.}$
- "approximate" normality **fixed** in sample size

The Normal-Error Model

will follow a distribution w/ n-k degrees of freedom & two small stds \rightarrow stds of t_{n-k}

$t_{n-k} = -1.9 \rightarrow 10\% \text{ prob } B_1 \text{ occurs}$

95% confidence true effect b/w $-1.9 \rightarrow 10$ [5% prob B_2 occurs]

say p-value

just writing

- Confidence Intervals
- From the previous expression one can obtain a confidence interval for B_1 as follows:
 - $P\{t_{(a/2, n-k)} < B_1 - S_{\text{e}}[B_1] < t_{(1-a/2, n-k)}\} = 1-\alpha$
 - $P\{t_{(a/2, n-k)} < B_1 - S_{\text{e}}[B_1] < t_{(1-a/2, n-k)}\} = 1-\alpha$ (quick example), i.e.:

$$\Pr\{t_{(a/2, n-k)} < B_1 - S_{\text{e}}[B_1] < B_1 + S_{\text{e}}[B_1]\} = 1-\alpha$$

$$\Pr\{t_{(a/2, n-k)} < B_1 - S_{\text{e}}[B_1] < B_1 + S_{\text{e}}[B_1]\} = 1-\alpha$$
 which is a confidence interval for B_1 (do example versus Excel)

Step 13

Hypotheses Testing: t-Tests

- All so, recall that:

$$t_1 = \frac{(B_1 - B_0) / S_{\text{e}}[B_1]}{\sqrt{B_1}} \sim t_{(n-k)}$$
 And note that the ratio above would only follow a distribution if the true population value (for B_1) is used to compute the ratio t_{n-k} . And note that the ratio above would only less the ratio would tend to be consistent with a draw from a $t_{(n-k)}$.
- The more the B_1 value used to compute the ratio departs from the true population value the less the ratio would tend to be consistent with a draw from a $t_{(n-k)}$.
- Therefore, the null hypothesis (H_0) that $B_1 = C$ (where C is any particular number) versus the alternative (H_a) that $B_1 \neq C$ can be tested by:

$$\text{Comparing } t = (B_1 - C) / S_{\text{e}}[B_1]$$
 and see if the test statistic falls in the rejection region.

to assume t_{n-k} distribution
use t_{n-k} distribution
compute t_{n-k} distribution
compute t_{n-k} distribution
compute t_{n-k} distribution
compute t_{n-k} distribution

$$\begin{aligned} t_{n-k} &= 2.131 \rightarrow \text{true} \\ \text{small stds} &\rightarrow \text{large } t_{n-k} \end{aligned}$$

suspecting os

is freedom

- Hypotheses Testing: t-Tests
- Therefore, the null hypothesis (H_0) that $B_1 = C$ (where C is any particular number) versus the alternative (H_a) that $B_1 \neq C$ can be tested by:

$$\text{Comparing } t = (B_1 - C) / S_{\text{e}}[B_1]$$
 and see if the test statistic falls in the rejection region.
 - Find the table value for a $t_{(n-k)}$ associated with that particular level of significance ($t_{(n-k)}$).
 - Select the desired α -level of significance for the test to have been drawn from a $t_{(n-k)}$, specifically:

$$P\{ |t| > t^*_{(n-k)} \text{ reject } H_0 \text{ in favor of } H_1 \text{ and conclude that } B_1 \neq C \text{ with a } \alpha \text{-probability of being mistaken (example)}$$
 - If $|t| < t^*_{(n-k)}$ one can not reject H_0 in favor of H_1 , but the probability that H_0 is correct is unknown (example)

thanks we can do

$B_2 = 0 \rightarrow x_2$ no effect on $y \rightarrow$ now
 x_2 & x_1 should be included
 with regular null turn don't know
 if x_2 has effect
 result of first step: in a
 [Don't be a idiot!]

at least

we can see that H_0
 we have two null hypotheses
 $H_0: B_j = 0$ (i.e. X_j has no
 effect on y)
 $H_1: B_j \neq 0$ (i.e. X_j has an effect
 on y)

- The basic most common use of the t-test is to test whether any given model parameter (B_j) impiles that the corresponding independent variable (X_j) likely affects y :
- The null hypothesis is $H_0: B_j = 0$ (i.e. X_j has no effect on y)
- The alternative hypothesis is $H_1: B_j \neq 0$ (i.e. X_j has an effect on y)
- This is called a two-tailed test because (under H_0) it can be positive or negative. I.e., if X_j has an effect on y , it could be positive or negative.

Statistical Hypothesis Testing

Y-axis

$H_0: B_j = 0$ → null hypothesis
 $H_1: B_j \neq 0$ → alternative hypothesis
 if x_1 & x_2 not ok
 + test valid w/o
 OLS must fail *

sample size

→ better w/ n big

small sample size

if x_1 & x_2 not ok

+ test valid w/o

OLS must fail *

- Note that, strictly speaking, the previously described t-test (and the rule of thumb) is only valid if, in addition to the main four OLS assumptions, the error term is normally distributed, i.e., for the purpose of the t-test can be reliably applied even if the error term is not normally distributed, i.e., for the reliability of the t-test is not guaranteed even if the sample size is relatively large (100-200 observations), the t-test can however, it has been shown that if the sample size is relatively small sample sizes
- However, the four main OLS assumptions are essential for the reliability of the t-tests

Statistical Hypothesis Testing

* detail look @ later

- Logic of t-test (assume $\alpha=0.05$, $t_{(60)}=2$)
- ▷ If $C=B_j$ the t-value (1) would be drawn from $t_{(60)}$, that $C=B_j$ comes from such a distribution, that is, it is unlikely an absolute value of greater than 2 is $\alpha=0.05$, it is unlikely that comes from such a distribution, that is, it is unlikely that $C=B_j$, since the probability that a draw from $t_{(60)}$ takes an absolute value of greater than 2 is $\alpha=0.05$, it is unlikely that there is an $\alpha=0.05$ probability that $C=B_j$
- ▷ Thus, $H_0: B_j = C$ can be rejected in favor of $H_1: B_j \neq C$
- ▷ However, since there is an $\alpha=0.05$ probability that it does come from a $t_{(60)}$, there is a 5% probability of being wrong when rejecting H_0
- ▷ Only if H_0 is rejected, α is the degree of statistical certainty that one can have in that conclusion
- ▷ Alternatively if, for example, $|t|=1.3$ one could only reject H_0 which an $\alpha=0.20$ probability of being wrong
- ▷ If H_0 is not rejected, one does not know the probability of being right or wrong on this decision

$S_{\text{sb}} = \text{postime}$

$\chi^2_d \rightarrow \text{Ward's test} \rightarrow \text{adj. seafoule}$

of two out → value out
pairwise comparison to rule out
turing a knowledge to rule out

$H_0: \beta_i > 0 \leftarrow 0$

Final adj:

[$\Delta \text{seafoul of } x_2 \leftrightarrow \Delta \text{values}$]
scale of other xs
assum of x_2 in same units
final model
to realize all → input in
if durable to take B's out now
Assumptions → OLS + normal (u)
rule for two tail adj

initial Muttels + casuar to reject null

o 2.2.5.1

- Since H_0 is the same, opting for a one-tailed alternative does not affect how the t -value has to be calculated
- The critical table value with which the calculated t value is to be compared, however, has to be adjusted to account for the fact that one of the tails of the t -distribution has been ruled out
- Alternatively, if the same critical table value is used, the actual level of significance would be affected by the number of observations.

Statistical Hypothesis Testing

- This is generally done when the knowledge of the issue at hand or previous research strongly suggests that one of the "tails" in the alternative hypothesis can be ruled out a priori (example)
- In some cases researchers are interested in conducting one-tailed alternative hypotheses ($H_0: \beta_j = C$) but t-tests for a similar null hypothesis ($H_0: \beta_j < C$) but under one-tailed alternative hypotheses (either $\beta_j < C$ or $\beta_j > C$)
- Standard $H_0: \beta_j = 0$ can be rejected in favor of a two-tailed alternative hypothesis ($H_a: \beta_j \neq 0$) (example)
- Alternatively, most computer programs calculate the exact/lowest values (labelled p-values) at which this standard $H_0: \beta_j = 0$ can be rejected in favor of a two-tailed alternative hypothesis ($H_a: \beta_j \neq 0$) (example)

Statistical Hypothesis Testing

of statistical certainty
se must turn x_2 B_j turn
reject H₀
rule of
ratio of
ratio of β_j $\leftarrow \frac{\beta_j}{S_{\beta_j}}$ $\leftarrow \chi^2 / \text{test}$
se must reject

- The t-statistic for this basic test reduces to:

$$Z_1 = \frac{\beta_j - 0}{S_{\beta_j}} = \frac{\beta_j}{S_{\beta_j}} \quad (\text{where } \beta_j \text{ is the value of } \beta_j \text{ estimated using the OLS formulae and } S_{\beta_j} \text{ is the standard error estimate associated with } \beta_j)$$
- The remainder of the test is carried out as explained before (this is where "rule of thumb" comes from).
- In practice, a values of either 0.10, 0.05 or 0.01 can be selected, depending on the nature and objectives of the research (discuss)
- These indicate three possible levels of statistical certainty when rejecting H_0 (90, 95 and 99%), respectively).
- In general, the three possible levels of statistical significance are often referred to as 1%, 5% and 10%.

Statistical Hypothesis Testing

$S \text{ will be outside}$
 $\text{do not do it in 95\%}$

$U_0 \sim N(\mu_0, \sigma^2)$
 confidence set $[U_0]$
 $\text{know } E[U_0] = \mu_0$
 summaries

- Outline process step by step
 $\Rightarrow (Y_0 - E[Y_0]) / S_e(Y_0) \sim N(0, 1)$
- Briefly discuss confidence interval for $E[Y_0] = X_0 B$ - hint: What's the distribution of $E[Y_0]$?
- Briefly discuss confidence interval for Y_0 is thus given by:
 $\Rightarrow Y_0 \in [E[Y_0] - t_{n-k} S_e(Y_0), E[Y_0] + t_{n-k} S_e(Y_0)]$
- A confidence interval for Y_0 is thus given by:
 $\Rightarrow Y_0 \in [t_{n-k} S_e(Y_0) \sim N(0, 1) \text{ and } U_0 / S_e(U_0) \sim t_{n-k}]$
- Thus $U_0 / S_e(U_0)$ is $N(0, 1)$ and $U_0 / S_e(U_0) \sim t_{n-k}$

Confidence Interval for Y_0

$y_0 = x_0 \beta_0 + \epsilon_0 \quad \therefore \quad U_0 = y_0 - \hat{y}_0$
 ci for the parameters + predictions
 Predictions for y
 used for prediction on y \rightarrow ci for values that just give predictions

- It can be shown that:
 $\hat{y}_0 = y_0 - \hat{\epsilon}_0$ is the prediction error at X_0
 $\hat{\epsilon}_0 = y_0 - \hat{y}_0 =$ the predicted value of y given X_0
 $\hat{y}_0 = X_0 B$ - the value to be taken by y given X_0
 $x_0 = \text{a } k \times 1 \text{ vector of specific explanatory variable values}$
 $y_0 = \text{the value of the random variable } y$
 $\hat{y}_0 = \text{the value of the random variable } y$
 $\hat{\epsilon}_0 = \text{the residual } y - \hat{y}_0$
- Notation (with running example):
 $\hat{y}_0 = \text{predicted value for } y$
 $\hat{\epsilon}_0 = \text{residual for } y$

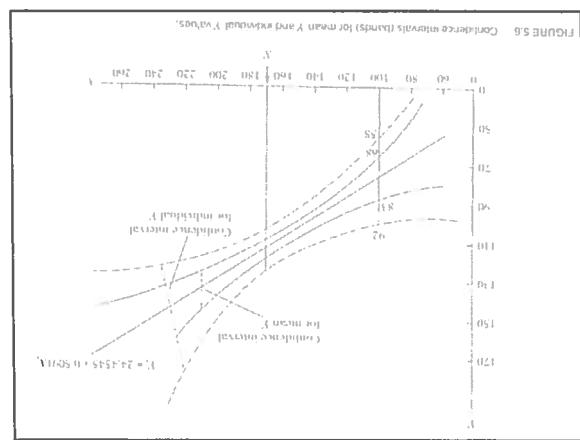
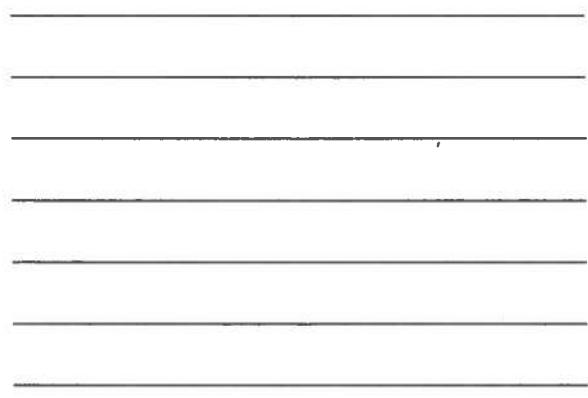
stur + sept 14

used to show $S_e \neq \sigma$ \leftarrow they're not equal
 parameter est \leftrightarrow (S_e) sum + n
 parameter est to denote two
 of significant

- When reporting the results of a regression analysis it is customary to report either the standard errors or the t-values in parentheses below the corresponding parameter estimate
- It is also customary to always conduct a "basic" test for the statistical significance of each of the model's parameters (i.e. to test $H_0: B_j = 0$ for $j = 1, \dots, k$)
- A "rule of thumb" is that if $|B_j| > 2 \times S_e(B_j)$ (i.e. $|t| > 2$)
- B_j is statistically different from zero at the 95% level of statistical certainty ($\alpha = 0.05$ level of statistical significance)

Final Remarks

L



```

    let data[21,6] = [179.3, 13760, 1, 2, 3, 4, 132.1, 142.1, 302.5, 15547, 5, 4, 14691, 14242, 359, 159.8, 152.1, 151.1, 366.9, 157.9, 155.4, 335.5, 15349, 4, 3, 14691, 14242, 359, 159.8, 151.1, 155.4, 335.5, 15349, 4, 5, 466.9, 147.9, 155.4, 335.5, 15349, 4, 6, 151.1, 257.9, 15186, 6, 7, 122.7, 255.5, 15099, 7, 8, 112.7, 277, 15076, 8, 9, 110.9, 259.8, 15299, 10, 10, 116.1, 248.1, 17090, 12, 11, 16427, 229.5, 15954, 108, 12, 104.8, 247.9, 17667, 16, 13, 105.3, 252.8, 17507, 104, 14, 110.236.3, 17804, 14, 15, 113.9, 224.7, 18193, 98.4, 16, 104.7, 247.9, 17667, 96.1, 17, 237.1, 17225, 17, 18, 16985, 215.1, 16985, 95.4, 19, 208.4, 17524, 17524, 19, 20, 17929, 17929, 20, 21: 5
    rows(data) { sets n equal to the number of rows in the dataset (i.e., observations)
    y=dataset[1]: y; wait;
    defines y as the first row of the dataset and prints it to the output
    X=ones(n,1) ~dataset[,2 3 4 5 6]: x; wait;
    defines X as an nx6 matrix with the first column being an nx1 vector of ones
    X=Bhat=inv(X'*X)*X'*y; Bhat; wait;
    computes Bhat using the OLS formula and prints it, then temporally stores it
    Y=Yhat; wait;
    prints Y and what side by side
    SY=sumc(y); SY; wait;
    computes and prints the sum of the values in y
    what=SYsumc(yhat); SYhat; wait;
    computes and prints the sum of the values in yhat
    what=y-X*Bhat; what; wait;
    computes and prints what
    what=yhat; what; wait;
    computes and prints what
    
```

sets n equal to the number of rows in the dataset (i.e., observations)

defines y as the first row of the dataset and prints it to the output

defines X as an nx6 matrix with the first column being an nx1 vector of ones

Bhat=inv(X'*X)*X'*y; Bhat; wait;

computes Bhat using the OLS formula and prints it, then temporally stores it

Y=Yhat; wait;

prints Y and what side by side

SY=sumc(y); SY; wait;

computes and prints the sum of the values in y

what=SYsumc(yhat); SYhat; wait;

computes and prints the sum of the values in yhat

what=y-X*Bhat; what; wait;

computes and prints what

what=yhat; what; wait;

computes and prints what

Computes and prints the sum of the values in $Uhat$.
 $Suhat2 = \text{sumc}(Uhat.^2)$; $Suhat2$; wait ;

Computes and prints the sum of the squared values in $Uhat$.
 X_Uhat ; wait ;

Computes and prints $X^T Uhat$ should be 0
 $X^T Uhat$; wait ;

Computes and prints $Uhat^T$, $Uhat$ should be 0
 $Yhat^T, Uhat$; wait ;

Computes estimate for error term variance
 $S2 = (Uhat^T, Uhat) / (n_rows(Bhat))$; wait ;

Computes estimate for error term variance
 $SEEst = \text{sqrt}(\text{diag}(VBhat))$; $Bhat \sim SEEst$; wait ;

Computes observations
 $YO = Y[i, .]$;
 $XO = X[i, .]$;
 $\text{defines } YO$
 $\text{defines } XO$

Computes $Yhat$:
 $Yhat = XO * Bhat$;

Computes the standard error of the forecasted
 $SehUoh = \text{sqrt}(S2 + S2 * XO * \text{inv}(X^T X) * XO^T)$;

Estimates t-table value for 15 df and $\alpha_{fa} = 0.056$
 $t = 2.13$;

Computes the confidence interval with yo in the m
 $(Yohat - (t * SehUoh)) \sim yo \sim (Yohat + (t * SehUoh))$;

Ends;

spec 20

intcept = 0
 Why \rightarrow throw y into $y = mx + c$
 $mx = 0 \Rightarrow m = 0$
 all $x = 0$ must be same
 all x must be same for
 intercept

$$\text{No } R^2 \quad TS = ESS + RSS$$

goods you want

- Regression through the origin is only true population value of the intercept is actually zero and a proper functional form can be identified if one has very strong evidence that the regression is only
- A "raw" R^2 (Corr Y, \hat{Y}) which only takes values between zero and one can be used instead, but this interpretation is not the same as the regular R^2 because bounded between zero and one and its usual interpretation is not applicable
- The R^2 is not appropriate in this case as its value is obviously different if the model does not have an intercept
- Obviously, the slope parameter estimates will be different if the intercept is not zero and this case is very similar to the raw R^2 which only takes values between zero and one.

Regression Through the Origin

- Computer programs often provide and option for estimating the model without an intercept.
- Such a model can also be estimated using OLS, by simply excluding the first column of ones.
- In hypothesis simple regression models, for example, if the price of a commodity is zero no quantity should be supplied, if there is no income a commodity could not be consumed that stands for the intercept in the X matrix.
- It is used when theory dictates that the intercept should be zero, i.e. that when all X 's take a value of zero y must also be zero
- In hypothesis simple regression models, for example, if the price of a commodity is zero no quantity should be supplied, if there is no income a commodity could not be consumed that stands for the intercept in the X matrix.

Regression Through the Origin

- spec 20
- Alternative Functional Forms
 - Standardized Regression
 - Chapter 6
 - Quantitative Techniques in Agricultural Economics
 - AAEC 6610

- $\beta_1 = .5 \rightarrow \text{every 1sd } y \uparrow .5\text{sd}$
- $\beta_0 = -1.5 \rightarrow \text{base } y \downarrow 1.5\text{sd} \downarrow \bar{y}$
- $y_i^* = \bar{y} - \beta_1 x_i^*$ in sd from mean
- $\text{mean unit of measure, measure}$ of the relationship if x in
- $\text{standardized } x + y \rightarrow \text{get a mean}$

$$\beta_x = \bar{y} - \beta_1 \bar{x}$$

$$y_i^* = \bar{y} - \beta_1 x_i^*$$

$\left. \begin{matrix} \bar{x} \\ \bar{y} \end{matrix} \right\} \text{sum} \quad \left. \begin{matrix} s_x \\ s_y \end{matrix} \right\} \text{var (all x's)}$

and 1 σ

standardize $x + y \rightarrow$ get a mean

- Standardized regression is a regression where all of its variables have a mean of zero, because all of its variables have a standard deviation of one.
- The interpretation of the beta coefficients follows from the fact that both Y and the X 's are now being measured in standard deviations, i.e. they measure the standard deviation change in Y when X changes by one standard deviation.
- The interpretation of the beta coefficients of the standardized regression is that both Y and the X 's are now being measured in standard deviations, i.e. they measure the standard deviation change in Y when X changes by one standard deviation.
- Standardized regression interpretation of the beta coefficients appealsing interpretation

Standardized Regression

- A standardized regression is a regression estimated with all the variables (Y and the X 's) being standardized to have a mean of zero and a variance of one.
- For each variable, this is accomplished by subtracting its mean and dividing by its standard deviation.
- The parameter estimates in this type of regression are known in the literature as the beta coefficients.

Standardized Regression

- A change in the scaling of the dependent variable by a fixed multiplicative factor of w_x only changes the value of all model parameters and their corresponding standard error estimates by a factor of w_x (example).
- A change in the scaling of the independent variable by a fixed multiplicative factor of $(1/w_x)$ (example) does not change anything else except estimates by a factor of w_x (example).

Scaling and Units of Measurement

$\$ \leftrightarrow \text{#} \leftrightarrow \text{#} \leftrightarrow \text{#}$

$$\begin{aligned}
 & \text{Time series data} \rightarrow T = \text{observation} \\
 & \text{Y} = \left[\begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_T \end{array} \right] = \left[\begin{array}{c} \ln t + x_1 \\ \ln t + x_2 \\ \vdots \\ \ln t + x_T \end{array} \right] = \left[\begin{array}{c} \ln t \\ x_1 \\ x_2 \\ \vdots \\ x_T \end{array} \right] + \left[\begin{array}{c} 1 \\ 1 \\ \vdots \\ 1 \end{array} \right] \\
 & \text{where } \left[\begin{array}{c} 1 \\ 1 \\ \vdots \\ 1 \end{array} \right] \text{ is called } \text{Intercept} \\
 & \text{and } \left[\begin{array}{c} \ln t \\ x_1 \\ x_2 \\ \vdots \\ x_T \end{array} \right] \text{ is called } \text{Explanatory variable} \\
 & \text{and } \left[\begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_T \end{array} \right] \text{ is called } \text{Lag variable}
 \end{aligned}$$

long effect \rightarrow sometimes value of x
of previous time period that a currency
- time periods aren't separate

$$\frac{f_{\text{HS}}}{S} = \frac{f_{\text{HS}}}{S} \cdot \frac{1}{1} = x_M$$

$$\frac{sy}{T} = h \mu$$

↑

$\frac{sy}{T} = h \mu$

↑

OLS = penultimate step

Sum of abs. cost \rightarrow larger std more effect

standard regression cost

which was more effective

Lagged Variables

- The OLS method is perfectly suitable for estimating the parameters of models that include lagged values of one or more independent variables
 - It is only necessary to rearrange the data in such a way that the value of Y at time period t coincides with the value of X at time period $t-1$
 - Notice that one observation will be lost for each lag included in the model

Lagged Variables

- When working with time series data, the value of y in time period t sometimes explained by the value taken by x in the previous time period:
- $y = b_0 + b_1 x_1 + \dots + b_t x_t$ ($t=1, \dots, T$) (T =most recent year)
 - For example, a retailer's current year investment decisions might be based on the previous year prices, since the current year prices are not known when making these decisions
 - Here, y is said to be dependent on lagged values of the explanatory variable x
 - It can be assumed that y is affected by more than one lag of x .

Lagged Variables

- The main use of standardized regression is that, since all the Xs are being measured in the same relative scale, the magnitudes of the beta coefficients are comparable measures of the relative importance of each of the explanatory variables in explaining Y.
 - A larger beta coefficient indicates that the contribution X contributes relatively more to explaining the variation in Y.
 - Correlation between the beta coefficients and the dependent variable is called multicollinearity.
 - Standardized regression coefficients can be compared to the OLS parameter estimates to determine the relative influence of each independent variable on the dependent variable.

* $[A \Delta x + Y]$ measurey relateed?]

~~But model now we use variable~~

~~LS~~ Linear in parameters
~~LS~~ linear in variables
~~LS~~ $\Delta x + \Delta y$

Unstable over time \rightarrow How to use
 Accuracy can be lost
 not yet follow short process

Var that follows short process

\rightarrow $\Delta x + \Delta y \rightarrow$ ill condition of some
 changes B :

$\Delta x \rightarrow$ get D first
 $\Delta y \rightarrow$ get after

derivation: for $y \rightarrow$ first order
 \rightarrow issue of passibility

can forecast?

$\frac{\partial y}{\partial x} = (B_2 + B_3) x_2 - B_2 x_{t-1}$

$B_2 (x_t - x_{t-1}) + B_3 x_t$

first use both $\Delta x + \Delta y$

$\Delta x - t_1 \rightarrow \Delta y$

lose 1 obs \rightarrow create Δx

difference between lost & to work
 only or any x

- When there is marked non-linearity in a given $Y-X_j$ relationship, it is not appropriate to assume a linear relationship between $Y-X_j$.
- The regression models studied so far assume that the relationships between Y and the independent variables (X_2, X_3, \dots, X_k) are linear. However, in many cases theory or experience indicates some of the individual $Y-X_j$ relationships are likely non-linear.
- The alternative Model Specifications also be estimated using the OLS method.

$$\Delta x + \Delta y = \text{parallel}$$

- Notice that one observation is always "lost" when using first difference models to address a non-stationarity problem.
- First differences on the dependent variable can be used as well, on the basis of theory or knowledge about the problem at hand or to rates more than by the rate levels themselves be affected by changes in long-term interest rates more than by the change in the short term interest rates.
- Some capital investments, for example, could be affected by changes in long-term interest rates more than by the rate levels themselves be affected by changes in long-term interest rates.

Examples of First Difference Models

- Another common specification in time series models involves letting the dependent or the independent variable, or both, be specified in first differences from one period to the next.
 - The first difference of a variable is its change in value from one period to the next.
 - An example of a first difference model is:
- $$Y_t = B_1 + B_2 X_t + U_t \quad (t=2, \dots, T)$$
- ↓ Δx
- \rightarrow Δy
- In this case, the researcher has a reason to believe that it is the change in X which affects the value taken by Y , in a linear fashion.

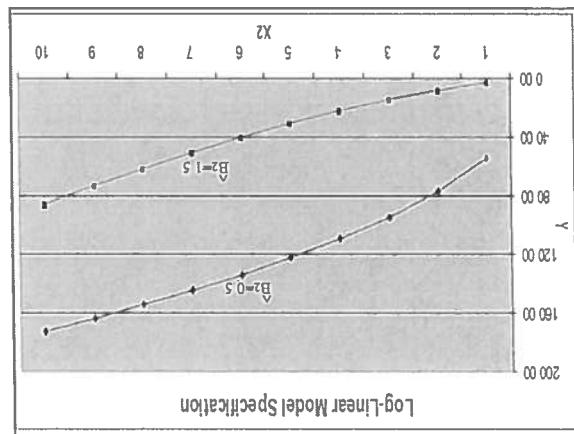
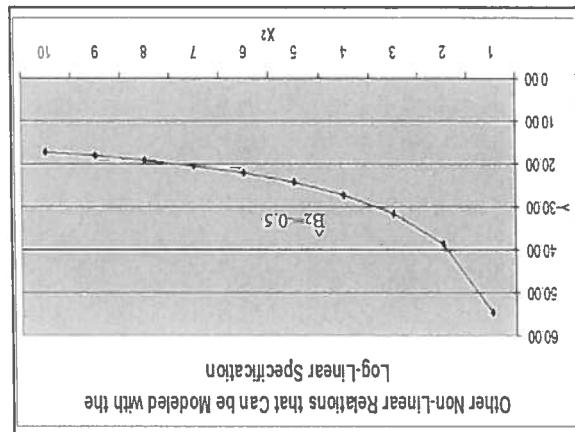
What effects $y \rightarrow$ value of x or Δx

log-log or double log
linear residual error

(Diff Non linear model)

ln(x) vs ln(y) is straight line

~~No decr in
decrease in growth rate~~



- The LOG-Linear Specification
- A special type of non-linear relations become linear when they are transformed with logarithms
- Specifically, consider $y_i = e^{B_1 X_2^{\beta_2} X_3^{\beta_3} \dots X_k^{\beta_k}}$
- Note that $e=2.7182\dots$ is the anti-natural logarithm of $1(\ln(e)=\ln(2.7182\dots)=1)$, therefore e^{B_1} is simply a multiplicative constant
- For example, if $B_1=4$, $B_2=0.5$, $B_3=1.5$ and $X_3=10$, the model is $y_i = e^{4 X_2^{0.5} 10^{1.5}} = 1726.55 X_2^{0.5}$
- This type of model can accommodate a variety of non-linear shapes, as depicted in the following figures

Percentiles \rightarrow distributions

not applicable

natural log (ad hoc)

in a distribution \rightarrow add small to everything

don't advise but can have it

use same form; lower or non linear

- A disadvantage of the Log-Linear model is that all X and Y values must be positive since the natural log of a non-positive number is not defined
 - An important feature is that β_j directly measures the elasticity of Y with respect to X_j ; i.e. the percentage change in Y when X_j changes by one percent (show)
 - Another disadvantage of the Log-Linear model is that, since one needs to take the \ln of Y , all of the $(Y-X)$ relations in the model have to be assumed to be non-linear
 - A disadvantage of the log-Linear specification is that, since one needs to take the \ln of Y , all of the $(Y-X)$ relations in the model have to be assumed to be non-linear
- The Log-Linear Specification

every thing

have to take natural log of

percentiles

tells

one case for obs w/ non linear

$[N_0 + U_i \rightarrow \text{constant obs}]$

$y_i = e^{B_0 + B_1 x_{i1} + \dots + B_k x_{ik}}$ \leftarrow error term + assumed

slope

slope const w/ $x^2 \rightarrow$ but Δx^2

$(D) = \text{no relation}$

$(-) \rightarrow \uparrow \text{ or } \downarrow \text{ slope}$

- The former implies that standard linear regression procedures can be used to estimate the coefficients of a Log-Linear model
 - regression coefficients can be used to estimate dependent and independent variables all i 's and then use the latter as your $X_1 = \ln(X_{i1}), X_2 = \ln(X_{i2}), \dots, X_k = \ln(X_{ik})$ for $y_i = \ln(Y_i), \text{ and } X_1 = \ln(X_{i1}), \dots, X_k = \ln(X_{ik})$, instead of $\ln(X_i)$, $\ln(X_{i2}), \dots, \ln(X_{ik})$ (i.e. let $Y_i = \ln(Y_i)$)
 - specification, but they are applied to $\ln(Y_i)$
- The Log-Linear Specification

- This is also known as the Log-Log or Double-Log specification, because it becomes a linear relationship when taking the natural logarithm of both sides:
 - $\ln(y_i) = \ln(e^{B_0 + B_1 \ln(x_{i1}) + \dots + B_k \ln(x_{ik})})$ (note that $\ln(e) = 1$ and, thus, the intercept above is B_0 and the error term is $\ln(e)$)
 - $\ln(y_i) = \ln(e^{B_0} + B_1 \ln(x_{i1}) + \dots + B_k \ln(x_{ik})) + \ln(e)$ (note that $\ln(e^x) = x \ln(e)$)
 - If $B_j > 1$ as X_j increases y increases at an increasing rate
 - If $B_j = 1$ as X_j increases y increases at a constant rate
 - If $0 < B_j < 1$ as X_j increases y increases at a decreasing rate
 - If $B_j < 0$ as X_j increases y increases at a decreasing rate
- The Log-Linear Specification

curvature is different

$$x_1 \times 100$$

$\% \Delta y$ when $x \Delta 1$ unit

percent change in Y when X changes by one unit

proportion of the unit

$$\ln(y_i) = \beta_0 + \beta_1 x_i + \epsilon_i$$

true log y not log x?

predictions gives y_i

allow: use e ϵ from y_i

gives and some sort of y_i/y_j

comparable to linear model

$$y \propto w_0 + 1$$

$y \rightarrow$ compare to $w_0 + 1$ look up

always use est. percent increase

is equivalent my model

Answers often in this form

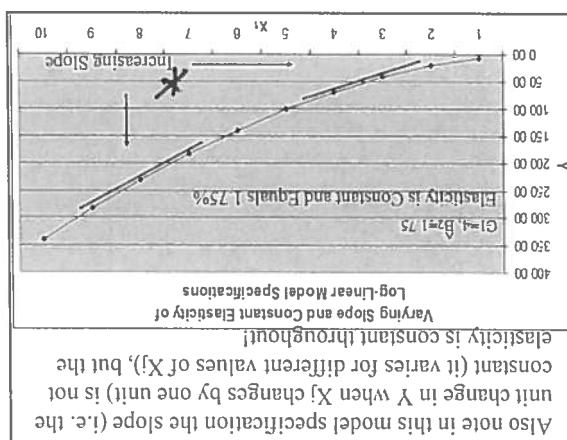
E^2 not same; bounded by $w_0 + 1$

usually log or linear

- Semi-Log Models: The Log-Lin Model
- In the Log-Lin model the logarithm of the dependent variable is taken but the explanatory variables remain in their original (non-log) form
 - Therefore, the regression parameters (β_j 's) measure the relative change in y when the corresponding X_j changes by one unit
 - That is, if multiplied by 100 they measure the absolute (i.e. one unit) amount
 - When the corresponding X_j changes by an increase in y when the corresponding X_j increases by one unit
 - As X increases y may increase at a decreasing rate
 - If $\beta_1 < 0$ or decrease at a decreasing rate

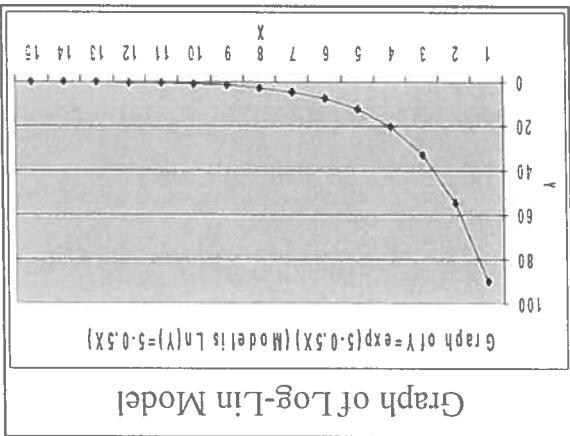
decrease
decrease

- The Log-Linear Specification
- Final notes on the Log-Lin specification:
 - Since the dependent variable is y the R^2 of the estimated log-linear model is not comparable to that of the linear model (why?)
 - In the case of the log-linear model, a more comparable goodness of fit measure would be the square of the correlation coefficient between the observed and the predicted values of y
 - Note that the predictions from a log-linear model have to be obtained from the estimated model that is different from a linear model
 - Equation, not directly from the estimated model

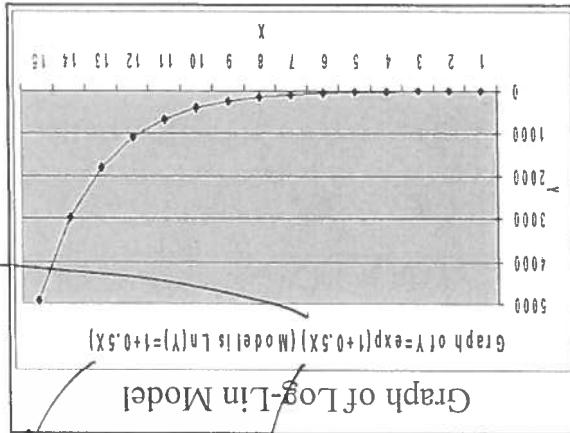


down $\times 2\%$
 write 8s \rightarrow 160 \rightarrow get unit 1
 so unit Δ 4 for 90%
 log y + normal x
 can do ap

- In the Lin-Log model the logarithm of the explanatory variables is taken but the dependent variable remains in its original (non-log) form
 - Therefore, the regression parameters (B_j 's) measure the absolute change (i.e. unit) change in Y when the corresponding X (X_j) changes in a relative (i.e. proportional) manner
 - That is, if divided by 100, they measure the unit change in Y when the corresponding X (X_j) changes by one percent
 - As X increases Y may increase at a decreasing rate ($B_j < 0$) or decrease at a decreasing rate ($B_j > 0$) (see next two graphs)
- Semi-Log Models: The Lin-Log Model



If y can't be neg: x can't be neg
 no \rightarrow e \rightarrow part [LIMIT]
 pos; \rightarrow e \rightarrow eate
 also have component e^x
 $y_i = e^{(B_1 + B_2 x_{i2})} = \ln(y_i) - B_1 + B_2 x_{i2}$



asymptotic effect

Slope is op. sign

$$\frac{dy_i}{dx_i} =$$

Intercept is max value

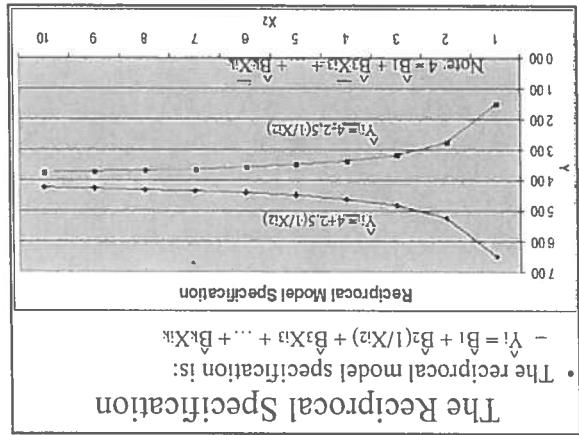
$\therefore B_0 \rightarrow Y \rightarrow \uparrow \text{rate}$

$\Delta X \rightarrow B_X \uparrow \text{approach intercept}$

val curv true ≈ 4

. B_0 pos \rightarrow \downarrow result but lowe

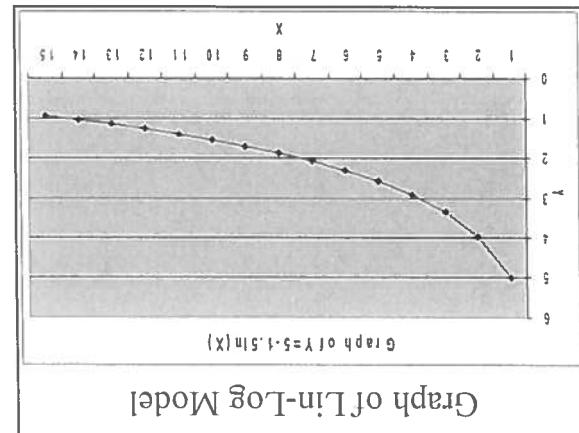
No null today but do *



No \downarrow \rightarrow rate
y curve neg x curv being

$-B_0 \rightarrow \uparrow \text{rate}$

$\uparrow \text{rate} \rightarrow$ rate

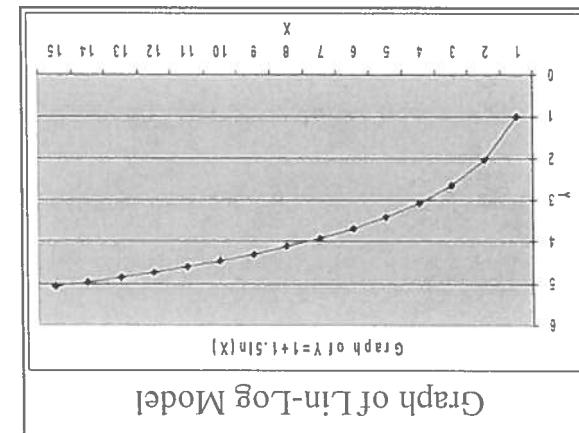


y curv being x curv being neg

* no \rightarrow rate

$\uparrow \text{rate} \rightarrow$ rate

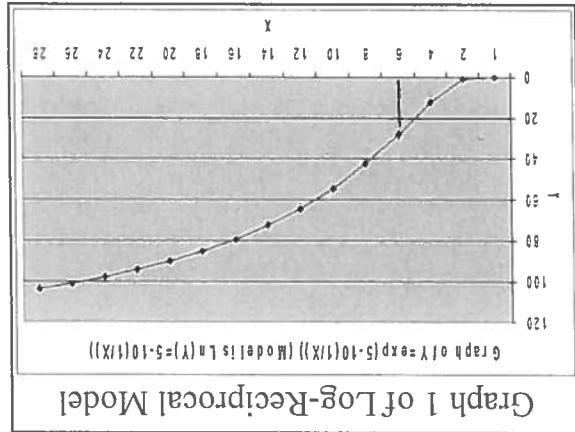
$B > 0$



8. (→)

get \rightarrow rate
 \rightarrow own significant point
 get is slope

$$\ln y_i = \beta_0 + \beta_1 \frac{1}{x_i}$$

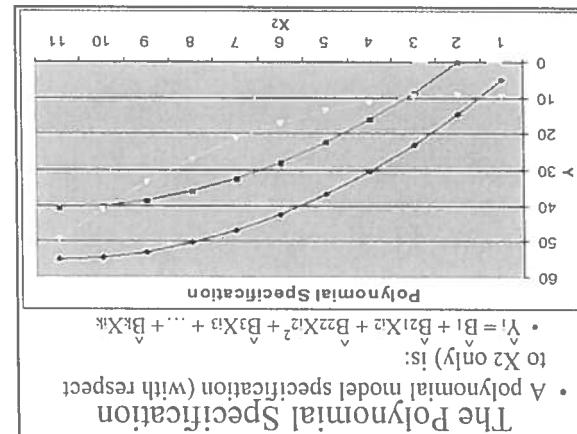
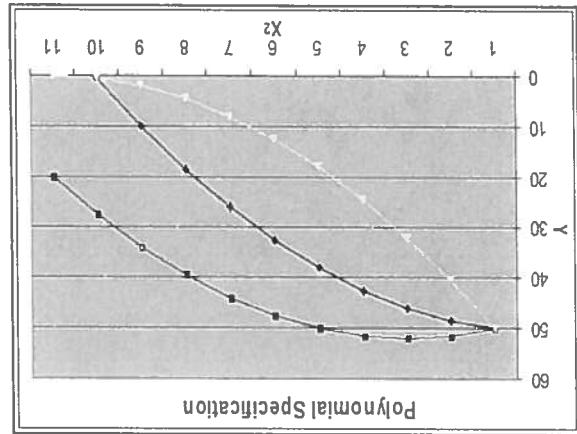


- Note that, as in the case of the Log-Linear model, the Log-Log and Reciprocal models can be estimated by standard OLS procedures
 - In the case of the Log-Log model one has to take the natural logarithm of the y_i data before applying OLS (i.e. use $\ln(y_i)$ instead of the y_i data)
 - In the case of the Log-Linear model one has to use the natural logarithm of the x_i data before applying OLS (i.e. use $\ln(x_i)$ instead of the x_i data)
 - In the case of the Reciprocal model one has to use $(1/x_i)$ as the explanatory variable(s) (i.e. $(1/x_i)^2$) instead of the x_i data
- A Note on OLS Estimation

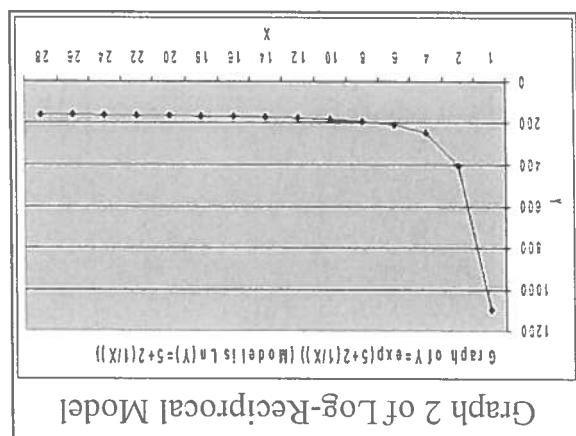
curv. name ~~curve~~ with concave
 curv. name reciprocal + reg x's

- The Reciprocal Specification
- A reciprocal model specifies that look like in the previous graph $y - x_j$ relations that look like in the previous graph decreases, but always at a decreasing rate
 - As x_j (x_2 in the graph) increases, y increases or decreases (which equals 4 in the graphed example)
 - The slope of a reciprocal model specification is: $-dy/dx_j = -\beta_j(1/x_j^2)$, which can be positive or negative depending on the sign of β_j
 - Unlike the linear model specification, the slope is different depending on the value of x_j

x_{ij} pos or (-)
 linear + $\square = \text{rule of polygons}$
 keep linear rule
 convex polygon \rightarrow slopes



parabola pos \leftarrow the result



$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_n x_{in}$

can do any exponent $\rightarrow x_i^k$

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_n x_{in}$$

test & these if $b_3 = 0$

use for test of non-linearity

- In the example before a polynomial regression viewing X_2^2 as any other independent variable is a polynomial model can be estimated by OLS.
- A polynomial model can be used for OLS estimation variables in the data set used for OLS estimation X_2 and X_2^2 would be included as independent specification with respect to X_2 is desired both X_2 and X_2^2 would be included as independent variables in the data set used for OLS estimation
- One of the independent variables can be similarly estimated by OLS.

The Polynomial Specification

- Another advantage of the polynomial specification is that it can easily test for polynomial non-linearity, at the pre-established certainty level
- If the parameter corresponding to X_2^2 is not statistically different from zero, the null hypothesis of linearity can not be rejected in favor of the alternative of polynomial non-linearity, at the pre-established certainty level
- A disadvantage of the polynomial specification is that it can easily used to test non-linearity in the relationship between Y and any given independent variable X_j .
- If the parameter corresponding to X_2^2 is not statistically different from zero, the null hypothesis of linearity can not be rejected in favor of the alternative of polynomial non-linearity, at the pre-established certainty level
- A disadvantage of the polynomial specification is that it can easily used to test non-linearity in the relationship between Y and any given independent variable X_j .

The Polynomial Specification

- However, one needs to be aware of the fact that the relationship between Y and X_j (examples) might eventually imply a reversal in the direction of the relationship between Y and X_j (examples).
- An advantage of the polynomial model specification is that it can combine situations in which some of the independent variables are non-linearly related to Y while others are linearly related to Y.
- Note that this advantage is shared by the Lin-Log model and the reciprocal specification.

known theory or widely known
 theory
 [Frequent signs]
 narrow but significant area
 different pattern \rightarrow Adjective
 theory \rightarrow current use/Adjective
 word that is
 ↪

best formulae are often
 used you don't need anything
 else

formula
 same

- Choice of Functional Form
- The underlying theory and/or practical knowledge of the problem at hand can often help decide on the choice(s) of functional form.
 - The user aspects to consider include:
 - The level of statistical significance of the parameters in the alternative models (i.e. functional forms)
 - The signs of the parameter estimates being consistent with theory and/or practical knowledge of the problem
 - The R^2 or, better yet, the adjusted R^2 as long as the dependent variable is the same across model choices
 - Some use a residual analysis technique to ensure selection of an appropriate functional form (do)

Thus

$y_i - \hat{y}_i \rightarrow n-k$ degrees of freedom
 $\hat{y}_i - \bar{y} \rightarrow k-1$ degrees of freedom
 $n-k \rightarrow n-1$ degrees of freedom
 random error introduced in calculation
 something to do w/ # of indiv
 theory \rightarrow need to remember df
 What are degrees of freedom?

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

- Notice that the degrees of freedom (DF) are additive, i.e.: $DF(TSS) = DF(ESS) + DF(RSS)$
- The ESS has $k-1$ degrees of freedom
- The RSS has $n-k$ degrees of freedom
- The TSS has $n-1$ degrees of freedom
- $TSS = RSS + ESS$, where
- Also recall that

F Test

proportion of vary not explained
 after ESS + sum ESS \rightarrow why
 $RSS: E(y_i - \hat{y}_i)^2$
 $ESS: E(\bar{y} - \hat{y})^2$
 $TSS: E(y_i - \bar{y})^2$

- Recall that the TSS (Total Sum of Squares) is a measure of the total amount of variation in Y
- The ESS (Explained Sum of Squares) measures the amount of variation in Y that is explained by the regression model
- The RSS (Residual Sum of Squares) is a measure of the amount of variation in Y that is not explained by the regression model
- The greater the proportion of the variation in Y not explained by the model
- The larger the proportion of the variation in Y not explained by the model
- The ESS ($E(\hat{y}_i - \bar{y})^2$) is a measure of the variation in Y that is explained by the model
- The RSS ($E(y_i - \hat{y}_i)^2$) is a measure of the variation in Y that is not explained by the model
- The proportion of the variation in Y that is explained by the model is called the coefficient of determination, denoted R-squared.

F-Test

AAEC 6610
 Quantitative Techniques
 in Agricultural Economics
 Chapter 8
 F Tests

Other types

- There are other types of F tests that are useful in multiple regression analysis;
- Joint tests on several regression coefficients specifically:
- Tests involving linear functions of the regression coefficients
- Tests involving different regression equations

Other Types of F Tests

- "Significance"*
- equation of F test*
- The F-test is carried out by comparing the F-statistic (i.e., value) calculated for a given model with a "critical" F-table value (example, logic)
 - Computer programs provide F* and the exact probability (i.e., level of statistical significance) with which it allows to reject H₀
 - Under H₀, if all four major OLS assumptions are met and ϵ is normally distributed, F* follows an F-distribution with (k-1) and (n-k) degrees of freedom
 - The F-test is calculated as:
- $$F^* = \frac{ESS/k-1}{RSS/(n-k)}$$
- *Handwritten notes*

- multiple joint effects*
- out not indep*
- cur jointy have effect*
- out not indep*
- multiple cell \downarrow S_c: would be (w/ success)*
- cur that have effect on but*
- don't know what happens.*
- multiple joint effects*
- one of the independent variables included in the model affects Y (back to residual analysis)*
- The alternative hypothesis H_a is that at least one of the independent variables included in the model affects Y
 - Thus, if H₀ is not rejected, the model as a whole can not be said to be useful to explain Y statistically shown to have an effect on Y
 - H₀ means that none of the independent variables included in the model can be statistically shown to have an effect on Y
 - $H_0: B_2=B_3=\dots=B_k=0$
 - The F-test is used to evaluate the null hypothesis that all of the model's parameters, with the exception of the intercept, are zero:

F Test

some
 $\hat{U}_R^2 = U_R^2 \Rightarrow$ all parameters
 $[U_{rest} + 2 U_{unrest}]$

of restrictions ! same # less
 KFC rule: # of param. <= # of restr.

$U_R^2 = \text{original model}$

MDN w/ tests

looks similar stuff \rightarrow no significant

x_5, x_6, x_7

\hookrightarrow sufficiency of var \rightarrow (out)on; dummy variable

large # dataset \rightarrow large # exp var

This type of test is used to evaluate if a sub-

group of explanatory variables as a whole

explains some of the variation observed in the dependent variable (examples)

Consider the following, to be called the unrestrictive (UR) model, since no assumptions have been made about any of its coefficients:

$\hat{Y} = B_1 + B_2 X_2 + B_3 X_3 + \dots + B_k X_k + u$

Suppose that we want to test if a subset of q ($< k$) of the regression coefficients is jointly equal to zero

To do so, let's rewrite the unrestricted model as:

$\hat{Y} = B_1 + B_2 X_2 + \dots + B_{k-q+1} X_{k-q+1} + \dots + B_k X_k + u$

- To test this H_0 we need to estimate both the unrestricted (UR) and the restricted (R) models (RSS_R) must be equal to greater than the unrestricted (RSS_U) model (RSSUR) Notice that the RSS of the unrestricted model would have little effect on the explanatory power of the model and the RSS_R should be only slightly higher than the RSS_U .
- If H_0 is correct, dropping the q variables would have little effect on the RSS_R and the RSS_R would be equal to the RSS_U .
- To test this H_0 we need to estimate both the unrestricted (UR) and the restricted (R) models (RSS_R) must be equal to greater than the restricted model (RSSUR) Notice that the RSS of the restricted model would have little effect on the explanatory power of the model and the RSS_R should be only slightly higher than the RSS_U .

Joint Tests on Several Coefficients

- The null hypothesis is that at least one of those coefficients does not equal zero
- The alternative hypothesis is that at least one of those coefficients does not equal zero
- The null hypothesis is:

$$H_0: B_{k-q+1} = B_{k-q+2} = \dots = B_k = 0$$
- The last q coefficients are all equal to zero, the correct model will be the restricted (R) model:

$$\hat{Y} = B_1 + B_2 X_2 + \dots + B_{k-q} X_{k-q} + u$$
- If the last q coefficients are all equal to zero, the correct model will be the restricted (R) model:

$$\hat{Y} = B_1 + B_2 X_2 + \dots + B_{k-q+1} X_{k-q+1} + \dots + B_k X_k + u$$
- To do so, let's rewrite the unrestricted model as:

$$\hat{Y} = B_1 + B_2 X_2 + \dots + B_{k-q+1} X_{k-q+1} + \dots + B_k X_k + u$$

Joint Tests on Several Coefficients

- This type of test is used to evaluate if a sub-group of explanatory variables as a whole explains some of the variation observed in the dependent variable (examples)
- Consider the following, to be called the unrestrictive (UR) model, since no assumptions have been made about any of its coefficients:
- Suppose that we want to test if a subset of q ($< k$) of the regression coefficients is jointly equal to zero
- To do so, let's rewrite the unrestricted model as:

$$\hat{Y} = B_1 + B_2 X_2 + B_3 X_3 + \dots + B_k X_k + u$$
- Consider the following, to be called the unrestrictive (UR) model, since no assumptions have been made about any of its coefficients:
- Suppose that we want to test if a subset of q ($< k$) of the regression coefficients is jointly equal to zero
- This type of test is used to evaluate if a sub-

Joint Tests on Several Coefficients

loss > error

Computer is not different

curve of \hat{y}_i
instead of y_i

$$\text{loss } \hat{y}_i - y_i x_{i1} = B_1 + B_3 (x_{i2} - \bar{x}_{i2})$$

$$\text{loss } \hat{y}_i = B_1 + B_2 x_{i2} - B_3 x_{i3} + B_4 x_{i4}$$

$$\text{loss } \hat{y}_i = B_1 + B_2 x_{i2} + B_3 x_{i3}$$

(impose)

$$\text{loss } \hat{y}_i = B_1 + B_2 x_{i2} + B_3 x_{i3}$$

if $B_2 + B_3 = 1$ then same CTS (CTS)

$$\hat{y}_i = c + B_1 x_{i2} + B_2 x_{i3}$$

- Recall that in a Cobb-Douglas production function the sum of the (exponentially) coefficients indicates the returns to the scale of production, i.e. the % change in production if the use of all inputs is increased by 1%.
- Production economists often want to test the hypothesis of constant returns to scale, i.e. that the sum of those coefficients equals one
- This is an example of an F-test involving linear functions of the regression coefficients (derive the restricted model for this example)

F Tests Involving Linear Functions of the Regression Coefficients

October 4

- Notice that this type of F test is not equivalent to conducting individual t-tests for the statistical significance of each of the variables involved because of multicollinearity, it is possible that all t-tests result in significant while the joint F-test allows for the rejection of Ho (i.e. it results in significant) - F-test is imprecise to multicollinearity
- Because of multicollinearity, it is possible that all t-tests result in significant while the joint F-test allows for the rejection of Ho (i.e. it results in significant) - F-test is imprecise to multicollinearity
- Notice that this type of F test is needed for this test to be valid
- As in the general F test, the four major OLS assumptions plus error term normality are needed for this test to be valid
- Notice that this type of F test is not equivalent to conducting individual t-tests for the statistical significance of each of the variables involved because of multicollinearity, it is possible that all t-tests result in significant while the joint F-test allows for the rejection of Ho (i.e. it results in significant) - F-test is imprecise to multicollinearity

see also

not so en f test
 F test + t test Multicollin \rightarrow use dt

t test \rightarrow not significant

$$\begin{aligned} \text{loss } (n-k) &= n - k + \text{loss} \\ \text{loss } (n-k) &= \text{loss} - \cancel{(n-k)} \\ \text{loss} &= \cancel{(n-k)} \end{aligned}$$

test in \rightarrow loss is significantto loss \rightarrow loss is sameif loss \rightarrow loss then real $\alpha_f = \# \text{ of restrictions}$

Mechanics on the same

- The test statistic for $H_0: B_{k+1} = B_{k+2} = \dots = B_q = 0$, is given by $F^* = ((RSS_R - RSS_U)/q) / ((RSS_U)/(n-k))$
- If H_0 is correct F^* will follow an F distribution with q degrees of freedom in the numerator and $n-k$ in the denominator
- Therefore, the rest of the test is carried out as in the case of the basic F test for the significance of the regression model as a whole, i.e. by comparing F^* with the critical F table value of the regression model.
- Notice that this type of F test is not equivalent to conducting individual t-tests for the statistical significance of each of the variables involved because of multicollinearity, it is possible that all t-tests result in significant while the joint F-test allows for the rejection of Ho (i.e. it results in significant) - F-test is imprecise to multicollinearity

Joint Tests on Several Coefficients

dry fruits \rightarrow use same product function
different curves! \rightarrow different functions
time series \rightarrow parabolic
ex: data over time into 2
deep learning architecture
devars should be closely rel +
can have 2 models
two sep model or can use the same

- These two models might have a different number of variables but closely related, and their independent observations, their dependent variables have to be same values (examples)
- For this (Chow) test, consider these two models:

$$\begin{cases} Y_j = \alpha_1 + \alpha_2 X_{j1} + \alpha_3 X_{j2} + \dots + \alpha_n X_{jn} + u_j & (j=1, \dots, m) \\ Y_i = B_1 + B_2 X_{i1} + B_3 X_{i2} + \dots + B_n X_{in} + u_i & (i=1, \dots, n) \end{cases}$$
- Sometimes one wants to test if the same model applies to two somewhat different data sets or if two separate models are needed (one for each)
- Sometimes one wants to test if two separate models must be the same, but not necessarily take the same values

F Tests for the Equality of Coefficients of Different Regression Equations

param
in MLST which only use one
each layer has each linear function

- Notice, however, that each linear function implies one restriction only, regardless of the number of coefficients involved in the function
- Cobb-Douglas example discussed before

$$Y = B_1 + B_2 X_1 + B_3 X_2 + \dots + B_n X_n + u$$
- The concepts can be easily extended to more restrictions involving linear functions of more than two coefficients ($B_2 + B_3 + B_4 = R$; such as the Cobb-Douglas example discussed before)
- The F* statistic in all of the previous cases is the same as before:

$$F^* = \frac{(RSSR - RSSU)/q}{(RSSU)/(n-k)}$$

same as saying 2 param are =

- Another example of an F-test involving linear functions of the model coefficients involves the restriction that $B_2 - B_3 = 0$ (i.e., that $B_2 = B_3$)
- The unrestricted and restricted models would be:

$$\begin{cases} Y = B_1 + B_2 X_2 + B_3 X_3 + \dots + B_n X_n + u & (UR) \\ Y = B_1 + B_2(X_2 + X_3) + \dots + B_n X_n + u & (R) \end{cases}$$
- Alternatively, consider:

$$\begin{cases} Y = B_1 + B_2 X_2 + B_3 X_3 + \dots + B_n X_n + u & (UR) \\ Y = B_1 + B_2 X_2 + B_3 X_3 + \dots + B_n X_n + u & (R) \end{cases}$$
- Make sure that you know how estimate all these, which implies that $B_2 + B_3 = 0$ (i.e., that $B_2 = -B_3$)
- Restricted models!

- $\text{df}_{\text{sum}} \text{ not } \rightarrow \text{use dummy var.}$
- $\text{intervened step} \rightarrow \text{allow some steps}$
- $\text{want to NOT reject}$
- $\text{reject null} \rightarrow \text{use step.}$

y_5 is after when before
we split data

- Note that, with the use of dummy variables, one could still estimate a single model where some of the coefficients are not equal pair-wise
- If H_0 is rejected the data can not be pooled and two separate regressions must be estimated
- $n-m-k = \{(n-k)+(m-k)\} = n+m-k-n-m+2k = k$ = the number of restrictions imposed on the numerator equals $\text{df}(RSSR) - \text{df}(RSSUR)$
- Notice that the degrees of freedom (df) for the numerator equals $\text{df}(RSSR) - \text{df}(RSSUR) = (n-m-k) - \{(n-k)+(m-k)\}$

$$\text{df}_{\text{sum}} = n+m-k - (n-k) - (m-k)$$

- In our F testing framework the RSS of this "merged" model is considered the RSSR
- The previously advanced F^* statistic is also applicable for this test with the appropriate degrees of freedom (k and $n-m-2k$)
- $F^* = ((RSSR-RSSUR)/k) / ((RSSUR)/(n+m-2k))$

- Here H_0 is that these two models are identical (i.e. that their coefficients are equal, pair-wise) and thus can be merged into a single model
- Has at least one pair of coefficients are not equal
- The unrestricted model consists of the two separate equations, with $\text{RSS}_1 + \text{RSS}_2$ (the sum of the RSS of the two separately estimated equations) the degrees of freedom of RSSR is the sum of the degrees of freedom of RSS1 and RSS2, i.e. $(n-k)+(m-k) = m+n-2k$

Salt column for 4 seasons

Salt religion? → need 5 dum vars

- In general, this type of independent variables have to be modeled through dummy variables
- A set of dummy variables is created for each categorical independent variable X in the model, where the number of dummy variables in the set equals the number of categories in which that independent variable can be classified
- How many categories in each of the previous examples?

Use of Dummy Variables

- Month explaining cumulative rainfall
- Month explaining food availability in forest
- State of residence explaining party affiliation
- Season explaining demand for electricity
- Breed explaining dogs' intelligence
- Religious affiliation explaining income
- Examples include:

- This means that they can only take on few values often with no natural/logical ordering
- In many models, one or more of the independent variables is qualitative (i.e., categorical) in nature

Use of Dummy Variables

Dummy Variable Regression Models Chapter 9

in Agricultural Economics
Quantitative Techniques
AAEC 6610

act 4

$\text{B}_1 + \text{B}_3 - \text{B}_1$
difficult fields below vent + L + 3
B3 diff b/w vent + 3 & vent + 1
fields b/w y1 + y2
2x is at field LM
[S+II study this]
compest intercept

survive
survive
survive

- For this year's yields and the 1st variety: $y_i = B_1 + B_2 X_{i1} + B_3 X_{i2}$
- For this year's yields and the 2nd variety: $y_i = (B_1 + B_2) + B_3 X_{i1} + B_6 X_{i5}$
- For this year's yields and the 3rd variety: $y_i = (B_1 + B_3) + B_5 X_{i1} + B_6 X_{i5}$
- For last year's yields and the 1st variety: $y_i = (B_1 + B_2) + B_3 X_{i1} + B_6 X_{i5}$
- For last year's yields and the 2nd variety: $y_i = (B_1 + B_3) + B_5 X_{i1} + B_6 X_{i5}$
- For last year's yields and the 3rd variety: $y_i = (B_1 + B_2 + B_3) + B_3 X_{i1} + B_6 X_{i5}$
- For this year's yields and the 2nd variety: $y_i = (B_1 + B_2 + B_3) + B_5 X_{i1} + B_6 X_{i5}$
- For this year's yields and the 3rd variety: $y_i = (B_1 + B_2 + B_3) + B_5 X_{i1} + B_6 X_{i5}$
- For last year's yields and the 1st variety: $y_i = B_1 + B_5 X_{i1} + B_6 X_{i5}$
- For last year's yields and the 2nd variety: $y_i = (B_1 + B_2) + B_5 X_{i1} + B_6 X_{i5}$
- For last year's yields and the 3rd variety: $y_i = (B_1 + B_2 + B_3) + B_5 X_{i1} + B_6 X_{i5}$

why
have to drop our cat

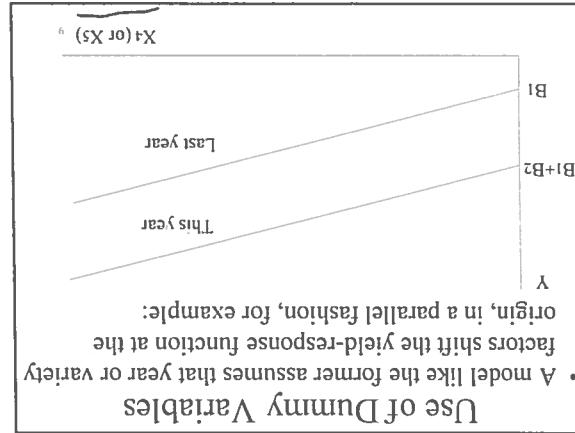
- Notice that the dummy variables corresponding to one of the categories of X_2 and X_3 (D_{22} and D_{33}) were excluded from the model to avoid perfect multicollinearity. Any one dummy category can be included, it makes no difference.
- Notice that this model actually estimates a different intercept for each year-variety combination, while maintaining the same slope parameters for all of the other independent variables in the model (X_4 and X_5).
- Notice that this model actually estimates a different intercept for each year-variety combination, while maintaining the same slope parameters for all of the other independent variables in the model (X_4 and X_5).

ok

$$\begin{aligned}
 D_{31} &= 1 - D_{21} - D_{33} \\
 D_{21} &= 1 - D_{22} \\
 D_{22} &= 1 - D_{31} \\
 D_{32} &= 1 - D_{31} \\
 D_{33} &= 1 - D_{21} - D_{22}
 \end{aligned}$$

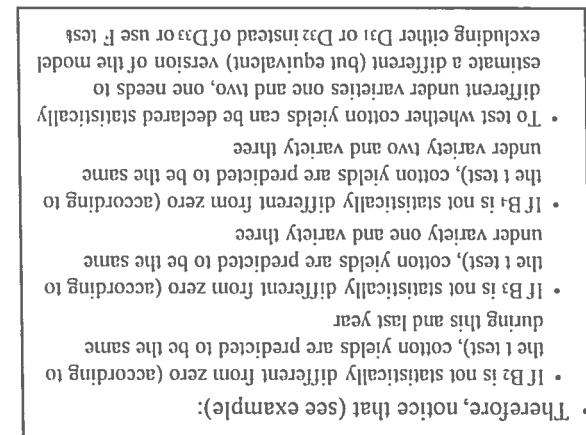
- For example let Y_i be cotton yields in farm i , X_3 indicates whether yields are from the current or the previous year (two categories), X_2 indicate which of three varieties was planted (three categories), and X_4, \dots, X_{10} be other variables that affect farm yields and D_{22} and three for X_3 (D_{31}, D_{22} and D_{33}). Two dummy variables will be created for X_2 (D_{21} and D_{32}) and D otherwise; $D_{32} = 1$ if X_2 and X_3 were from the same year and 0 otherwise; $D_{21} = 1$ if X_2 and X_3 were from different years and 0 otherwise; $D_{31} = 1$ if the 2nd variety was planted and 0 otherwise, and $D_{33} = 1$ if the 3rd variety was planted and 0 otherwise (see dataset).

assume that yr
varieties don't affect
responses less than 15%

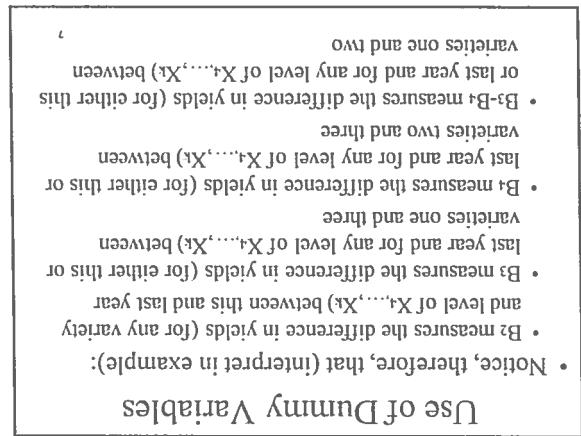


- A model like the former assumes that year or variety factors shift the yield-response function at the origin, in a parallel fashion, for example:

Use of Dummy Variables

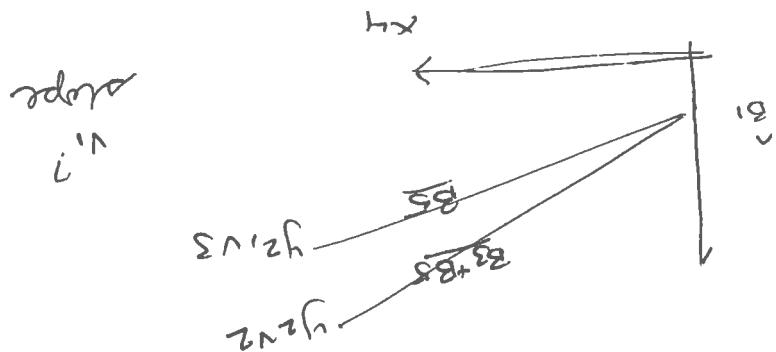


- To test whether cotton yields can be declared statistically under variety two and variety three
- If B_2 is not statistically different from zero (according to the F test), cotton yields are predicted to be the same under variety one and variety three
- If B_3 is not statistically different from zero (according to the F test), cotton yields are predicted to be the same during this and last year
- Therefore, notice that (see example):



- Notice, therefore, that (interpret in example):

Use of Dummy Variables



Yield Response to Fertilizer Use $X_4 (X_5 \text{ constant})$	
B_1	Slope = B_1
$B_1 + B_2$	Slope = $B_1 + B_2$
$B_1 + B_2 + B_3$	Slope = $B_1 + B_2 + B_3$
This year, Variety 2 Last year, Variety 3 Used as both a slope and an intercept shift	

- $B_3 + B_2$ is the X_4 slope when variety two is planted
- $B_3 + B_2 + B_1$ is the X_4 slope when variety one is planted
- Notice that the same categorical variable can be used as both a slope and an intercept shift
- When X_4 -fertilizer use-changes by one unit) when variety three is planted (why?)

allow slope rel to same var
such as year or variety can cause "slope shifts"
For example, consider the following model:

$$Y_i = B_1 + B_2 D_{i1} + B_3 D_{i2} + B_4 D_{i3} + B_5 X_{i4} + B_6 X_{i5} + e_i$$
 such as year or variety (or in addition), categorical factors such as year or variety can cause "slope shifts"

• For that reason, the dummy variables in models like the former are often called "intercept shifters".
• Alternatively (or in addition), categorical factors such as year or variety can cause "slope shifts".
• Here, the year effect is accounted for by an intercept shift, as before, but the variety effect is considered through an X_4 slope shift (take derivative).
• B_3 is the X_4 slope (it measures the change in yields through an X_4 slope shift) (take derivative)
• Notice that the same categorical variable can be used as both a slope and an intercept shift
• When X_4 -fertilizer use-changes by one unit) when variety three is planted (why?)

AAC 6610 Quantitative Techniques in Agricultural Economics

Chapter 10 Multicollinearity

- When two or more independent variables in a regression model are highly correlated to each other, it is difficult to determine if each of these variables, individually, have an effect on Y, and to quantify the magnitude of that effect.
- Intrinsically, for example, if all farms in a sample that use a lot of fertilizer also apply large amounts of pesticides (and vice versa), it would be hard to tell if the observed increase in yields is due to higher fertilizer or to higher pesticide use.

Multicollinearity

- In economics, when (nominal) interest and inflation rates are used as independent variables, it is often hard to quantify their individual effects on the dependent variable because they are highly correlated to each other.
- Mathematically, this occurs because, every time a given OLS parameter estimate is more highly correlated to the other independent variable than to a given constant, the standard error associated with that being constant will be higher if the corresponding independent variable is more highly correlated to the other independent variable in the model.

Multicollinearity

multicollinearity
when two or more independent variables in a regression model are highly correlated to each other

get var you think is important
 your will double if 3
 $y = x_1 + x_2 + x_3$ higher
 of variance
 effect or see if it's effect
 [under variation yes needed]
 not get a sc of

$\sum_{i=1}^n (x_i - \bar{x})^2$ is regression
 \rightarrow $(x_i - \bar{x})^2$ residual

- Multicollinearity
 - From the formulas for computing the standard error estimates in the two independent variable case:
 - $\hat{V}[B_1] = \frac{\sigma^2}{\sum_{i=1}^n (1-x_{1i})^2 (X_{2i}-\bar{X}_2)^2}$
 - $\hat{V}[B_2] = \frac{\sigma^2}{\sum_{i=1}^n (1-x_{2i})^2 (X_{1i}-\bar{X}_1)^2}$
 - where r_{12} is the correlation coefficient between X_1 and X_2 , note that:
 - The larger r_{12} , the larger the standard errors of the dependent variable do not result statistically significant when conducting the basic t-tests.
 - It could be the reason why independent variables that are believed to be key in determining the value of the independent variable do not result statistically significant when comparing the basic t-tests.
 - This potential problem is known as multicollinearity.

- Multicollinearity
 - It is more common in time-series data models because time often affects the value taken by many of the independent variables in these types of models because they are highly correlated to each other.]
 - Undesirable characteristic of the data violation of an OLS assumption, but rather an assumption of an OLS specification or a perfect multicollinearity occurs when there is a perfect linear correlation between two or more independent variables, i.e. when an independent variable is actually an exact linear function of one or more of the others; for example:
 - When one of the dummy variables in a set is not omitted, because the value of one of these variables always equals one minus the sum of the others]
 - When an independent variable takes a constant value in all observations during each of the seasons as independent variables in a time series model during including total annual rainfall as well as rainfall one minus the sum of the others]
 - When an independent variable is a set is not omitted, because the value of one of these variables always equals one minus the sum of the others]

- Perfect Multicollinearity
 - Perfect multicollinearity occurs when there is a perfect linear correlation between two or more independent variables, i.e. when an independent variable is actually an exact linear function of one or more of the others; for example:
 - When an independent variable takes a constant value in all observations during each of the seasons as independent variables in a time series model during including total annual rainfall as well as rainfall one minus the sum of the others]
 - When an independent variable is a set is not omitted, because the value of one of these variables always equals one minus the sum of the others]
 - When an independent variable is a set is not omitted, because the value of one of these variables always equals one minus the sum of the others]

Perfect Multicollinearity

$\Delta \text{out} + \text{se} \Delta \text{out}$
 take very out toward parameter
 Δout small # absr.

but unhappy w/ what get? ✓
 which all will recalc ~~these~~ set
 one signif; don't want issue
 if se row + all est

redundant info
 is a mistake! included

- Some independent variables from the regression or become statistically significant when excluding some independent variables from the regression
- Parameter estimates drastically change values and statistical significance tests **change**
- High R^2 , highly significant according to the F test, but few or no statistically significant variables are explained by the model's dependent variable are not explaining the model(s) considered critical in problem:
- Independent variable(s) considered critical in key
- The following, when taken together, are considered symptoms of a multicollinearity

Symptoms of Multicollinearity

- Multicollinearity is considered severe and becomes a problem when this correlation is high and interferes with the estimation of the model's parameters at the desired level of statistical certainty
- A certain degree of correlation between the independent variables is normal and expected in most cases

Severe Multicollinearity

- All cases of perfect multicollinearity are the result of making a mistake when specifying the model and can be easily corrected by properly specifying the model (i.e., excluding redundant error estimates)
- If there is perfect multicollinearity, the OLS method can not produce parameter or standard error estimates

Perfect Multicollinearity

multiple to scale A
 true: condition # is not
 gives overall read
 buying + loses + gain
 more correlated the larger the
 matrix will have larger value
 how covariation: any effect
 shows condition # effects

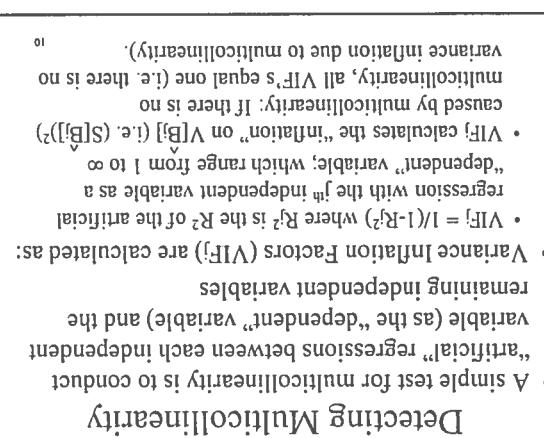
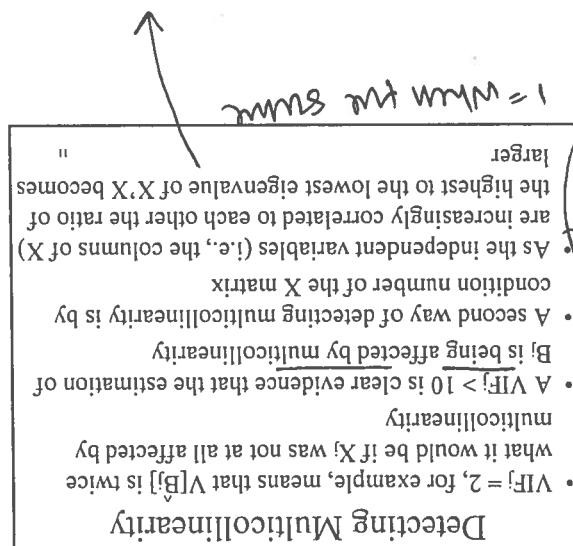
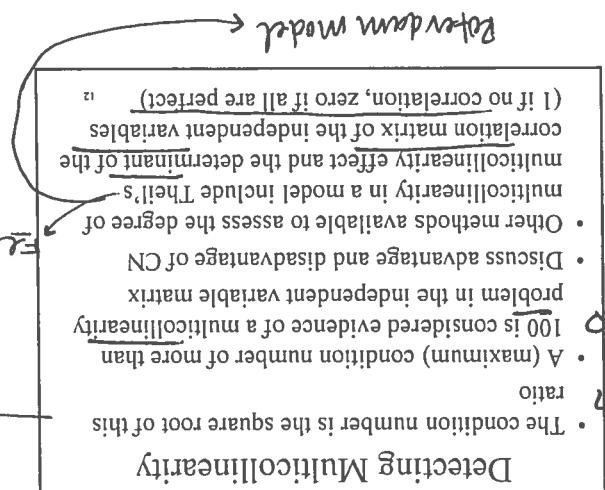
there is a
 from multiple sources but all
 see a significant factor
 $(1 - R^2) \rightarrow$ factor in our
 what percent does it in our
 for in this
 $R^2 = 0.5$
 what would say ~ 10 and
 $-VIF = 2$, our covars B_1 is twice

all other explain $VIF = 12$
 est regression w/ no drop +
 $VIF_{x2} = 1/(1 - R^2_{x2})$ is all else

one for each explained.

samples? $\rightarrow VIF$

only first few variables
 and second few
 could have 2
 same sample



Addressing Multicollinearity

• Although it is useful to be aware of the presence of multicollinearity (why?), it is not easy to remedy severe (non-perfect) multicollinearity.

• If possible, adding observations or taking a new sample might help lessen multicollinearity.

• Excluding the independent variables that appear to be causing the problem (can use VIF to identify).

• However, these might be variables that are of interest to the researcher ~~secondarily~~ ~~secondarily~~ with make the OLS estimates biased relevant independent variables from the model will even if this is not the case, recall that omitting

• Even if this is not the case, recall that omitting irrelevant independent variables from the model will not worry about it as long as the statistical differences are not disappearing.

real data only for analysis → ~~curly~~

tried up: so when used non
more more multicoll

Addressing Multicollinearity

- Remember that there is usually a certain degree correlation across the X's
- Using ~~real~~ instead of nominal economic data sometimes helps, for example:
- Using a polynomial instead of a polynomial specification on a given independent variable forms might reduce the level of overall linear when justified, using various non-linear functional forms might reduce the level of overall linear
- Remember that there is usually a certain degree of multicollinearity in any model, and one should not worry about it as long as the statistical differences are not disappearing.

Modifying the model specification can sometimes help, for example:

Key point: what kind of α ? \leftarrow resulting from
to full to reject incorrect
reject against true of heteroske

$H_0: \text{homosked}$

$H_A: \text{heteroske}$

conslst w/ α^2 \leftarrow don't reject H_0

conslst w/ α^2 \leftarrow 25% prob
of having
conslst w/ α^2 \leftarrow 25% prob
use high $\alpha \rightarrow 25\% \rightarrow 1 - \alpha$

- careful w/ atm vir \rightarrow don't atm.

- similar to func form

- org $y_i(u_i)$ term squares + axpnd \leftarrow inclp

Residual \leftarrow resid^2 's sum/dp

+ comput resid [squared]

ALWAYS have whic test \rightarrow est model

[assumption iid]
[computers]
delay is const
nowhere + nowhere (id)
 $\text{if } \text{var } u = \sigma^2 I$
Heteroske.
 $\text{BASE} \leftarrow (\mathbf{x}' \mathbf{x})^{-1} \mathbf{x}' \mathbf{u}$
 $\text{Var } \hat{\mathbf{B}} = E[(\mathbf{x}' \mathbf{x})^{-1} \mathbf{x}' \mathbf{u} \mathbf{u}' (\mathbf{x}' \mathbf{x})^{-1}]$
Find $\text{var } \hat{\mathbf{B}} = (\mathbf{x}' \mathbf{x})^{-1} \mathbf{x}' \mathbf{u}$
 $\text{var } \hat{\mathbf{B}} = E[\hat{\mathbf{B}} - E(\hat{\mathbf{B}})]^2 (E[\hat{\mathbf{B}} - E(\hat{\mathbf{B}})])'$
standard error est same bias

structure excluded \leftarrow minus one
the number of parameters in the auxiliary regression
variable with p degrees of freedom ($\chi^2_{(p)}$), where p is
regression is distributed as chi-square random

\Rightarrow Under H_0 : Homoskedasticity, n times the R^2 of that
and exclude any redundant right-hand-side variables
This "auxiliary" regression must include an intercept
litter squares, did cross products as the independent
dependent variable versus the explanatory variables,
Regression the squared OLS residuals as the
known as the White test, which is conducted by:
The most common test for heteroskedasticity is

Tests for Heteroskedasticity

[No longer BLUE]
NBUS but not uncorr
use OLS but not OLS but
most efficient (i.e., minimum variance), even if the
error term is normally distributed

- Also, the OLS parameter estimates are no longer the
intercept that uses the OLS standard error estimates
will be incorrect on the average
- This means that any statistical test of confidence
and discusses correct way to compute OLS variance
- The OLS standard error estimates are still unbiased, but
constant variance across observations (see handout)
- Heteroskedasticity occurs when the error term (and
thus the dependent variable Y) does not have a
constant variance across observations (see handout)

Heteroskedasticity

AAEC 6610
in Agricultural Economics
Quantitative Techniques
Chapter 11
Heteroskedasticity
and Generalized Least Squares

↑ incorrect accept of null
↓ a type 2 error
↓ /

require more power to detect a difference in white wine

Student work: The wine taste test

suspects

$$t_{\text{test}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 / n_1 + s_p^2 / n_2}}$$

$$t_{\text{test}} = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

null: there is no difference in average taste

exp variance in white wine

exp variance in white wine

OLS residual $\hat{e}_i^2 = \bar{e}^2$

diff regression but ΔY

now compare to compare

now

can take out from heteroscedasticity

$$\text{const. } \ln y_i + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \dots + \beta_n \ln x_n$$

$$\text{of } y_i + \dots + y_n$$

$$H = \text{some linear } \rightarrow \ln y_i = \beta_0 + \beta_1 \ln x_1 + \dots + \beta_n \ln x_n$$

and multivar.

same $f(\cdot)$ of answer + (exp out)

the more restricted form of H

Performance: do both tests

white is general

an null hypothesis w/o heterosce

viel assume $1/2$ stronger rejection

null in white test

can be other reasons why reject

disadvantage

As usual, H_0 is tested by comparing the computed statistic with the table value of $\chi^2_{(1)}$ at the desired level of significance. $t_{\text{test}} > t_{\text{crit}}$

$\chi^2_{(1)}$ is the dependent variable (where $\chi^2 = \sum_i (y_i - \hat{y}_i)^2 / \sigma_i^2$) and the independent variables in χ^2 , as the explanatory variables (y is the dependent variable) is the explained sum of squares of a regression with

H_0 , $\text{BTS} = \text{ESS}_{AR/2}$ follows a $\chi^2_{(S-1)}$, where $\text{ESS}_{AR/2}$ is the multiple test

• The Breusch-Pagan is an older, longer range

Tests for Heteroskedasticity

all y except y_1 and y_2

$y_i = 0$, where y_i includes all other parameters in H_0 : hypothesis of homoskedasticity is equivalent to H_0 :

H_0 : Since an intercept (y_1) is included in χ^2 , the null hypothesis of homoskedasticity is equivalent to H_0 :

$\chi^2 = \sum_i (y_i - \hat{y}_i)^2 / \sigma_i^2$, etc.

Possible functional forms for $\sigma_i^2 = h(z_i)$ include:

H_0 : In this test, the specific alternative hypothesis of heteroskedasticity is $H_a: g_1 = h(z_1)$ (expanded), where g_1 is any possible nonlinear function of z_1 , z_1 is an imprecise, and y is an $(S \times 1)$ vector of parameters

multivariable nonlinearity function of z_1, z_2 , etc.

• The Breusch-Pagan is an older, longer range

Tests for Heteroskedasticity

first two major OLS assumptions hold

heteroskedasticity if the researcher is confident that the H_0 , it can only be used as a definite test for

for these other conditions:

H_0 (although it should be noted that there are better tests for correlation between the independent variables and the error term, could also cause rejection of

misspecification of the random component (exclusion of relevant explanatory variables or use of incorrect function form) or correlation between the independent variables and the error term, could also cause rejection of

rejecting H_0 , indicates likelihood of heteroskedasticity

It is important to note that, in the White test, the objective should be not to reject at 0.20 or so

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

level of significance (briefly discuss selection of a -

8

Least squares fit $\hat{y}_i = B_0 + B_1 x_i + \epsilon_i$ for all i
 $\sum \epsilon_i = 0$ and $\sum \epsilon_i^2$ is minimum
 $\text{Var}[\epsilon_i] = \sigma^2$ for all i
 After terms + n residuals
 $y_i = B_0 + B_1 x_i + \epsilon_i$ $\text{Var}[y_i] = \sigma^2$
 $B_0 = \bar{y}$, $B_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$
 $\text{Var}[B_0] = \sigma^2 \frac{1}{n}$, $\text{Var}[B_1] = \sigma^2 \frac{1}{\sum (x_i - \bar{x})^2}$
 Therefore estimate by the left:
 $\hat{y}_i = \bar{y} + B_1 x_i + \epsilon_i$ $\text{Var}[\hat{y}_i] = \sigma^2$
 can be extra var.
 our compound error
 $y_i = \text{depand var} \rightarrow \text{Z of obs result} = \text{est res}$
 $y_i^2 = \text{depand var} \rightarrow \text{Z of obs result} = \text{covars}$
 some can bring the less does not have
 standard devps. ! just could loves + loves + my self
 weill fix stds
 to signs same \leftarrow do it

- Generalized Least Squares (GLS)
 - Generalize regression models when the error term is not i.i.d., i.e. when $\text{Var}[U] = \sigma^2 I \neq \sigma^2 I_n$
 - It is based on the fact that if ψ is a positive definite symmetric matrix (as it is the case for all proper covariance matrices), there exist smoother (nrxn) matrix P such that $P\psi P^T = I_n$
 - Therefore, if $\text{Var}[U] = \sigma^2 \psi$, $\text{Var}[PU] = P\text{Var}[U]P^T = \sigma^2 P\psi P^T = \sigma^2 I_n$
 - Also note that $\psi = P^{-1}(P)$, and thus $\psi^{-1} = P$

Generalized Least Squares (GLS)

✓ 2018-02-22 14:47:22

- If heteroskedasticity with respect to more than one variable is discovered, one can use the **White** heteroskedasticity with respect to correct for heteroskedasticity.
- One variable is discovered, one can use the **WLS** regression to obtain estimates of the error term variances for each of the observations.
- Since the dependent variable in that regression is the squared OLS residuals (\hat{U}_j^2), the predictions from it are (proportional) estimates for the σ_i^2 's.
- Those predictions might need to be transformed in order to obtain the appropriate weights to correct for heteroskedasticity.
- Although acceptable, this is not the optimal method to correct for heteroskedasticity.

Dealing with Heteroskedasticity

Dealing with Heteroskedasticity

- If a single „suspect“ variable can be identified, the heteroskedasticity problem might be lessened by using a weighted least squares (WLS) estimator.
- In WLS, before applying OLS, all of the data is multiplied by a set of weights (w_j), where the weights are set of n values that when multiplied by the error term make it have a constant variance (i.e., $V[w_i u_i] = \sigma^2$).
- When there is only one variable (Z), involved in the heteroskedasticity, one can try to use different functions of i (such as $w_i = Z_i^2$, Z_i^3 , $1/Z_i$, $1/Z_i^2$, $\ln(Z_i)$, $\ln(Z_i^2)$, etc.).
- The residuals of the WLS regression must be re-tested before one can consider the problem solved (see example **left for $B_1 + B_2$**).

decrease Δ after P effect
Param name traversal

use white text to see if our model is better

- multicollinearity helps
 $P = \text{transformation matrix}$
 $\hat{Y} = PY + BU$ or $Y = X + U$
- thus if OLS is used to estimate parameters, it will produce inefficient parameter estimates and unbiased standard error estimates
- thus, if OLS is applied to transformed data, it will produce efficient parameter estimates and unbiased standard error estimates
- multicollinearity helps
- apply this to transformed model
- $P^{-1} = (P^T P)^{-1} P^T$
- thus sample \rightarrow not levered about difficulty: many issue is see can writing
- if the inefficiency of the OLS parameter estimates is not a concern, a simple way of dealing with heteroskedasticity is to estimate heteroskedastic-consistent standard errors, as follows:
- $\hat{H}CSE = (X^T X) Q(X^T X)^{-1}$ where Q is a matrix with zeros elsewhere (where does it come from?)
- This yields consistent estimates for the standard errors, meaning that they can be considered correct and confidence intervals if the t -tests and F -test are used for hypothesis testing.
- OLS f(s) . Predicted from white test
- as No longer effective \rightarrow OLS are incorrect
- WLS f(s)
- need est of σ^2 then apply formulae
- $\text{Var}[U] \leftarrow \text{square of residual endogenous variables}$
- $\text{symmetric matrix} \rightarrow \text{place in } Z^T Z$
- $\text{Var}[E] = (X^T X)^{-1} \text{Var}[U] (X^T X)^{-1}$
- Not unique, only consistent
- \downarrow
- OLS \rightarrow OLS residuals
- \downarrow
- consistency
- \downarrow
- OLS f(s)
-

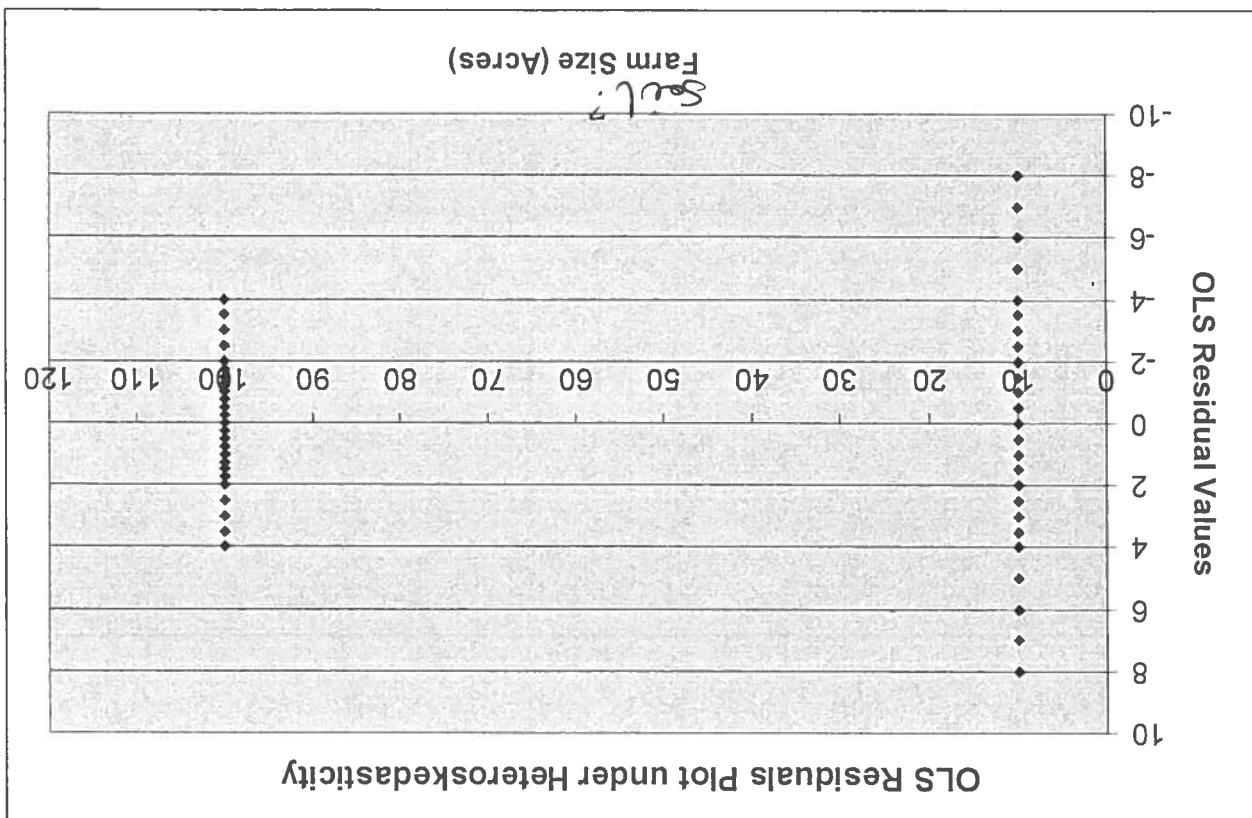
Dealing with Heteroskedasticity

- If the inefficiency of the OLS parameter estimates is not a concern, a simple way of dealing with heteroskedasticity is to estimate heteroskedastic-consistent standard errors, as follows:
- $\hat{H}CSE = (X^T X) Q(X^T X)^{-1}$ where Q is a matrix with zeros elsewhere (where does it come from?)
- This yields consistent estimates for the standard errors, meaning that they can be considered correct and confidence intervals if the t -tests and F -test are used for hypothesis testing.
- The sample size is relatively "large" (at least > 100) and building confidence intervals if the t -tests and F -test are used for hypothesis testing.

Generalized Least Squares (GLS)

- The following model has i.i.d. errors:
- Thus, if OLS is used to estimate parameters, it will produce inefficient parameter estimates and unbiased standard error estimates
- Application of OLS to this model yields the GLS estimator:
- $B = (X^T P_{XX}^{-1} X)^{-1} P_{YY} = (X^T Y - X^T \mu)^{-1}$
- $\hat{V}[B] = Q(X^T P_{XX}^{-1} X)^{-1} P_{YY}(X^T Y - X^T \mu)^{-1} Q$
- Note that the previously discussed WLS method to correct for heteroskedasticity is in fact a consequence of the OLS residuals in the diagonal of the square of the OLS residuals for the standard errors.
- Particular application of the GLS estimator to correct for heteroskedasticity is in fact a consequence of the OLS residuals in the diagonal of the square of the OLS residuals for the standard errors.

In practice, this means that, everything else being constant, random coffee yield variability is higher in smaller size farms than in large farms. If, in addition, average (i.e. expected) coffee yields increase with farm size, we would observe the following plot of yields versus farm size:



In our example, if $\sigma^2 = 16$ when farm size is 10 acres and $\sigma^2 = 4$ when farm size is 100 acres, we would observe the following when plotting the OLS residuals:

A higher error term variance means that the range of the observed errors is also higher. If, for example, the error term variance is $\sigma^2 = 16$ (and the error term distribution is close to normal) about 95% of the errors would fall within $\pm 2\sigma$ i.e. within -8 and 8. Alternatively, if the error term variance is $\sigma^2 = 4$, most of the error terms would fall within $\pm 2\sigma$ i.e. within -4 and 4.

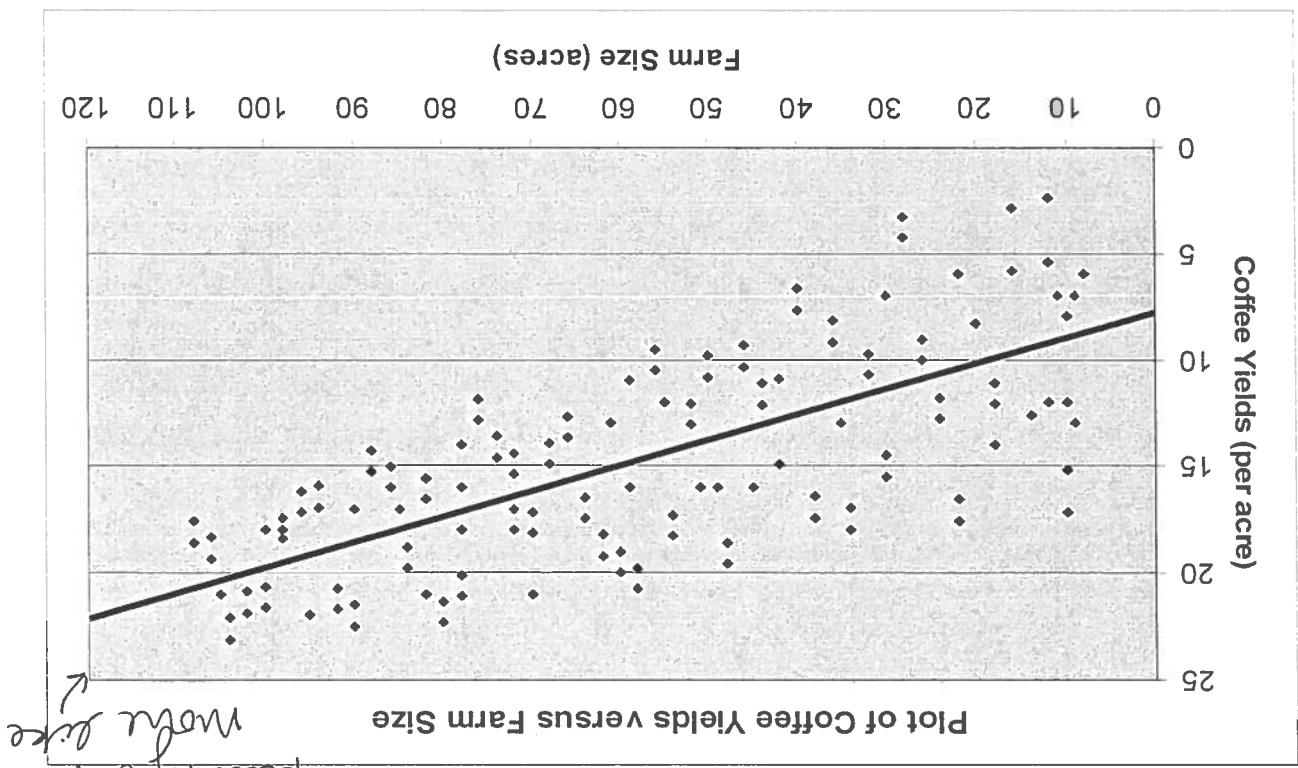
For example, in a model where the dependent variable is per-acre coffee yields and one of the explanatory variables is total farm size, the error term variance (i.e. random yield variability) might be higher when farm size is small than when farm size is large.

In the presence of heteroskedasticity the error terms corresponding to different observations have different variances. This difference in variance is usually associated with one or more of the different variables. For example, in a model where the dependent variable is per-acre coffee yields and one of the explanatory variables in the model, $y_{it} = \alpha + \beta x_{it} + u_{it}$

Heteroskedasticity most often occurs when the observations comprising the data set are taken from different units (such as farms, businesses, states, countries) but during the same time period, i.e. when analyzing cross-sectional data. However, panel and time series data might also suffer from it.

Key Points on Heteroskedasticity

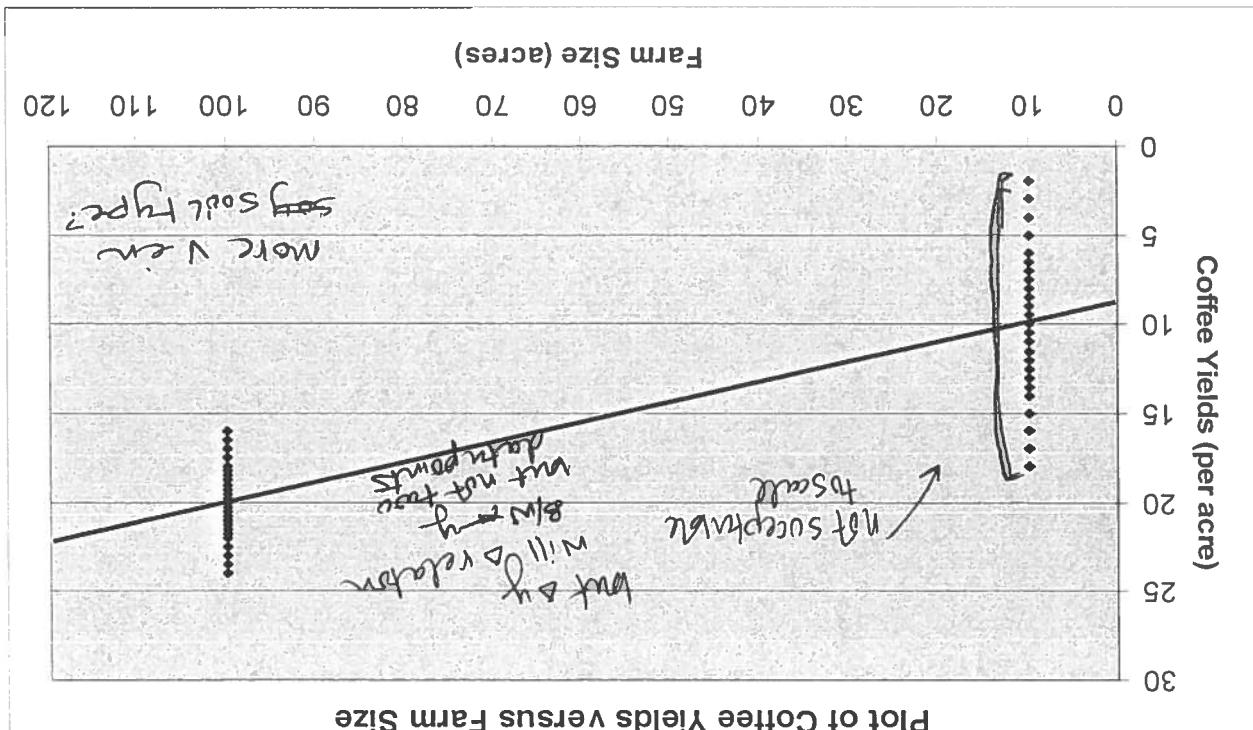
problem of cross sectional data



The former plot assumes that we have numerous yield observations, but from farms of size 10 and 100 only. In reality, we are more likely to have numerous observations from farms of a variety of sizes. Under such scenario, the plot would look like:

So, as farm size increases average coffee yields increase (the parameter estimate corresponds to the error term variance is lower at higher farm sizes).

farm size in the model would be positive), but the variability in coffee yields decreases (i.e. the error term variance is lower at higher farm sizes.



In a model with more than one explanatory variable, the OLS residuals can be plotted according to the order of any of the explanatory variables in the model, and the plots — of course — would all look different. It is possible that when viewed according to the order of another explanatory variable (such as farm size), error term variability (i.e. variance) is decreasing (or increasing) while it is relatively constant when examined in the order of another explanatory variable (such as the age of farms, the age of the plantation has no effect on yield variability, i.e. yield variability is about the same for plantations of all ages).

The pattern of a decreasing error term variance as farm size increases should still be obvious in the previous plot.

In the previous example, this would mean that although per-acre yield variability is lower at larger farms, the age of the plantation has no effect on yield variability, i.e. yield variability is about the same for plantations of all ages.

The main OLS assumption of no heteroskedasticity, however, requires that the error term variance is constant (i.e. the same) when viewed in relation to any and all of the explanatory variables in the model. This is also known as the homoskedasticity (i.e. constant variance) assumption.

↓↓↓↓↓
↓↓↓↓↓
↓↓↓↓↓
↓↓↓↓↓
↓↓↓↓↓

→ No pattern! Regardless

↓↓↓↓↓
↓↓↓↓↓
↓↓↓↓↓
↓↓↓↓↓
↓↓↓↓↓

~~most~~ ~~most~~

as long as $P_{t+1} < P_t$
 process comes back to 0
 closer to 1 = more dynamics
 in my! yet after which

- Discuss Excel example of AR(1) process
- error-term $\{u_t = \rho u_{t-1} + v_t, v_t \sim i.i.d N(0, \sigma^2)\}$
- This usually translates into an autocorrelated followed by periods of "depressed" prices
- there are several periods/years of "good" behavior over time, i.e. exhibit random cyclical behavior, commodity prices often series data, for example, commodity prices often
- It is most common in (but not unique to) time-independent assumption is violated
- Autocorrelation occurs when this error-term

Autocorrelation/Introduction

← can also happen for location
 autocorrelation
 near-time is one sometimes highly
 E_1, E_2, \dots, E_T nothing to do w/ each other
 organize in this time!
 usually a timeseries issue
 most → Autocor

- Previous (and all other) error-terms is independent of the size of the error-term
- of any particular observation is not related to (i.e. space), the sign and magnitude of the error-term particular order, usually time (but it could be that when the observations are organized in a independent distribution error term is independent and identically distributed)
- An independently distributed error term means the population error term is independent and of any particular order, usually time (but it could be that when the observations are organized in a space), the sign and magnitude of the error-term particular order, usually time (but it could be that when the observations are organized in a space), the sign and magnitude of the error-term
- Recall that when estimating the parameters of a regression model using OLS, it is assumed that the population error term is independent and identically distributed

Autocorrelation: Introduction

AAEC 6610
 Quantitative Techniques
 in Agricultural Economics
 Chapter 12
 Autocorrelation

WOMI
SWOT

Consequences

- The consequences of autocorrelation are exactly the same as those of heteroskedasticity:
- ▷ OLS parameter estimates remain unbiased but are no-longer the most efficient, i.e., GLS (not OLS)
- ▷ The OLS standard error estimates are biased, i.e., they are incorrect on average, which invalidates the results from the t-tests and confidence intervals
- ▷ F-test is invalid as well

~~Stoch Estimation~~

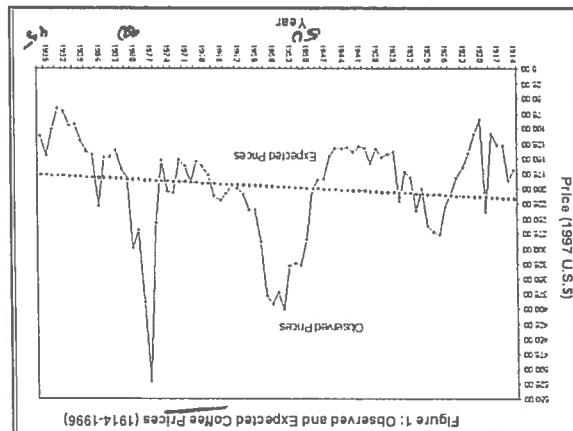
$$\begin{aligned} * & \rightarrow \text{corr}[U_t, U_{t-1}] = \rho \\ * & \rightarrow \text{corr}[Cov[U_t, U_{t-1}]] = \rho^2 / (1 - \rho^2) \\ * & \leftarrow \text{Var}[U_t] = \sigma_u^2 = \frac{\sigma^2}{1 - \rho^2} \\ & \text{Thus to be } > -1 + < 1 \end{aligned}$$

$$\begin{aligned} & E[U_t] = 0 \\ & \text{Main conclusion: Autocorrelation process} \\ & \quad \text{is AR(1) process, } Q \neq \Psi \text{ matrix, Gauss example} \\ & \quad \text{of a MA(1) process, see Excel \& Gauss example} \\ & \quad \text{Briefly discusses higher-order AR(p), MA(q) and} \\ & \quad \text{ARMA(p,q) processes, see Gauss example} \\ & \quad \text{Discusses random walk, stationary issue (AR vs} \\ & \quad \text{MA) and Dickey-Fuller and Phillips-Perron tests,} \\ & \quad \text{see Gauss example} \\ & \bullet \text{Structural vs time series (forecasting) models} \\ & \bullet \text{Discusses forecasting using Gauss example} \end{aligned}$$

- Derive the variance and autocorrelation function for an AR(1) process, see Example 2
- Derive variance and autocorrelation function for an AR(1) process, 2 & 4 matrix, Gauss example
- Briefly discusses higher-order AR(p), MA(q) and ARMA(p,q) processes, see Gauss example
- Discusses random walk, stationary issue (AR vs MA) and Dickey-Fuller, stationarity issue (AR vs MA) and Phillips-Perron tests, see Gauss example
- Structural vs time series (forecasting) models
- Discuss forecasting using Gauss example

Autocorrelation/Introduction

be observed in our time period
 - pos + lumped together (weeks can't
 ↓ vice versa
 if error is ↓ next forward
 PV for comodity



Visual Sig → using autocorrl does not really help

White test for autocorrl

assumes structural time series

comp DLS stat

$E(U_t - U_{t-1})^2 / E(U_t)^2$

No autocorrl (i.e. random)

→ null review

- More precisely, let H_0 : no autocorrelation vs H_a :
- Positive autocorrelation (a one-tailed alternative)
- Compare the calculated d^* with the "significance points" in the Durbin-Watson statistic Table at the number of independent variables in the model (k)
- If $d^* > d_u$ conclude H_0 ($d^* < d_l$)
- If $d^* < d_l$ conclude H_a ($d^* > d_u$)
- Otherwise the test is inconclusive
- Do example of DW test in Excel
- A two-tailed alternative is also possible (discuss)

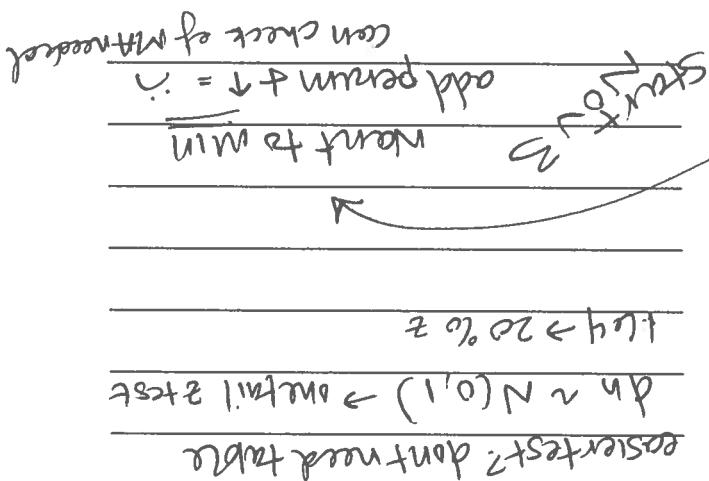
- ### Autocorrelation/Diagnostics
- The basic Durbin-Watson statistic tests for first order autocorrelation; i.e. correlation between the dependent variable values of previous period and the present period residual (i.e. $d^* = \frac{E(U_t - U_{t-1})^2 / E(U_t)^2}{k}$)
 - If there is no first order autocorrelation the value of d^* will be close to 2 (briefly discuss why) ↗
 - If would be smaller than 2 is there is positive first order autocorrelation and greater than 2 if there is negative first order autocorrelation (briefly discuss)
 - It is calculated as:

- ### Autocorrelation/Diagnostics
- for normal test
- The best known formal test for autocorrelation is the Durbin-Watson statistic
 - Residuals with clearly alternating signs indicate negative autocorrelation, which is very rare
 - Groups of several positive residuals followed by several negative residuals indicate possible positive autocorrelation, which is most common
 - Residuals versus time can be used as a preliminary diagnostic tool for autocorrelation
 - A visual inspection of a plot of the OLS residuals versus time can be used as a visual sign of autocorrelation

~~multiple time series~~

transform then use OLS

var of ϵ_t



- The solution for autocorrelation is analogous to that of heteroskedasticity: Transforms the original regression equation with the autoregressive error term into one with a non-correlated error term so as to permit the use of the OLS procedures; let:
- $y_t = b_1 + b_2 x_2 + \dots + b_k x_k + u_t$, $t = 1, \dots, T$
- $u_t = p u_{t-1} + v_t$, ($|p| < 1$) where both u_t and v_t have zero unconditional expected values and constant variances throughout time, v_t is autocorrelated but u_t is not
- The former defines a standard first-order autoregressive model; p is the correlation coefficient between errors in time-period t and errors in time period $t-1$

Correcting for Autocorrelation

- ~~positive autocorrelation~~
- ~~negative autocorrelation~~
- ~~autocorrelation~~
- ~~Durbin-Watson~~
- Durbin has shown that under H_0 : No autocorrelation using a standard normal table conducted using a standard normal table distributed with unit variance; thus the test can be conducted using a standard normal table for higher-order autocorrelations, AIC and SBC, go back to Gauss example
 - Model selection, Barlett and Box-Pierce tests if $T(V[B]) > 1$ (why?) $N \rightarrow$ take $\frac{1}{2}(V[B])$
 - A modified version of the test has to be used since significance is $\alpha/2$ *
 - Notice that if the alternative hypothesis is positive autocorrelation the level of significance is $\alpha/2$
 - Model selection, Barlett and Box-Pierce tests for higher-order autocorrelations, AIC and SBC, go back to Gauss example

~~lags really do explain~~

~~structural time series + add lags as~~

~~α ! tends not to reject null~~

- ~~Autocorrelation/Diagnostics~~
- The Durbin-Watson test only detects autocorrelation when no lagged dependent variable has been included in the model
 - If a lagged dependent variable is included, d^* would tend to be close to 2, even if there is autocorrelation present in the model
 - The Durbin h statistic has to be used in this case:
 - $h^* = \{1 - (d^*/2)\} V_t / \{1 - T(V[B])\}$
 - Where $V[B]$ is the square of the estimated standard error of the parameter of the lagged endogenous variable and T is the number of observations

Multicollinearity of previous form
subset DLS
 $y_t = f(y_{t-1} + \epsilon_t)$
autocorrelation
all xs are int diff isn't
 $t - p_{t-1}$

- If p were known, the following procedure could be used to correct for autocorrelation
- Notice that the former model implies:
- $\hat{y}_t = (\hat{B}_1 + \hat{B}_2 X_2 t_1 + \dots + \hat{B}_k X_k t_1) + v_t$ Use OLS
- Notice that the error term (v_t) on the last equation, which is a transformed model, satisfies the OLS assumptions
- Therefore, if p is known, applying OLS to estimate the transformed model given above yields BLUE parameter estimators and unbiased standard error estimators
- Relate to GLS estimator

- This involves using an estimate of p (\hat{p}) to implement the previously discussed procedure
- The simplest estimate is the sample correlation coefficient between the error terms:
- $\hat{p} = \frac{\sum_{i=1}^n v_i v_{i+1}}{\sum_{i=1}^n v_i^2}$ residuals
- This (approximate) GLS estimator is also known as the Cochrane-Orcutt procedure
- A second alternative is to regress v_i on v_{i+1} etc.
- Without an intercept and use the parameter estimate from that regression as an estimate for p (without an intercept) and use the parameter estimate from that regression as an estimate for p

- ### GLS Estimation
- Yet a third alternative is to use $\hat{p} = 1 - 0.5d$ where d is the calculated Durbin-Watson statistic
 - It is also possible to use the previous procedure on an iterative manner, as follows:
 - Estimate model by OLS $\hat{y} \rightarrow \hat{v} \rightarrow \hat{p} \rightarrow \hat{y}$
 - Compute residuals and obtain a first estimate for p
 - Oblin parameter estimates using GLS
 - Re-compute the residuals with GLS parameter estimates and raw data) and obtain a second estimate for p
 - Continue process until the estimate for p stops changing

$$\hat{p} = 1 - \frac{d}{2}$$

and residual GLS

$$\hat{y} \rightarrow \hat{v} \rightarrow \hat{p} \rightarrow \hat{y}$$

$$\hat{p} = 1 - \frac{d}{2}$$

$$\hat{p} = 1 - \frac{d}{2}$$

- An alternative to GLS/NLS is to use the Maximum Likelihood (ML) estimation method (to be discussed next)
- As NLS, this method usually involves numerical optimization search procedures
- If the error term is normally distributed, GLS/NLS and ML yield asymptotically equivalent results (to be shown later)

Sec Notes

Jointly find $\hat{\beta}_0 + \hat{\beta}_1 x_1$

\downarrow
P

$$\text{Min} (U, P_U) \rightarrow \text{Min} (U, \hat{U}) \leftarrow$$

$$\text{Min} ((y - XB)^T (y - XB))$$

OLS \rightarrow GLS \rightarrow Minimize ($(P_U)^T P_U$)

$$P_U = P(X^T X)^{-1} X^T Y$$

Exact generalization

- Here a large number of possible values for P are used to estimate the transformed model, and the estimated equation with the lowest RSS is selected
- Usually the P values to be tested are selected to be evenly spaced between -1 and 1
- Once the neighborhood for the RSS-minimizing P value is identified, a "refined" search can be conducted near that neighborhood
- Simple search is only feasible when Y involves one or two parameters, otherwise numerical optimization search procedures are needed

- ### GLS Estimation
- In fact, none of the previous alternatives is the preferred estimation method
 - More efficient parameter estimates are obtained by minimizing the sum of the squared residuals of the transformed model (do with matrices)
 - However, this is a non-linear minimization problem (NLS) that can only be solved by numerical search procedures
 - GLS \rightarrow Minimize ($(P_U)^T P_U$) helps obtain more efficient parameter estimates
 - However, it is not clear that iterating helps

Is A an
attribute
of the choice
model choosing one or
yes or no? And

Is yes or no
Did it have much time will pay; response
depends on the election
y to quality attribute

numerically discrete + continuous

11/10/2016

Nov 10

- In addition, binary-choice models are used to predict the probability that the dependent variable is in one category or the other given a set of values taken by the explanatory variables
 - An objective of these models is to quantify the impact of different factors (i.e. independent variables) on the probability that the dependent variable takes one value or the other
 - Primary choice models are used when the dependent variable can only take two mutually exclusive values (i.e. is a dummy 0/1 variable)
 - Binary choice models are used when the dependent variable can only take two mutually exclusive values (i.e. is a dummy 0/1 variable)
- Binary-Choice Models**

- Which type of vehicle (car, minivan, truck, SUV or domestic, Japanese, European) a consumer purchases
 - Whether or not a farmer decides to use systems (racitors, etc) a farmer decides to use which of four different available types of irrigation
 - Technology adoption studies: whether or not farmer adopts to adopt a recommended production practice(s)
 - Whether or not a (potential) visitor is willing to pay a particular fee to enter a park NPs
 - These models are most valuable in the analysis of survey and econometric experimental data:
 - These models are more opposed to quantitative choices (as opposed to qualitative values)
 - Qualitative choice models involve two or more dependent variables in which the dependent variables are involved in more qualitative surveys and economic experiments
- Introduction to Qualitative Choice Models**

Qualitative Choice Models

AAEC 6610
in Agricultural Economics
Qualitative Techniques

Nov 10

Linear Regression

DO NOT assume: Normality (y_i is not normal)

$Y = XB + U$, where X is an $n \times k$ matrix of independent variables, which can be of any kind, Y is an $n \times 1$ vector of qualitative dependent variables realizations (i.e. zero or one values), and U is assumed to be an $n \times 1$ vector of independent variables distributed random variables with zero mean.

The Linear Probability Model above is estimated by OLS, which yields inefficient parameter estimates and biased standard error estimates (briefly explain why, will specifically show later).

The Linear Probability Model is denoted by $E[Y_i|X_i] = \pi_i$

$P[Y_i=1|X_i] = \pi_i$

$P[Y_i=0|X_i] = 1 - \pi_i$

Discrete Random Var: Prob. prob
of add up.

Getting prob
of add up.

Used to $> 1 \leftarrow 1 < 0 \rightarrow 0$

Those prob out are less.
 $\therefore E[\pi_i] \neq \pi_i$

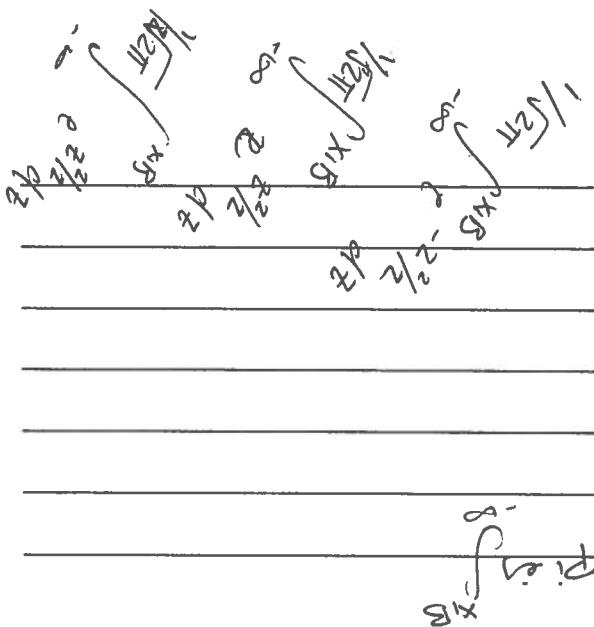
- By convention, since Y_i is a Bernoulli (0-1) random variable with $E[Y_i]=\pi_i$ (implied by $P[Y_i]=0$ being $(1-\pi_i)$, then $E[Y_i]=\pi_i$ being π_i and $E[Y_i]=X_iB+u_i$ that particular model)
- Now, since in the LPM we define $Y_i=X_iB+u_i$ where $E[u_i]=0$, it follows that $E[Y_i]=X_iB$ (implied by $E[X_iB]=X_iE[B]$ and $E[u_i]=0$ since $E[B]=\pi_i$)
- Thus, in the case of the LPM, π_i is defined to be equal to X_iB and $X_iB = \pi_i$ is the LPM prediction for π_i

Summary

- To be able to interpret $Y_i = \pi_i$ as the predicted probability that Y_i takes a value of 1 given X_i , the following convention has been adopted:
- Model is that $Y_i = \pi_i = X_iB$ could take values greater than one or less than zero. OLS ~~predicts~~ ^{estimates} $E[Y_i|X_i]$ is thus the LPM estimate for π_i ($\hat{\pi}_i$)
- Also since $E[u_i]=0$, $E[Y_i]=X_iB+\pi_i$ and $E[Y_i]=0$ by π_i ; then $E[Y_i]=1-(\pi_i)+0(1-\pi_i)=\pi_i$
- If we denote the probability that Y_i takes a value of 1 by π_i , then $E[Y_i]=1-(\pi_i)+0(1-\pi_i)=\pi_i$
- Therefore, another disadvantage of the LPM model is that $\hat{Y}_i = \hat{\pi}_i = X_iB$ could take values greater than one or less than zero. OLS ~~predicts~~ ^{estimates} $E[Y_i|X_i]$ is the following convention has been adopted:

The Linear Probability Model (LPM)

- The Linear Probability model is the most elementary binary-choice model:
- The linear probability model is the most elementary binary-choice model:
- $Y = XB + U$, where X is an $n \times k$ matrix of independent variables, which can be of any kind, Y is an $n \times 1$ vector of qualitative dependent variables realizations (i.e. zero or one values), and U is assumed to be an $n \times 1$ vector of independent variables distributed random variables with zero mean.
- The Linear Probability Model above is estimated by OLS, which yields inefficient parameter estimates and biased standard error estimates (briefly explain why, will specifically show later).



- The Probit and Logit Models
- This requires that $E[Y_i]$ is expressed as a non-linear function of X_{iB} , specifically a cumulative probability distribution function (CDF), because a CDF only takes values between 0 and 1 regardless of the value of X_{iB} .
 - That is, predicted probabilities are unbiased and lie in the $(0, 1)$ interval.
 - These models address the formerly discussed drawbacks of the linear probability model and insure complete the specification of the model above.
 - Numerous alternative CDF's could be used to normalize CDF , while the logit model is standard normal CDF, assuming a logistic CDF.
 - $E[Y_i] = P_i = \text{CDF}(X_{iB})$ (draw picture)

- The Probit and Logit Models
- More specifically, in the probit model:
 - Assuming a logistic CDF
 - Where Z is a normally distributed random variable with zero mean and unit variance (draw pictures)
 - By construction, P_i will lie on the $(0, 1)$ interval
 - As X_{iB} increases, the probability that Y_i takes a value of 1 (P_i) also increases

Now if $0.25 < 0.5$ in approach \Rightarrow approach

- The Linear Probability Model
- | | | | |
|-----------|-------------------|-----|----------------------|
| $P_i = 1$ | when $X_{iB} > 1$ | 0 | when $X_{iB} \leq 1$ |
|-----------|-------------------|-----|----------------------|
- It can be shown that the error term of the LPM is heteroskedastic: $\{g_i^2 = E[u_i^2] = P_i(1-P_i) = X_{iB}(1-X_{iB})\}$
 - If it has lower variances when P_i is closer to 0 or 1 and higher variances when P_i is closer to 0.5.
 - What are the consequences of heteroskedasticity?
 - A third weakness of the LPM is that the model's predictions (as above) are biased

Octavio Ramirez, 11/18/2014
OR1 Substitute for μ_1 and recognize that $E[Y|Z]=\mu_1$ also.

when $\alpha \neq \beta$ not ranked

Step 1: $\alpha > \beta$

$F(\text{tip coin}) = \frac{1}{2}$

Inhardt $P(A) = CDF$! discrete RV
don't need cat

count do this for cont. random var
but yes for discrete
we can

if y_i is independent (we assume)

* Note! logit function
just plotted CDF! could have
preceded continuity line etc!

CDF has different form (Now will go thru)
no threshold

$$\frac{1}{1 + e^{-\beta_0 - \beta_1 x}}$$

- Notice that in some of the examples of qualitative choices that more than two mutually exclusive choices represent more than two different variables can be discussed before, the dependent variables are individuals (farmers, consumers, businesses, etc), face a set of more than two alternatives.
- Here, individuals (farmers, consumers, businesses, etc) choose between four types of irrigation systems ranked "more" than the other.
- Those simpler multiple choice models apply when clearly imply "more" than the other.

Multiple Choice Models

$P_i = \frac{\text{prob model } i}{\text{sum prob all models}} \rightarrow \text{marginal probability} \rightarrow \text{marginal}$

- And the log-likelihood function to be maximized is:
- $L(Y|ln(P_i)+(1-Y)ln(1-P_i))$ where P_i has been defined before for both the probit and logit models (example).
- The likelihood function is:
- $P_i^{Y_i} (1-P_i)^{1-Y_i} \rightarrow \text{joint PDF}$
- Therefore, the PDF for any given observation is:
- Notice that in both, the probit and logit models, the probability that Y_i takes on a value of one is P_i and the probability that Y_i takes on a value of zero is $(1-P_i)$.
- Notice that in both, the probit and logit models, the PDE for any given observation is:

Probit and Logit Models

Maximum Likelihood Estimation of

- The best way to estimate either of these models is by maximum likelihood procedures

Also, the computations to estimate the logistic model are simpler.

The probit and logit models are fairly similar, with the difference being that the logistic distribution has slightly fatter tails.

In the case of the logit model:

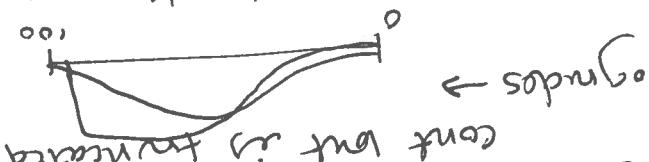
$$P_i = \text{CDF}(X_i B) = 1 / (1 + e^{-X_i B})$$

The Probit and Logit Models

with the difference being that the logistic

with the difference being that the logistic

• Higher the truncation the less OLS works.



↳ Tobit model: used when dep var is truncated

out kommen model

11/12

↳ \rightarrow count + var \rightarrow Poisson

↳ use known means + result true
counts are correlated

↳ num counts are

↳ also Neg binomial regression

ordinal + discrete numerical

↳ final option

• Poisson regressions can be used for count variables

multiple

↳ 70% 60% not
continuous only buy

also es max likely hood
expansion of probit / logit

- Multiple Choice Models
- Multiple choice (multinomial) versions of the probit and logit models are available in the literature and can also be estimated using maximum likelihood procedures
- More sophisticated multiple choice models such as the ordered probit and logit and the poisson count regressions allow for dependent variable alternatives that follow a clear ranking.
- An example of the latter would be a farmer's choice of crop insurance levels, since these can be unequivocally ranked from low to high.

Joint PDF

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

Sum of observations $y = y_1 + y_2 + \dots + y_n$

Joint PDF for y_1, y_2, \dots, y_n

Good way to solve for μ is max

Sample of n observations y_1, y_2, \dots, y_n from unknown

Depend on parameter μ & σ^2

$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$

Probability Density Function

$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$

Joint PDF is product of individual PDFs

Maximum Likelihood Estimation

- If the population error term u_i is i.i.d. normal, the probability that the i th observation $(Y_i, X_{i1}, \dots, X_{ik})$ occurs given a set of values of parameters (B_1, \dots, B_k) , and the error term variance σ^2 is given by the normal probability density function:
- $f(Y_i | X_{i1}, \dots, X_{ik}, B_1, \dots, B_k) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i - \mu)^2}{2\sigma^2}}$, where $\mu = (X_{i1}B_1 + \dots + X_{ik}B_k)$, and the error term variance σ^2 occurs for any given set of values of the model's parameters (B_1, \dots, B_k) .
- If each observation Y_i is independent and identically distributed by $f(Y_i | X_{i1}, \dots, X_{ik}, B_1, \dots, B_k)$ (also discuss the case of a non-linear model), the model's parameters equals the product of the single-observation probabilities, i.e., the product of PDF's (one observation probability that the whole sample for each observation) as defined above

Maximum Likelihood Estimation

- Maximum likelihood estimators for μ and σ^2 can be estimated by ML methods as well
- Many other types of "specialty" models can be accommodated by ML methods
- It can also be used to estimate models with non-i.i.d. errors and (as GLS/NLS) is able to accommodate any conceivable form of heteroskedasticity and/or autocorrelation
- It can be used to estimate regression models that are not linear in the parameters (do example later)
- Maximum likelihood is the most general method for estimating econometric models
- Discusses pdf and cdf of $u_i \sim N(0, \sigma^2)$, $y_i \sim N(\mu, \sigma^2)$, joint pdf for Y_i and ML estimators for μ and σ^2

Maximum Likelihood Estimation

Quantitative Techniques
in Agricultural Economics

AAEC 6610

Joint PDF $f(y)$

Normal distribution (μ, σ^2)

Probability Density Function

$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$

Joint PDF $f(y)$

will be more efficient.

of error Normal: $m + \sigma z$ same

not normal ML won't like ML

log likelihood function

- When $u_i = Y_i - B_1 - B_2 x_{i2} - \dots - B_k x_{ik}$ (i.e. a linear model), the system of k equations obtained by taking the first k derivatives of LLF with respect to B_1, \dots, B_k is identical to the system obtained when B_1, \dots, B_k are the parameters of the OLS and ML estimators taking the derivatives of the RSS with respect to those parameters; thus the OLS and ML estimators are the same in this case
- Notice that, as in OLS, in maximum likelihood estimation the Y and X observation values (i.e. the sample data) are considered constant and the parameters are the variables which needs to be maximized
- The likelihood function needs to be maximized with respect to which sample data are considered constant and the estimation that, as in OLS, the Y and X observation values (i.e. the sample data) are considered constant and the parameters are the variables which needs to be maximized

Maximum Likelihood Estimation

$$-(\ln(L)) = \ln(\ln(L)) + (\ln(L)^2)$$

- The usual procedure in maximum likelihood estimation is to maximize the natural logarithm (\ln) of the likelihood function instead of the log.
- This is often easier and yields the same results since the \ln is a monotonically increasing transformation of the likelihood function instead of the original function
- Taking the \ln of the LLF (for the case of the i.i.d. normal error term model) and ignoring the constants we obtain:
- $L(L(Y_i | B_1, \dots, B_k, \sigma^2)) = -(\ln(2\pi))^n / \sigma^2 (1/2)^n (\ln(\sigma^2))$
- $L(L(Y_i | B_1, \dots, B_k, \sigma^2)) = -(\ln(2\pi))^n / \sigma^2 (1/2)^n (\ln(\sigma^2))$, which is the function to be maximized w.r.t. B_1, \dots, B_k and σ^2 .

Maximum Likelihood Estimation

$\ln(L) \rightarrow \min$ RSS sum values

$-\frac{1}{2}\ln(\sigma^2) \text{ won't do anything} \rightarrow \text{const}$

thus \rightarrow get same result

$\min / \max val \text{ of function via monotonic}$

$-\frac{1}{2}\frac{\partial \ln(\sigma^2)}{\partial \sigma^2} + -\frac{1}{2}\frac{\partial \ln(\sigma^2)}{\partial \sigma^2} = \frac{1}{2}\frac{\partial \ln(\sigma^2)}{\partial \sigma^2}$

$$\begin{aligned} & \frac{\partial \ln(\sigma^2)}{\partial \sigma^2} = \frac{1}{2} - \frac{1}{2} \frac{\partial \ln(\sigma^2)}{\partial \sigma^2} \\ & \ln(\sigma^2)_{1/2} + \ln(\sigma^2)_{1/2} + \ln e^{-1/2} \frac{\partial \ln(\sigma^2)}{\partial \sigma^2} \leftarrow \text{RSS} \\ & \ln(\sigma^2)_{1/2} + \ln e^{-1/2} \frac{\partial \ln(\sigma^2)}{\partial \sigma^2} \\ & f(\sigma^2) = (2\pi\sigma^2)^{-n/2} \exp(-\frac{1}{2}\frac{\partial \ln(\sigma^2)}{\partial \sigma^2}) \leftarrow Y_1 - B_1 - B_2 x_{i2} - \dots - B_k x_{ik} \\ & \text{take nat'l. } \frac{\partial \ln f}{\partial \sigma^2} = \max \\ & U_i = Y_i - M_i \end{aligned}$$

- In the case of the (linear or non-linear) i.i.d. normal error term model, the likelihood function is:
- $L(F(u_i | B_1, \dots, B_k, \sigma^2)) = (2\pi\sigma^2)^{-n/2} \exp(-0.5 \frac{\partial \ln(\sigma^2)}{\partial \sigma^2})$.
- The maximum likelihood estimates for B_1, \dots, B_k and σ^2 are the individual set of values of the parameters ($B_1, \dots, B_k, \sigma^2$) that maximizes LLF given our n sample observations on Y and the X 's

Maximum Likelihood Estimation

How to compute SE of est
estimator
so we can sample! ML estimator be good b/c
everything hinges on it! it's hard to when
it's not consistent
of f(.) correct all else are consistent
so why sample? ML estimator be good b/c
asymptotically efficient \rightarrow Min var if
a is large

$E[\hat{G}^2]$ gets to be so close to
Prob \rightarrow not to \hat{G}^2
 $Bias \leftarrow E[\hat{G}^2] - G^2$

- Maximum Likelihood Estimation
- The Gauss ML procedure is one of them
- Developed to numerically find the set of parameter values that maximize a given likelihood function
- Therefore, specialized programs have been developed, specifically programs have been to estimate the parameters of a given likelihood function
- A heteroskedastic model
- An i.i.d. linear model
- Do example (program 3) of using the ML procedure to estimate normal log-likelihood to the case of non i.i.d. errors and explain how to use it!

Nov 10

Sum of all of linear

- The formula states that LLF is only correct for an i.i.d. normal-error regression model
- In most cases (other than the i.i.d. standard-normal error regression model), the system of equations obtained when taking the derivatives of the likelihood function with respect to the parameters to be estimated can not be easily solved
- The maximum likelihood estimates for the variance of the error term of the linear regression model, the standard-normal
- Thus its diagonal elements provide statistically sound estimates for the variance of the maximum likelihood estimators of B_1, \dots, B_k and G^2
- Consistent estimate for the covariance matrix and thus the inverse of this information matrix is a normal-error regression model
- The formula states that LLF is only correct for an i.i.d. normal-error regression model
- In most cases (other than the i.i.d. standard-normal error regression model), the system of equations obtained when taking the derivatives of the likelihood function with respect to the parameters to be estimated can not be easily solved

- Solving for the $k+1$ derivative of L.F (with respect to G^2) yields a slightly different estimator for G^2 :

$$\hat{G}^2 = S^2 = \frac{1}{n} \sum_i \hat{Y}_i^2 / n$$
- Under very general conditions the ML estimators for a given model parameters are consistent and most efficient (asymptotically) as long as the model specification is correct
- Minus the expected value of the matrix of second partial derivatives of LLF w.r.t. the $k+1$ parameters evaluated at their LLF-maximizing values is known as the information matrix

Likelihood Ratio Tests

- Notice the logic of the test:
- MVURLF~MVRLLF implies that the likelihood occurrence of the sample values under $(Y_i, X_{i1}, \dots, X_{ik}, i=1, \dots, n)$ is nearly the same under the unrestricted and restricted models
- Therefore, the restricted model is nearly as adequate to explain the sample values as the unrestricted model and H_0 should not be rejected
- Do examples of likelihood ratio tests

Likelihood Ratio Tests

- Under the null hypothesis that the restrictions are valid, LRTS is distributed as a χ^2 random variable with d degrees of freedom, where d is the number of restrictions being tested
- If the calculated value of the LRTS exceeds the value of a χ^2 random variable with d degrees of freedom at the desired level of statistical certainty (α), H_0 is rejected in favor of the H_a : at least one of the restrictions being tested is not statistically justifiable

[can be used for joint twings like f-test]
 unlikely as diff pos.
 weights are not an issue.
 weight of unres + & does not work
 max rule of unres + log like max
 H₀: unrestricted
 H_a: restricted
 may not be as reliable
 useful of se (w/o Hessians)
 type test on param

- If a model parameters are being eliminated by maximum likelihood methods, the likelihood ratio tests are a straightforward way of conducting ratio tests about those parameters
- As in the case of the F tests, the likelihood ratio tests are based on estimating an unrestricted, and a restricted model that incorporates the restrictions about the parameter values that one wants to test
- A likelihood ratio test statistic is then calculated as:
- LRTS = $2(\text{MVURLF}-\text{MVRLLF})$; where MVURLF and MVRLLF are the maximum values reached by the log-likelihood functions in the case of the unrestricted and the restricted model, respectively
- Likelihood ratio tests statistic is then calculated as:

Likelihood Ratio Tests

logarithmic function (just know about)

F test (linear)

just know about

- When linear models with normally distributed error terms are involved, the F test (also known as the Wald test) is easily applied.
- Likelihood ratio tests, however, might be required when working with more general types of models.
- The two tests are asymptotically equivalent, i.e., they will produce the same results when large samples are involved.
- When working with OLS models and small samples, however, the F test will always be more prompt to reject H_0 .
- A more complicated but also asymptotically equivalent test is the Large range multiplier test.

1911
July 1
Afternoon

~~highly~~ ~~mark~~

and IV doesn't affect; price w/o end day
but aggregate $\leftarrow P + Q$ our jointly determined

another equation for nutrient food

$$\text{ex: } \# \text{ of animals in park} + \text{out}\text{of area available} = \text{total population}$$

\rightarrow no feed back effect \leftarrow not jointly
if $y \in$ trust doesn't affect

$z \leq 510$

- When working with simultaneous equations systems, one can think of the model as consisting of a set of equations, with each equation stating a known interrelation between the endogenous variables, however, each equation can be estimated separately
 - A simple example is a two-equation model of market demand and supply, where price and quantity are both endogenous variables:
 - $Q_d = Q_d^e = Q_d^r = Q_d^s$ (Q_d^e equilibrium condition)
 - $Q_d^e = Q_d^r = P_2 + 2P_1 + P_3 Y_1 + u_1$ (demand, P_1 could be in RHS)
 - $Q_d^s = Q_s^r = C_1 + \alpha P_1 + \alpha P_2 - 1 + e_1$ (supply, P_1 could be in LHS)

can focus on our but results and add est tequeur.

- A single equation model assumes that a set of independent variables affect the value taken by “feedback” relationship between the dependent variable and any of the independent variables (i.e. they are not jointly determined)
- Multiple simultaneous equations models can account for the interrelationships within a set of jointily determined dependent (i.e. endogenous) variables as well as include exogeneity (i.e. exogenally determined independent variables)

Introduction

Multivariate Equation Models An Introduction to

in Agricultural Economics

AAEC 6610

2/2

→ overcount measurement

sideway cover w/ error term

applicable only where $E+$

* both within var + structure less ts.

put est. quantities

more difficult

- Instrumental variables and two-stage least
- Three-stage least squares (instrumental variables, two-stage least squares) that yield consistent parameter estimates
- Although, under these circumstances, OLS is not an acceptable estimation method, there are other procedures (instrumental variables, three-stage least squares, and maximum likelihood) that yield consistent parameter estimates
- Simultaneous Equations Systems

as an answer

is → does not matter if for

determine which eq. +

pre determine var →

endogeneity problem

- Simultaneous Equations Systems
- y_i is called endogenous variable: its value is supposed to be determined outside of the system, and it is not affected by Q_i and P_i .
- The "endogeneity" (i.e. simultaneity) of P_i and Q_i because of this endogeneity, applying OLS to changes in Q_i or P_i would affect both P_i and Q_i can be better perceived graphically, noting that can be better perceived graphically, noting that because either the demand or the supply estimate either the demand or the supply equation (or the corresponding price equation)
- Because of this endogeneity, applying OLS to would produce biased and inconsistent parameter estimates
- Simultaneous Equations Systems

exog.

det. exogenous \neq $Q_i + P_i$

lagged endog. var → pre

$y_t = \text{int} + \sum \text{exogenous}$

- Simultaneous Equations Systems
- Those equations determine the market price and the quantity supplied and demanded when the market is in equilibrium
- Q_i and P_i are called endogenous variables because their values are determined from within the model causes the movement of the endogenous variables within the system (graphically)
- P_i and Y_i are pre-determined variables that help Note that the values of P_i are also determined within the system so it is a large endogenous variable, but it is pre-determined at time period t within the system so it is a large endogenous variable, but it is pre-determined at time period t

use econometrics to rep the structure

- A model such as the supply-demand system is discussed before it is derived from the underlying economic theory.
- A structural model contains endogenous variables on the left-hand side of its equations and exogenous/pre-determined variables in the right-hand side of its equations.
- The equations in a simultaneous structural model can be solved for each of the endogenous variables as a function of the predetermined variables only.

Reduced Form Model

- This is called a reduced form model (or solution), and consists of one equation for each of the endogenous variables which is only a function of the structural model's parameters and predetermined variables.
- For example, the reduced form model for the structural demand-supply system discussed before is represented by:
- $P_i = \pi_1 + \pi_2 Y_i + \pi_3 P_{i-1} + \pi_4$ (equilibrium price)
- $Q_i = \pi_1 + \pi_2 Y_i + \pi_3 P_{i-1} + \pi_4$ (equilibrium quantity)
- Note that the reduced form model parameters (π_{ij}) are functions of the structural model parameters (π_{ij}) and vice versa.

The Identification Problem

- Suppose that the values of the parameters of the reduced form solution of a structural model are known; is this sufficient to determine unique values for the parameters in the corresponding set of structural equations?
- Unfortunately, this is not always the case.
- A given structural equation is identified if it is possible to calculate its parameters from knowledge of the reduced form model.

parameters

knowledge of the reduced form model

possible to calculate its parameters from given structural equation is identified if it is

knowledgeable to calculate its parameters from given structural equation is identified if it is

knowledgeable to calculate its parameters from given structural equation is identified if it is

knowledgeable to calculate its parameters from given structural equation is identified if it is

knowledgeable to calculate its parameters from given structural equation is identified if it is

knowledgeable to calculate its parameters from given structural equation is identified if it is

knowledgeable to calculate its parameters from given structural equation is identified if it is

knowledgeable to calculate its parameters from given structural equation is identified if it is

unbiased / consistent measure

current records from left

are notes

SA

over 1d

over 1d 11

exact ID ✓

muskrat

muskrat

in some cases → mouth turn

The Identification Problem

- It is exactly identifiable if only a single set of parameters (as in the case of our example) can be inferred from the reduced equation
 - It is over-identified if more than one set of parameters for that structural equation can be inferred from the reduced equation
 - Otherwise it is not identified

- The Identical Problem As a simple example of the identical demand-supply model, which assumes that the market is in equilibrium every time-period so that $Q_d = Q_s$:
- $$Q_d = \alpha_1 + \alpha_2 P_t + \epsilon_t \quad (\text{Supply})$$
- $$Q_s = \beta_1 + \beta_2 P_t + \epsilon_t \quad (\text{Demand})$$
- Neither of the equations in the previous structural model is identified.
- There are an infinite number of structural models (i.e. sets of supply/demand equations) which are consistent with the same reduced form model,
- which represents the equilibrium values of P and Q .

The Identification Problem

- As a matter of fact, according to the previously assumed structural model, expected equilibrium price and quantity values remain the same through time
 - However, the actual observed values of Q_1 and P_1 change through time because of the error terms in the supply and demand equations making them shift randomly from year to year 
 - The presence of the error terms is actually the only reason why different values of P_t and Q_t are observed through time and the structural equations can thus be (inconsistently) estimated

price shifts

need to ID equation

Why we can't get supply up
know what's known

more P to SP → get over demand
add another SP to supply →
demand becomes over demand

Pt-1 Δs → ID demand

yte D D → IDS

original model

add shift to exogenous

- The Identification Problem
- The reduced-form equations (ILS method) if one knows the values of the parameters of the demand) equations can be calculated (i.e. recovered).
 - The parameters of both structural (supply and reduced-form equations (ILS method)
 - $P_i = \alpha_1 + \beta_2 Y_i + \beta_3 P_{i-1} + \nu_i$
 - $Q_i = \alpha_1 + \beta_2 Y_i + \beta_3 P_i + \nu_i$
 - The reduced form model in this case indicates the equilibrium values of Q and P as a function of the two predetermined variables Y_i and P_{i-1} .
 - The reduced form model in the supply curve is identified by the shifts in the supply curve in the demand curve and the demand equation will be identified.
 - The supply equation will be identified by the shifts in the supply equation will be identified by the shifts in the supply curve and the demand equation will be identified.

Step 11/13

- shift over time and can be identified
- If P_i and Y_i vary over time (and are not perfectly correlated) both the supply and demand curves will level at time period t
 - $Q_i = \beta_1 + \beta_2 P_i + \beta_3 Y_i + \nu_i$ (demand); where Y_i is income price in the previous time period
 - $Q_i = \alpha_1 + \alpha_2 P_i + \alpha_3 P_{i-1} + \varepsilon_i$ (supply); where P_{i-1} is the price in the previous time period
 - Now consider the original supply-demand model: equations
 - The supply equation is identified in this case: its parameters can be deduced from the reduced form equations

- The Identification Problem
- It is obvious that more information is needed in this case to identify the supply and demand equations.
 - Consider, then, the following (expanded) model:
 - $Q_i = \alpha_1 + \alpha_2 P_i + \alpha_3 P_{i-1} + \varepsilon_i$ (supply)
 - $Q_i = \beta_1 + \beta_2 P_i + \beta_3 Y_i + \nu_i$ (demand); where Y_i is income level at time period t
 - Changes in Y_i would make the demand curve to path underrating the supply curve, but it is still not possible to determine where the demand curve is
 - The resulting equilibrium values will trace out the systematical shift from year to year, allowing for the identification of the supply curve in this case
 - Consider, then, the following (expanded) model:
 - $Q_i = \beta_1 + \beta_2 P_i + \beta_3 Y_i + \nu_i$ (demand); where Y_i is income level at time period t
 - Changes in Y_i would make the demand curve to path underrating the supply curve, but it is still not possible to determine where the demand curve is

The Identification Problem

- More stringency (fail proof) test is the rank sufficiency condition: it is possible for it to be satisfied and (rarely) the equation to be unidentified and sufficient; if it is necessary but not sufficient for identification to exist *
- The rank condition is both necessary and sufficient for identification of the rank condition for a more advanced course

in large systems
for identification

The Identification Problem

- More formally, the order condition states that for an equation to be identified, the number of predetermined variables excluded from the equation must be equal to or greater than the number of endogenous variables included in the right-hand-side of the structural equation
- If either, the equation is exactly not a problem as we will learn next is really not a problem
- If neither, the equation is over-identified, which if equal, the equation is exactly identified

of predictor var in
system =

final demand ex 1D ex 2 ! + 2 for each exogenous predictor
price, income + wealth
structural return
JLR
Demand ex 1D ex 2 but supply in
over demand \rightarrow get parameter Demand
test is scale for parameter
Supply exogenous

The Identification Problem

- Finally, consider a situation in which the demand equation includes income and wealth as exogenous variables: its shifts over time would result from changes in two variables
- In this case, there are two pre-determined variables in the supply-demand system which are not in the supply equation: this makes it to be over-identified
- Over-identification implies that the reduced form supply equation: this makes it to be over-identified least two sets of values for the parameters of the parameters imply (i.e. can be used to calculate) at least two sets of values for the parameters of the supply equation
- Supply equation: this makes it to be over-identified in the supply-demand system which are not in the supply equation: this makes it to be over-identified

Plus simultaneous \rightarrow see notes section

$$Prm(B_{IV}) = B + Prb \ln(z_1 x_1 z_2)$$

$$B_{IV} = B + (z_1 x_1 z_2)$$

$$B_{IV} = (z_1 x_1 z_2) + (z_1 x_1 z_2)$$

$$B_{IV} = (z_1 x_1 z_2) = (z_1 x_1 z_2)(x_B + u)$$

② Next consider error term

① Consider using var(y) better
instead.

④ simultaneous \rightarrow

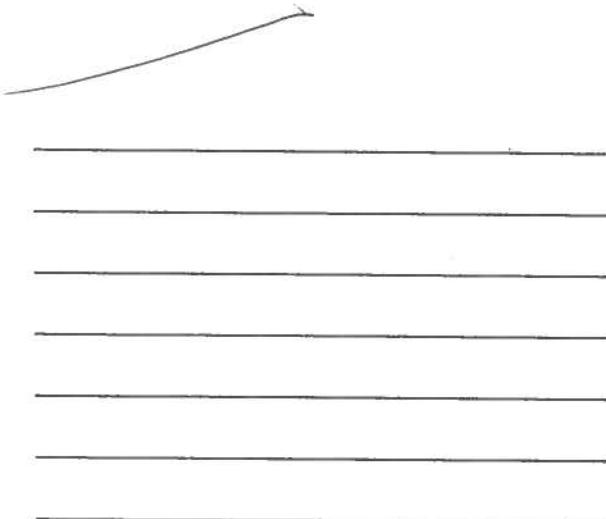
- LLS yielded exactly the same result
- However, if an equation is just identified, IV and II obtained with the IV estimator thus different sets of parameter estimates for Z and identified, there are multiple choices for Z and thus different sets of parameter estimates can be obtained with the IV estimator.
- As in the case of LLS, when an equation is over-identified with the endogenous variables in it correlated with the endogenous variables in it, they are likely to be correlated with the system (but not in the equation) can serve as excellent instruments for determining variables in the system (but not in the equation).
- Since they are in the system, they are likely to be correlated with the endogenous variables in it.
- Since they are considered pre-determined, they are not correlated with the endogenous variables in it.
- In the particular case of simultaneous equations since they are in the system, they are likely to be correlated with the endogenous variables in it because:

Consistent Parameter Estimation

- Under the above conditions, IV estimation produces consistent parameter estimates even if some of the variables in X are correlated with the error term (as long as Z is not).
- Note that Z can be based on X, with only the endogenous column(s) being replaced by instruments correlated with the error term.
- Note that Z can be based on X, with only the endogenous variables which are correlated with the instruments in the equation and are not correlated with the error term.
- $B_{IV} = (Z'X)^{-1}Z'Y$ and Z is an n by k matrix of instruments in simultaneous equations is Instrumental Variable (IV) estimation, where:
- A simpler technique to obtain consistent parameter estimates in simultaneous equations is Instrumental Variables (IV) estimation, which is the one possible set of parameters that the equation is over-identified, it will yield more efficient equations of the structural system.
- A disadvantage of the LLS approach is that, when estimating the reduced form system of equations by OLS and using the resulting parameter estimates to calculate parameter values for the structural equations of the system.
- The former implies a loss of information and calculating efficiency.

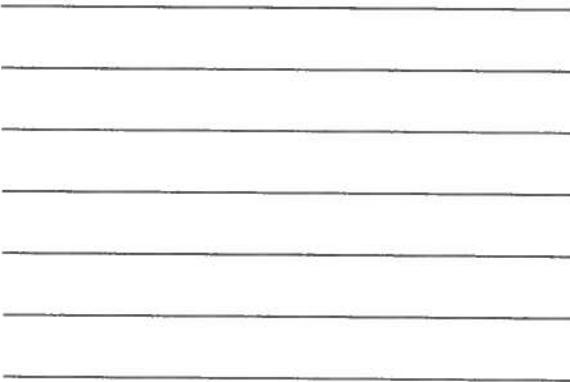
Consistent Parameter Estimation

- As previously mentioned, indirect least squares (ILS) is an alternative approach which is based on estimating the reduced form system of equations by OLS and using the resulting parameter estimates to calculate parameter values for the structural equations of the system.
- A disadvantage of the ILS approach is that, when the equation is over-identified, it will yield more efficient equations of the structural system.
- The former implies a loss of information and calculating efficiency.



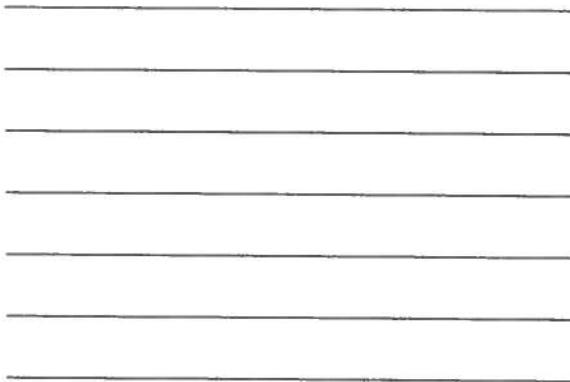
- In this case, the endogenous variables are correlated with the will be perfectly (linearly) correlated, but 3SLS equations could be identified and others may not
- Keep in mind that in a given system some equation (example)
- 2SLS is the most widely used method, but 3SLS and maximum likelihood are available as well.
- 2SLS is impossible
- Alternatively, note that (as ILS and IV) when the structural equation of interest is not identified

Two-Stage Least Squares



- Theoretically, under a large sample size, those estimated by OLS
- In the second stage, the observed values of the endogenous variables are replaced by their corresponding predicted values in the X matrix, and the structural equation parameters are then estimated by OLS, if over-identified
- Therefore, the 2SLS technique yields consistent estimates for the structural equation parameters
- If the structural equation is exactly identified, 2SLS is identical to IV or ILS, if over-identified
- 2SLS is a more efficient estimation method

Two-Stage Least Squares



- In the first stage, the reduced form equation corresponds to each of the endogenous variables in the system is estimated by OLS, and then used to predict the values of those endogenous variables for each of the observations in the sample
- The two-stage least squares (2SLS) technique can be used to estimate the parameters of over-identified equations in structural systems avoiding the loss of information and estimating efficiently the lost of information and estimating each structural parameter

Two-Stage Least Squares

- 3SLS includes a third stage which takes advantage of the correlations in the error terms across the structural equations to increase estimation efficiency
- In most cases, however, the estimation efficiency gains are marginal relative to 2SLS
- Maximum likelihood is a likelihood-based application of 3SLS where all parameters are jointly estimated in a single "stage"
- Asymptotically, 3SLS and ML are the same

3SLS and ML

Dec 1

```

format /rd 8,5; 2*NLF*n;
{ b,NLF,gx,cov,retcode } = CMLPrt(CML(data,vars,MLE,start));
retP(NLF); endp;

NLF=-(1/2)*sumc(ln(SIG.^2)+((Y-XB)./SIG).^2);
SIG=b[4]+b[5]*X;
XB=b[1]+b[2]*X.*b[3];
BXB=b[1]+b[2]*X+b[3]*(X.^2);@
X=(data[,1]-1912);@*
Y=(data[,2]);@*
n=rows(data);@*
proc MLE(b,data);@*
local NLF;
output=l; cm1_MaxIter=1000; cm1_CovPar = 1; cm1_DirToI = 0.00001;
clearg Y,X,XB,SIG,n;
etcode=0;
et vars[2,1]=1 2;
start=30110110;
start=26.510.510.0111210;
start=30110110;

```

www.WTP

Pos sign \rightarrow positive

$$P_i = \text{prob}$$

Good return?
Bad

such always + same value
want to know things
that $\rightarrow 1000x$
square \rightarrow don't want a x

$\hookrightarrow P_i$ is non linear
else not slopes or derivatives

Upward

Special

```

new! Cls: wait;
load data[801,12]=c:\files-c\prog5\data.txt;
library cml; #include cml.set;
n=rows(data);
set vars[12,1]= 1 2 3 4 5 6 7 8 9 10 11 12;
set start[12,1]=-0.2 -0.8 0 0 0 0 0 0 0 0 0 0;
clear y,x,z,p;
proc dcm(b,data);
cml_algorithm=3;
logcat lf;
lf=(y^ln(p)) + ((1-y)^ln(1-p));
z=x*b; p=1/(1+exp(-z)); logit;
y=data[,1]; x=ones(n,1)~(data[,2]/100)~data[,3:12];
do until i>9; bid[i]=5*i; endo;
bid=zeros(10,1); i=0;
do until i>9; i=i+1; bid[i]=5*i; endo;
XB=b[1]+b[2]*bid+b[3]*mean(data[,3])+b[4]*mean(data[,4])+b[5]*mean(
+b[6]*mean(data[,5])+b[7]*mean(data[,6])+b[8]*mean(data[,7])+b[9]*mean(
data[,9])+b[10]*mean(data[,10])+b[11]*mean(data[,11]);
P=1/(1+exp(-XB)); bid~P;

```