tsit in front

AAEC 6610 Quantitative Techniques in Agricultural Economics

Introductory Remarks What is Econometrics?

What is a Regression Model? The Population vs. the Estimated Model

Details +

" СУyish1

ју ууууум

Some Unsolicited Advice

• Graduate-level classes are much more demanding than

undergraduate-level courses: You are responsible for learning all of the material discussed in class

• Attend class and make sure to focus, follow and

understand all of the material being presented

• Might need to read the textbook beforehand

• The slides and book are not enough: Taking detailed

notes and understanding as well as memorizing the material is key to being successful in this class Kecp up with the material: 5-6 hours of studying per week and an additional 10-12 hours before an exam is needed to realize your full potential in this class TA hours - let me know if you need help!

How its affected by others

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

other variables conformpripuud af gors,

• It is a primary statistical tool for data analyses in many scientific and management fields, suitable

for analyzing a wide variety of problems and issues

• In economics, regression analysis the basis of a

discipline called Econometrics

• Econometrics also involves many other more advanced statistical methods that are suitable for analyzing economic data/relations

What is a Regression Model?

• Regression analysis starts by conceptualizing

a behavioral relation based on economic

theories/knowledge or plain reasoning

• This behavioral relation includes one -dependent variable (Y) and one or more

independent variables (X’s), which are believed to influence Y (economic examples: demand and production functions)

• In regression analysis, this relation is stated as

, a mathematical equation or model:

Y = B! + B2X2 + B3X3 + B4X4 + u

X Var that y deps on tust & preference time trend allows foo the accounts for (factor)

Prelated good; p of good, yd/GDP

prodo inputs

outp is fluortal labol

*vus)*

Preef ; '/d rete.

- Bz

१mstone

i nsbintaneous Reste efs

indip What is a Regression Model?

• In that model (continue with demand example):

→ Y and X2, X3 and Xt are the dependent and the

independent or explanatory variables, respectively >B1, B2, B3 and B4 are true parameters, i.e., constant cocfficients that quantify the relations between Y and X2, X3 and X4 (derive generic interpretation and give

an example) >B1, B2, B3 and B4 are estimated using data on Y and X2, X3 and Xt and the regression analysis techniquc u is a random crror or disturbance term, which recognizes that the relation between Y and X2, X3 and Xa is not exact ervar term

cunt explain wl exact relationship.

Unit I whinya to Beefconsompilblf846 xq = sep (116) B2 = 15 (Neg b/e price has negeffect on consumpt)

149 6-15 16/person/sayr

Prel goods (subst) is pos

\* тер*о*р

model says By = intercept value taken byy when all as une ø sometines duesn't make

sense thats okay

do issues

cant

nowevest*,* nose w/ lowy

What is a Regression Model?

• The error term takes into account other

factors that affect the dependent variable Y, such as (continue with demand example):

Relatively unimportant explanatory variables Random error in the measurement or Y (but not in llie X's) Not Alodays pos-Not DW]

>Pure chance an outcome or the proportion of at

outcome that can not be explained by any known

de factors – i.e. variables] Ramenez din.

& do best to indul all key exp. var.

LP ketchup on beef consumption..]

U can only handle small

Hects will enentually cancelout

Us expected to be o

hos i low

- you cut never be perfect

Other preliminary considerations... bzfudgement ca*ll*

• 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

• Two main uses of a regression model are:

To estimate the magnitudes of the effects of the explanatory variables on the dependent variable (i.e. estimate the parameters in the regression model) To obtain predictions for the dependent variable (10

be discussed next) [for purtulan set of waters

• Estimating models of behavioral relations, such as demand or production functions, is key to conducting economic analyses

markety factors.

i s**ubsuet-titü denotes** obsendon # k-, expl var so # of panasuster =k -fseyptar (BI)

Population vs Estimated Model

• Conceptually, it is very important to differentiates

between the population vs the estimated model 7

W The population model is denoted by (show data):

\*Yi - Bi+B2Xi2+...+BXiktui (i=1,...,n); } 7 where Yi is the observation on the dependent

variable, Xiz,...,Xik is įth observation on the k-1 independent variables, B1, B2,...,Bk are unklown

population parameters, and ui is the ith error term

• In this model, each value of Y (Yi) can be decomposed into a systematic, predictable component (E[Yi]=B1+B2Xiz +...+BXik) also known as the PRF, and a random, unpredictable

component called the error term (ui)

& expected value ty

Lo condival on oso

Popmodel: 1;= 3 + B X 12 + 2 ks ... BK Tnt to

ed i lsin qiven thosey

observed

*Val*ve *Oy*

obscheelth

The Estimated Regression Model

• If data on Yi and Xi2,...,Xik (i=1,...,n) are available, the unknown values of B1,...,Bk can be

estimated (show how to do this in Excel)

• The estimated values of B.,...,Bk will be denoted

by B1,....Bk and called parameter estimates

• The estimated regression model is denoted by: e fyi Bi+B:8+ BXik+ûi diff but knowing

> In thie estimated model, Yi again can be decomposed

into a systematic and a random component The systematic component (Ựi=B1+B2X12+ +BwXih) is also known as the SRF or predicted value of Y The random component (li) is called the residual

Bis unknown predict w/ yixidata

contingent on data More date is better

Not thie popullin pariter NOT BETA

paruter estimate: B

Data

equivo

applies toborn

Know*n*

assume is process that has genenited

yourdata

The Population Model

• 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 (Xi), it is assuined that the population model generates the Yi data by adding a random

error to E[Yi]

• For instance, in the case of a single-variable model, E[Yi]=B1+B2Xi, which graphically is a straight line

Graphical Representation of Population Model

= nowned distribe, va*ri*ance or

and mean of a

PRF E[Yi].....

IIIIIIIIIIIII

Assume that population

model generates the observations... could have many more!

B1 B2 unknown Data used to

(est Bit 8

[Yi]

50

55

X1 60

65

70

75

80

Graphical Representation of Estimated Model

Yi - B, + B2 y + U walnt to be close to population

SRF

55

Xi 60

Slide 12

OR6

Bih=5, B2h=0.12. Xi=58, Yih=12 Octavio Ramirez, 8*/2*0/2013

Two Independent Variable Case

3

yi= Bia Brxiz & Bzřiz + 3

mone B plane nothing

Estimated Model (Plane) în =B1+B2X12+B3X13

X3 slope measured

X2 slope measured by B2

by B:

Aug 25

ols req9 of 19

- Non math v reg

ef not met cunt do ors ♡ only for lineur in Panuiters

1 - can be won linear

en vart

Yi = Bit B2 X iz t ui non leur on x

can estimate Y;= Bit B2 x Brui cunt est ble

x2 use x2 ate exp var non linen en punitens

3

#obser> # parimeter

n = #obsen

ko #paniter le en es) \* only for als to work → ols will be bud est e least 4 obser per pameter, a little less me okay

but over 3

& Some var level in sumpre val of exprar

explan var NEEDS variabity la lot Dummy rart make sone ration woulou

xs

a = fixed const

6 not perfect multical linearity

dit d z \* 2 \* ... \*kaikto for alli

if = 0 = pery multicar " It is perfet unor function to othens

crop yeilds Ran full (sum, wint, full, spriny)

annual Rainful X12 + xis & xin tris-rice =0 i

Sp sum full w Annual - Class jr, fresh.soph, natsra

double counting

"joint determina*to*n"

Aug 25 cont -dep var feed back on effect on explevar

- fixed external - random: y cant effect & back

- ny xay. [#deer & food availibity]

i

[Mkt pro ] For

- ZSLS

[vor asus

to tow stuces least squad 2 least is

expected Uizu in practice Requines:

'no key relivent expvar excelded from systemet comp

to funct form es correct speelt

popmodel: Y; \*B, + B2712 TB3%; 3 TV

Elulo

BE: Yi =B11B 2 Xiztei

Bz xiz tui

Es unless X13-0 then

Elei] = By xiz

"violaton is incorrect specit .

:: 93 But By xiz

Some curvatuer still guess linear

uinean non linear some anias where I = (-) ;

or û = (+)

Aug 25

No consistan

ef either one doesn't hold? No consistancy or unbias.

ef est is unblas is consist but ops is not true

Avg 18

AAEC 6610 Quantitative Techniques in Agricultural Economics

Tues: Quiet How you come up w/ B & B

Topic II OLS Estination

Estimation of the Model Coefficients

obj come up w/ B, that is close

OLS:

\_

• It is obviously desirable that the estimated

regression model is as close as possible to the true population regression model; 1.e. that the estimated parameter values (B1, B2,...,Bk) are as close as possible to the true population

paraineter values (B1, B2,..., Bk)

• The Ordinary Least Squares (OLS) method to

estimate B1, B2,..., Bk (or a variation of it) is in many cases best at accomplishing this ,

objective

CAqu*ks lul5)*

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:

Minimize RSS = minimize 4.2

minimize S, - ,

- minimize E(Yi - B1-B2X1)?

• The RSS is a function like any other, in which

the Yi's and Xi's are known constants (the data) and Bi and B2 are to be viewed as the variables or unknowns which can take on different values and have to be “solved for"

mm sum af as a residuals

min resid sumay as AKA: the ui distance to line squared and summed want as small as possible

Рntүnіn а sом - So4 am neg tsome are pos cantake

Il resid/ or Ž resid. ?

= Bi & Rariztui

û = 4,9 flyi-Bi-Ba xiz

penilize winger diffences

Better

more. graded wt. I

Yitxi dutu ginen, thet as known

Bit B2 = variables

constant

2.

\* Chain Rule

Ignore somaton to do part dern.

de Hi-Bo-By x)2 = 2

The Ordinary Least Squares (OLS) Method

To find the values of those variables (Bi and B2) that

minimize the RSS (E(Yi-Bi-B2Xi)?) the partial derivatives of the RSS with respect to those variables have to be taken and set equal to zero:

- ORSS/DB1 = 0 [E(Yi - B1-B2X1)”]/BB170

- ORSS/ĐỐ2 = 0 {E(Yi - B1-B2X:)2)/2ß2 = 0

• The former partial derivatives are (chain rule):

- -23(Y.-B1-B2Xi) = 0 2214- X - :. Y-CB1-B2X =EY-B1(n)-B2(EXI)= 0 - -2EX:(Yi-B1-B2Xi) = 0 - :: EXY:-EXIBı-EX.B2Xi = EXi Yı-Bı(EXi)-B2(EX:-)= 0

40ҮМ*К Ruu*ti*o*n

-

ý ý tôn =P - System of 2 equations

| The Ordinary Least Squares (OLS) Method

• Notice that the two equations (in the boxes)

resulting from setting the two partial derivatives equal to zero make up the following system of two linear equations with two unknowns:

> (n) B1+ (EXi) B2 = {Yi See Back

1 L (EXi) B1 + (EXi2) B2 = EXiyi 1 Pi L: This can be easily solved for B1 and B2 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

somyn

4

moved

yo mer

The Ordinary Least Squares (OLS) Method

• The formulas are (discuss summation petation):

nEXY-EXiZY (Xi-X)(YY)

nXX--Xi)2 E(Xi-8) 2(XI-X)Y,

to the sum of in resid

(X-X)? *1)* and

two give you

I

Bi = Y-B:X (gel first!) The

• These are the formulas to calculate the OLS

estimates (Bi and B2) for B1 and B2

• Briefly discuss the multivariable case

The Ordinary Least Squares (OLS) Method

• Discuss matrix representation of population

and estimated inodel (make sure to use Xij

notation)

• Derive matrix algebra formula for OLS

estimator

• Show example in Gauss

resum of predicted -Sum of obeerval

OF 2(x,-ŷ) = {xi { fo.

The Ordinary Least Squares (OLS) Method

• A regression model estimated by OLS method

has the following key mathematical properties that always hold (show all in Gauss example): VEû = 0 (follows from the first normal equation) VEYi = £Yi (follows from the above result) \*\* It always passes through the point (Y, X) - why? The residuals (lli) are not correlated with the values of

the independent variables (Xi, for j=2, k) or with the ( predicted values of the dependent variable (Yi), ie. t

Xijùi=0 (j=2,...,k) (also from normal equations) and

Yiui 0(i=1,.,n) Tcasier to show with matrices),

I

equat of interest

ist.

moltivar

R3

OLS Requirements

• These are minimum "technical” requirements that

have to be satisfied in order to even be able to obtain parameter estimates by OLS: VOLS can only be used to estimate models that are linear

in the parameters, although the inodels can be non linear in variables (examples) ✓ The number of observations must be greater than the \*

number of paranieters to be estimated (rule of thumb)

✓ There must be some level of variability in the sample

values of the explanatory variables (the more the better ✓ More generally, there can't be perfect multicollinearity

(discuss in some detail, give example). \*4

equations Sexilyi-Bi-Bit) Exi (Oi)

\* 4 = B1 + B2 712 tui (ran var

Y = Bit B2

xs tui (\*)

\*eleast 4 parimeter \*3 dummy var 50/50 \*\* Crop Heild example Feasonal tannual

rainfull] ne x can bea funchon of other's why? def [xx'} 0 → [xx]\*= does not exist

cantest B

\* f not

B not best/wrong

PRI

OLS Assumptions Assumptions are things” that have to hold or be true in order for OLS to work best: \* 1. The values taken by the explanatory variables are obsevda

fixed in repeated sampling or, if randon, they are not correlated with the error term (Cov[U.Xj]=0, j=2,...k). In practice, correlation with the error term might occur if:

nee*d r*ainfallout none tell

· The explanatory variables are proxy variables or if

they are measured with error (show why, can usc IV)

tv The dependent variable has a “feedback” effect on

any of the explanatory variables (i.e. they arc jointly determined) (can test for endogeneity problem and address through IV, 2SLS, etc.)

use near buy tu

decided

fone,

hand

I

xit could becomel w/us

Matter of degree small is

*о*ил

Yean effecty

instrimentul

variation

32 stage - Lease

nes

back

-

"Joint detemind mat(p=0)

OLS Assumptions The expected value of the error term is zero (E[U]=0) which, in practice, requires that:

No key relevant explanatory variables have been excluded for the systematic, component of the

model (show) *ac*count for *a*ll relß\*

» The functional form is correctly specified, i.e. that

the true relationship between Y and X is linear *V*. eppur

✓ Show that, if assumptions 1 and 2 hold, OLS

yields an unbiased or at least a consistent

estimator for B. \* ✓ What is an unbiased estimator, what is a

consistent estimator? (briefly, more later) !

makes üi non ♡

istrincar os.esponetc.

i con [vidz]=0 Lin should hold for any

OLS Assumptions

combo

onestirme

situatur

is violated

3. There is no autocorrelation (Cov[ui,uj]=0 all

i+j), i.e. the error termıs are independently distributed i.e. the error terms associated with different observations are totally uncorrelated it with each other (focus on concept for now,

will discuss in detail later) 4. There is no heteroskedasticity (Var[ui]=0,

i=i...n), i.e. the error term variances are the same regardless of the values taken by the explanatory variables (focus on concept for now, will discuss in detail later)

var of ui= var uzz ... un

= 02 sometimes varane dif due to x attect og to affect var iny

x=

sample

mein

Some

**randon**

van

Covariance Matrix

Nx)eski-eck)"}= - measure of dispersal

oute

tugo

Tor 2 | 3 -

✓ Define the variance (of X1) and the

covariance and correlation (X1 and X2) -

and discuss how to estimate them ✓ Explain the concepts of covariance and

correlation ✓ Discuss error-terni covariance matrix:

Var[U]=E[(U-E[U]/(U-E[U])'] (expand)

✓ Define an iid error ✓ Show that if error-term is iid, Var[U]=0 |

where I is an nxn identity matrix

*ЕГе*

compute ECARE

И-

unlaus

estim*a*tor

cover (214 XL)- e{xi-Exu) (2 - 5%)]

me ad

dexaton af te from man

& dev. af xz".

Icov b/w war. ditself=variancesc

est:

ôž - Î (\*14-és)(xi2- XD) Efail-o

*(*n-1)

Calculating the OLS Standard Errors

• Discuss why B is a random variable

• Explain what the covariance matrix of B is

• Derive the OLS estimator for this

covariance matrix using matrix algebra

• Explain how to obtain an estimate for

oʻ=Var[ui] - i.e. the error term variance

• Explain how to obtain standard error

estimates from this estimated covariance

matrix (do it in Gauss)

• Emphasize difference between the true SE

and the SE estimates

The OLS Standard Errors

• Each parameter estimate has a standard error

(SE) associated with it

• The standard errors provide a measure of the

degree of precision with which the OLS parameter estimates (i.e. the B's) are being able to approximate the true parameter values in the

population regression model (i.e. the B's) – why?

• A rule of thumb is that the true parameter value

would be within + two standard errors of the

corresponding OLS estimate

• Go back and discuss example in Gauss and Excel

lowest

o af corelcofb/w kz fx3 you

*y ♡* -> Hugleist possible SEE

2 total corulla

undfu

A Note on the Standard Errors

• From the formulas for computing the standard error

estimates in the two independent variable case:

terror termuar n see nontonment 1

>ÛTÊ. Our cx21-Xay'*up Sca*ler oppee the square routa Standerrorest. >V[3] = G:/{(1-1)2X30-X3) ;

Sum of u deviaton of x2 + Er Numar - ľVar=&SEE allows to varez

Show effects. no var=undth

where r23 is the correlation coefficient between X2 and X3 (recall correlation coefficient). Note that: >The larger the error term variance (o?, i.e. the level of unexplained variation), the larger the standard errors

The larger run, the larger the standard errors The larger X(Xji-Xj)” (i.e. the amount of variation in Xji), the smaller the standard errors

về ở tu-.) (x )

**n**

**.1*/***

dfn propertes of

multipul de tusets

Properties of the OLS Estimators

• The Gauss-Markow Theorem states the properties of the OLS estimator; i.e., of the matrix formula used to calculate B given the sample data: \* It is unbiased (E[B]=B); this means that, on average, or

ge overvepeated samples

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 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

#mi,

sie most out

+

min var = lowest poss. SEE

good = prease most preelse

Properties of the OLS Estimators

• The Gauss-Markow Theorem also states that:

IFOLS assumptions I through 4 hold, the OLS estimator also has the minimum variance anrong all possibld unbiased estimators that are linear functions of the dependent variable Y (discuss why OLS is a linear estiniator) Combined with unbiasedness, this means that, if assumptions 1-4 hold, the OLS method is the most likely yield an estimate for B that is close to the actual unknown value of that parameter vector Discuss Unbiasedness-Efliciency Handout

cavayut among all possible UN BIAS estuu possile bas mone precise est; Blas v. eftiuency linear B = (c'>)"ty

kan nou

\*na"

Ŝ =mi me

Seenote

could be non linear estimator

that is more efficient I finalaugoti ef ernostermis normal L alstro then als es min var most ef of all unbias

linear or not linear

Suptostart

Best

in most

efficent

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 in that it is only best among all possible unbiased estimators that are linear

functions of Y

• However, if error term (ui) has a normal

distribution, the OLS estimator has the minimum variance (i.e. it is the most precise) among all possible unbiased estimators for B 19

a Best in this

**Cluss**

> minvar

Luneen funchon ofy Not Sano asmedel

being linen

Sumaf

*3*

30 septz

and and errors - MP3AD - og (1-yis da Cxs-x)%] square devahona

uf XZ 2X2 (varxa}

"Squar of cornel bw x2 +48

47. tvaru SEL

"No var= undfa

error termt; BEER [wantsomin]

yo lowest possible SEE

o per tres of oLs Estimafors -Ŝ is unblased: estm. expected to equal val- of unknown coef

of population

W AN; mone Bon average = B

- requires Assomp.lt2 - eg 1a-4 Holds: ons estim, alsso has min variance among all possible

UNBIAS estimators that are LINEAR function ofy

s More precise

- w/ unblas ons

is most likely

to yeild est

for B

close

to actual is

\* is uis normal distrito s is min var of all untas (linor not)

Chept 3

Septle start.

AAEC 6610 Quantitative Techniques in Agricultural Economics

Topic III easures of Goodness of Fit

How good is Best fitting model

to predicty?

Measures of Goodness of Fit

• 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, it's BLUE

• However, the question of how well can this best

fitting model predict the behavior of Y remains unanswered

• The main statistic use to quantify how well the

estimated regression model "fits" the Y data is the R?

The R2

• The R2 concept is based on the notion that each

Y observation (Yi) can be decomposed into total,

explained, and residual variation (wrt Y)

• 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 Yi-Y)

• Explained variation is measured as the square of

the difference between the predicted value of Y

and the mean of Y, i.e. by (Yi-Y) |

• Residual variation is measured as the square of

the difference between the observed and the predicted value of Y, i.e. bylYi-Yi)-1

unexplahed

I Nomentor for var u

The R2

• Specifically:

→The total sum of squares (TSS=E(Yi-Y)-) measures

the total amount of variation in the observed values of the dependent variable (numerator of variancc) The explained (regression sum of squares (ESS=E(Yi-Y)?) measures how much of that total amount of variation (i.e. of the TSS) is being explained by the estimated regression model > The residualsum of squares (RSS = (Yi-Yi)2=

£(ui)?) measures how much of the total amount of variation (TSS) is not being explained by the estimated model Mathematically show that TSS=ESS+RSS! SUNOT

>Thus RSS/TSS measures? ESS/TSS measures? +

minimize

\_TSS= ESS+ RSS

The R2

explained ss = Regressionss tegre residualss= Emorss

• Therefore, the R\* is calculated as:

R?=(Yi-Y)*?/ E* (Yi-Y)2 = ESS/TSS) >R2 = 1 - 1 Eủi? */* (Yi-Y)2} = 1 - (RSS/TSS)pr

• In the first case, note that ESS/TSS measures

the proportion of the total amount of variation in the dependent variable that is being

explained by the model

• In the second case, since (RSS/TSS) measures

the proportion that variation that is not being explained by the model, 1-(RSS/TSS) is also the proportion explained - why equal? :)

Plug into

• TSS=RSS+TSS

Both Suma Dame san units . So R2 has nounit. no a wonita

The R2

• Thus, the R? measures the proportion of the total

variation in Y that is explained by the model (i.e. X)

• The R2 is a ratio, it has no units, its value is the same

regardless of the units in which Y is measured

• It can only take values between 0 and 1 (why?):

R? = 0 (RSS=TSS, ESS=0) indicates no fit at all (Y and

X are not linearly related at all)

• R2 = 1 (ESS=TSS, RSS=0) indicates a perfect fit (there is |

a perfect linear relationship between X and Y - no error) -

• The higher the R2, the better the fit.

• In a simple regression the value of the R? equals (ryx)2

(i.e. the square of the correlation coefficient between Y and X), in multiple regression...

бе? 4)

-

Horizontal = R2 undth

ŷ, = Bit 87 x = Y sos Elyi-y) = {ly, ale

- all yi on to straight line -Elyi-o = z lŷi -ý)\*

simple=lvar

Corrlz=R2 time series: data on particular

entiney measured over time cross sectimeli yerlds; data

for one point of time, different

A final important note is that the R? is only\*\*\* enthes; same time

MPORTANYT!

HESS ] Pool panuel datu Mix of both

- 4 How well cun model predict Y

can explain how axiwillsy

predictions but

ß Low se] SET R2 are both

rady

important

The R2

• The value of the R’ must be assessed in light of the type of data being analyzed >In models estimated using time-series data a good R?

(i.e. a good fit) is perhaps 0.80 or higher In models estimated using cross-sectional data a good R? (1.c. a good fit) is perhaps 0.50 or higher >The Ris only a measure of the model's capacity to predict Y, so models with very low R2 but precise parameter estimates serve other useful purpose...

“valid” if the model includes an intercept, otherwise Yi+EYi (why?) and TSSERSS+ESS

uptg

The R2

• A disadvantage of the Ras 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

(or at least the same) RSS

• An increase in the RP as a result of adding an

independent variable to the model does not mean that the “expanded” model is better, or that that variable really affects Y in the population

quor R2 + still useful model? add cop ver, no matter what it is most likely ↑ R2

one way to account

put lots of var @ I time out still assess if of one is

*I unde*

The Adjusted R? 2 • The adjusted R2 denoted by R’ is better to assess

whether the adding of an independent variable likely

increases the ability of the model to predict Y: Modell • R2 = 1- Var(u)/Var(Y) = 1-[{RSS/TSS}{(n-1)/(n-k)}] ||

• Notice that the R’ is always less than the R-(why?

A-1 /

unless k=1 or RP=1

n

· Unfortunately, the R? does not have the same

straightforward interpretation as the R'; under unusual circumstances, it can even be negative

Submit Some argue that the R’ provides for a better measure mone

| of goodness of fit when the model is estimated with a This town isa penalty

Ts3

| lots of variables and few observations as K f k esep ba*r*u

diways useful

for

a

Anor 1x4

7

I duesht

-

gonor

These two

terms hawlto

*own*vide e*ac*hother

matter

toor?

. If y,= ý tŷ, =

Ý Rz undfo :TSS=0

ruse proxy

use indirect to est.

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]#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

those want to putin if believe

carily)<t ef tiney etteet + don't put ino

rinu

not always just time may be ttor logs

period I, Z...

tastetry

**Bhat**

-1.5391

TIME

(year)

**Xi6**

co joc cos são vou AWN

**Parameter Estimates *(*estimated from data)** B1hat B2hat

B3hat

**B4hat B5hat**

**141.5516** -0.1469

0.1771

0.0331 0.000175 **Observati**on BEEF CONS. BEEF PRICE CHICKEN PRICE PORK PRICE INCOME (year) (lbs*/*pers-year),(cents/lb) (cents/lb) (cents*/*lb) *($/*pers-year) - Yi ove Xi2

**Xi3**

**Xi4**

**Xi5**

118.80 431.10

179.30

381.40 13760

127.50 390.20

159.80

359.00 14242

124.00 366.90

151.10

315.20

14691

117.90 417.90

155.40

335.50 15349

105.50 466.90

142.10

302.50 15547

103.40 432.00

131.10 257.90

15186

104.20 393.50

122.70

255.50 15099

104.30 *3*76.50 112.70

*2*77.00

15076

106.40 358.20

110.90

259.80

15299

106.20 345.40 118.80

2*3*7.70 15954

106.90 323.80

108.10 229.50

16427

108.00 315.40 116.10 248.10

17090

104.00 319.80 105,30

252.80 17507

103.10 322.40 110.00

236.30 17804

98.40 326.60 113.90

224.70 18193

96.10 32*7.*70

104.80 247.90

17667

95.40 322.60 98.50 237.10

17225

95.00 309.10 94.40 215.10

16985

93.00 309.40 93.90 208.40

17524

20 96.40 290.90 93.50 203.70

17929

97.10 284.30 91.70

194.80 18143

\*Dataset is a sample from the population

♡ ♡ oo to co Ñ Í ö on voor AWNA

PRED. VALUE RESIDUAL UI (lbs*/*pers-year) (lbs*/*pers-year) Yihat

**Uihat**

123.44

-4.64.

123.80

3.70

122.78

1.22

115.29

2.61

103.14

2.36

103.24

0.16

105.78

-1.58

105.67

-1.37

105.97

0.43

107.10

-0.90

106.65

0.25

108.49

-0.49

104.62

-0.62

103.04

0.06

101.26

-2.86

98.62

-2.52

96.28

-0.88

95.23

-0.23

93.43

-0.43

94.46

1.94

93.31

3.79

add to ø

0.00

a will alwasss

21

" 1st subscript = obser

2m subscrpt= Dingx

online

I look

overtime

SUMMARY OUTPUT

*Regression Statistics* Multiple R 0.975576

95% Beetrar explabe

this model

quare 0*.*951748 Fano *e*ss*/*tss

w/5 factors

Adjusted F 0.935664 Standard E 2.434128 Observatic 21

means of

ESS

m70% i

this

th Brany

*Ra*ng

<speedo

*eu*

en

ANOVA

*df* SS *MS F ignificance F*

Regressio 5 1753.0114 350.6022 59.17356 2.44E-09 Residual 15 88.8747139 5.924981

Total

20 1841.8857)

SEB *Coefficientstandard Errct Stat P-value* L*ower 95% Upper* 9*5%*

Intercept 141.5516 19.725 7.176 0.000 99.508 183.595 B2 No.67 -0.1469 0.022 -6.779 0.0000062 -0.193 -0.101 B3 0.177*1*

0.088

2.021 0.062 -0.010 0.364

B4

0.0331 0.037 0.892 0.386 -0.046 0.112

B5 0.0002 0.001

0.162 0.873 -0.002 0.002

B6

-1.5391 0.358 -4.300 0.001 -2.302 -0.77*6*

*ower 95.0%pper 95.09*

99.508 183.595 -0.193 -0.101 -0.010 0.364 -0.046 0.112 -0.002 0.002 -2.302 *-0.*776

paruultor estimate

were obsero

#

w

RESIDUAL OUTPUT

**دا د**

**ل**

*Observatio. Observed Predicted Y Residuals*

1 118.80 123.441 -4.641 2 127.50 123.802 3.698 3 124.00 122*.*775 1.225 4 117.90 115.290 2.610 5 105.50 103.138 2.362 6 103.40 103.240 0.160 7 104.20 105.77*7* -1.577 8 104.30 105.673 -1.373 9 106.40 105.974 0.426 10 106.20 107.097 -0.897 11 106.90 106.649 0.251 12 108.00 108.492 -0.492 13 104.00 104.623 -0.623 14 103.10 103.040 0.060

98.40 101.258 -2.858 16 96.10 98.622 -2.522

95.40 96.282 -0.882

95.00 95.231 -0.231 19 93.00 93.431 -0.431 20 96.40 94.455 1.945 21 97.10 93.310 3.790

APPENDIX D

STATISTICAL TABLES 961

TABLE D.2

PERCENTAGE POINTS OF THE DISTRIBUTION Example Pilt > 2.086) = 0.025 Pilt > 1.725) = 005 for di = 20 Pret 1.725) = 0.10

0.05

. one**rai*l***

0

1725

0.25 0.50

0.10 0.20

0.05 0.10

0.025 0.05

0.01 0.02

0.005 0.010

0.001 0.002

1.000 0.816 0.765 0.741

3.078 1.886 1.638 1.533

6.314 2.920 2 353 2 132

12.706 4.303 3.182 2.776

31.821 6.965 4.541 3.747

63.657

9.925 5.841 4.604

318.31 22.327 10.214 7.173

0.727 0.718 0.711 0.706 0.703

1.476 1 440 1.415 1.397 1.383

2015 1 943 1.895 1 860 1 833

2.571 2.447 2.365 2.306 2.262

3.365 3.143 2.998 2.896 2.821

4.032 3.707 3.499 3.355 3.250

5.893 5.208 4.785 4.501 4.297

0.700 0.697 0.695 0.694 0.692

1.372 1 363 1.356 1.350 1.345

1.812 1.796 1.782 1.771 1.761

2.228 2.201 2.179 2.160 2.145

2.764 2.718 2.681 2.650 2.624

3.169 3.106 3.055 3.012 2.977

4.144 4.025 3.930 3.852 3 787

0.691 0.690 0.689 0.688 0.688

1.341 1.337 1333 1330 1.328

1.753 1 746 1.740 1.734 1.729

2.602 2.583 2.567 2.552 2.539

2.947 2.921 2.898 2.878 2.861

2.131 2.120 2.110 2.101 2.093 2.086 2.080 2.074 2.069 2.064

3733 3.686 3.646 3.610 3.579

0.687 0.686 0.686 0.685 0.685

1 325 1.323 1.321 1.319 1.318

1.725 1.721 1.717 1.714 1711

2.528 2.518 2.508 2.500 2.492

2.845 2.831 2.819 2.807 *2.*797

3.552 3.527 3.505 3.485 3.467

0.684 0.684 0.684 0.683 0.683

1.316 1.315 1.314 1.313 1.311

1 708 1.706 1.703 1.701 1 699

2.060 2.056 2.052 2.048 2.045

2.485 2.479 2.473 2.467 2.462

*2.*787 *2.77*9 *2.*771 2.763 2.756

3.450 3.435 3.421 3.408 3.396

1.310

1 303

0 683 0.681 0.679 0.677 0.674

1.697 1.684 1.671 1.658 1.645

2.042 2.021 2.000 1.980 1.960

1.296 1.289 1.282

2.457 2.423 2.390 2.358 2.326

2.750 2.704 2.660 2.617 2.576

120

3.385 3.307 3 232 3.160 3.090

1

N(0,1)>

*Note* The smaller probability shown at the head of each column is the area in one tail the larger probability

Is the area in both tails

S*ource* From E S Pearson and H O Hartley, eds. B*iometrika Tables for Statist cian*s vol 1,3d ed table 12 Cambridge University Press New York 1966 Reproduced by permission of the editors and trustees of *Biometrika*

Explan on

P2

**<a hң,**

Deter

-

1

*1*.1

19

Table 1: Illustration of Unbiasedness and Minimum Variance Properties:

Based on 40 Repeated Samples Containing 25 observations each

w*a* True Population Model is Yi = B1+B2Xi+Ui = 10+1.0Xi+Ui

Y= 10+lri\* Vi (B Not B)

Sample Number OLS Estimate Alternative A

Alternative B

0.95 0.82

0.93

1.07 0.68

0.99

0.89 0.98

0.90

0.94

0.92

1.17 1.28

1.04

0.76 0*.*72

0.83

1.11 0.76

1.01

1.15 1.48

1.03

9 1.05 1.28

0.98

10 0.88 1.62

0.89

11 1.03 1.1

0.97

12 0.82 0.58

0.86

13 0.97 0.78

0.94

14. 0.9 0.66

0.90

15 1.31 0.92

1.11

16 0.87

0.89

17 1.12 1.2

1.01

18 1.02 0.86

0.96

0.88 1.08

0.89

20 1.07 1.18

0.99

21 1.05 1.06

0.98

22 0.79 0.64

0.85

23 0.89 0.94

0.90

24 0.83 0.8

0.87

B2~

0.96 1.62

0.93

26 1.05 0.74

0.98

1.1 1.24

1.00

0.93 1.04

0.92

29 1.04 0.76

0.97

1.09 1.14

1.00

0.91 0.9

0.91

0.84 1.14

0.87

0.99 0.78

0.95

34

0.88

0.95

1.14 1.34

1.02

36 0.86 0.52

0.88

0.88 1.22

0.89

1.24 1.3

1.07

1.14

1.02

1.31 0*.*76

1.11

**Avera**ge:

0.95

**Stand. D**ev.:

**0.137** s 0.*2*74

0.068 <0

get se for each

**un tota laut**

on average of Blas)

Parum est wat

ressef

but lomer

szivar

25

27

28

3*0*

31

32

33

1

35

3*7*

38 39

1.1

40

**1**

the se

Table 1 illustrates the unbiasedness and minimum variance (i.e. efficiency) properties of estimators such as OLS. As discussed in class, if the error term is normally distributed 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 Bj).

Table 1 assumes that all those conditions are satisfied and OLS is used to obtain 40 estimates for Bj (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 Bj is known to be Bj=1 (which would not be the case in practice).

Note that the average of the 40 OLS estimates for Bj is equals 1, which means that OLS is an unbiased estimator for Bj (in reality, depending on the sample size, more than 40 samples might be needed for the average to be exactly equal to 1).

Alternative A illustrates another estimator (i.e. formula or procedure) to compute estimates for Bj 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 disperse 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 estimator A range form 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.

In fact, according to Gauss-Markow, OLS would have the lowest possible variance among all estimators that are unbiased and are linear functions of the Yi's. If error-term normality is added to the Gauss-Markow conditions, OLS would be minimum variance unbiased without qualification, i.e. there would be no other unbiased estimator with a variance that is lower than OLS's. The combination of these to 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.

Alternative B illustrates another alternative estimator which is biased but has a lower variance that 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*?*

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. This 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.

\*Note that in practice only one sample of size n and, therefore only one estimate for Bj is actually obtained.

Worley

Pexam atterthis

D

Scots

AAEC 6610 Quantitative Techniques in Agricultural Economics

Chapters 4 and 5

The Normal-Error Model Confidence Intervals and Hypothesis Testing

t

assume ols asumphad Normality

The Normal-Error Model

• In the normal-error model, in addition to the usual error-terni assumptions:

> Cov[U,Xj] = 0 >E[ui] = 0 for all i, Var[ui] = o’ for all i, and

Cov[ui,uj] = 0 for all itj the error term is also assumed to be normally distributed: Assump2

> U~ N(0,021) (briefly discuss what this means)

• This assumption is supposedly supported by the

Central Limit Theorem (discuss briefly)

The Normal-Error Model

• Recall that under OLS assumptions 1 and 2:

Cov[U,Xj] = 0 >E[U]= 0 and and the OLS formulas are unbiased and consistent estimators for B, that is: E[B) B and Plim[B]=BS

Basist » OLSHOW

• If, in addition, the error term is i.i.d. (i.e., OLS assumptions 3 and 4 hold), Var[U]=oʻl and:

Var[B] = 024X’X)-1

The Normal-Error Model

up

B = (x x x'y 8=13\* (\*\*\*)\*\*'U

kinemfuacomafu

***And*y**

• If all four OLS assumptions hold and, in 616 addition, the error term is normally distributed:

>B~N(B,04X’X)'), thus: unhas*l*inda >Bj ~ N(Bj, Var[B;]) j=1,...,k; where Var[Bj] is the

jih diagonal element of oʻ(X'X)"

• Normality results from the fact that the OLS

estimator is a linear function of U, and U is normally distributed (this result only holds asymptotically if the X's are random, in small

samples the distribution of B would be in “approximately” normal) x es fixed "

large

bout Raridom x?

sumpre

linear function of Normal r.)

Lame also N strictly speakinyx holdsfor Large Samples

Why dowe cave? basis of cift for paralter & Hypoth test I

The Normal-Error Model

• The former means that if multiple samples are taken

and the OLS estimator is used to obtain multiple estimates for a given parameter (Bj), one can think of these estimates as being drawn (i.e. coming from) a normal distribution withi mean B; and variance Var[Bị]; i.e. each individual estimate is being drawn from that distribution (back to ble handout) As such, any particular estimate will be expected to:

Be within Bj Se[B] with 68% probability Be within B; + 2rSe[Bj] with 95% probability >Be within Bj † 3.xSe[Bj] with 99.8% probability

Stund Normal

t distrib: Fatter tails urger af 400 Tapprochs N(001)

Confidence Intervals

• Also note than since B; ~ NABj, Var[B]):

>Z = (Bj-B;)/Se[Bi] - N(0,1), where Se[B;]=VVar[B]

• However, recall that o?, which is needed to

compute Se[B]], is unknown and has to be estimated by @?, which means that one only has access to an estimate for Se[Bj] (denoted by Se[B;])

• Fortunately, it can be shown that: tum N *Y*SK.

>1 = (Bj-Bj)/Se[Bj] - In-h) (a t distribution with n-k degrees of freedom, where k is the number of

parameters in the model (briefly discuss Z versus

i distribution)

liste

nok af

2 things we can do

say probe

Confidence Intervals

• From the previous expression one can obtain a confidence interval for Bj as follows: » Pr{-trav2.+k) < 1<trox2.1-x)} = 1-a (quick example), i.e.: » Pr{-txa/2, 1-10 < (Bi-Bj)/Se[Bi] <t/a/2.0\_k)} = 1-0 which, after multiplying all elements in the inequality by Se[Bj], subtracting B; and multiplying all terms by -1, which reverses the direction of the inequality signs; one obtains: »Pr{Bi-l(w2.nky'Se{B}]<Bj<Bj+1(0/2,1-k@e[B]}= 1-, wst

which is a confidence interval for Bj (do example versus Excel)

Not the ef VAAN

7

assful

s i

distribs

BYO

sept is

i poprtadory Parunter

lugin

Start

Hypothesis Testing: t-Tests

• Also, recall that:

m1 = (Bj-B;)/Se[Bj] ~Pek) Trahu

• And note that the t ratio above would only

follow a t distribution if the correct (i.e. the true population value) for B; is used to compute the ratio Need Sassumpt.

• The more the Bj value used to compute the

ratio departs from the true population value the less the t ratio would tend to be consistent with a draw from a t<n-k)

95% confident true effect b/w [ 19. 10] 5% poob B2 outside

this range compute toner seneral sets sets oft will follow at distrip w/ nok degre of

robe true

freeldu

*E*L BJJ

Jes unbias,

5o assume

> sonst

critical two tuiltest

table

t = 2.131 igofreedom

Hypothesis Testing: t-Tests

• Therefore, the hull hypothesis (Ho) that B;=C (were

C is any particular number) versus the alternative (Ha) that Bj+C can be tested by:

> Computing t = (Bj-C)/Se[Bj] and ascertaining how likely that particular t value is

to have been drawn from a tin-k), specifically: > Select the desired a-level of significance for the test >Find the table value for a tink) associated with that

partịcular level of significance (1\*(-k)) >If|t1> 1\*in-ky reject Ho in favor of Ha and conclude that

Bj+C with a a>probability of being mistaken (example)

>If|t1<t\*(n-kone can not reject Ho in favor of Ha, but the probability that Ho is correct is unknown (example)

USAN

\* Detuel

look @

later

• Logic of t-test (assume a=0.05, \*(60)=2)

>If C=Bj the t-value (t) would be drawn from tom, >If |t>2, since the probability that a draw from a to takes an absolute value of greater than 2 is a=0.05, it is unlikely that t comes from such a distribution, that is, it is unlikely that C=Bj Thus, Ho: Bj=C can be rejected in favor of Ha: Bj+C However, since there is an a=0.05 probability that does come from a tou), there is a 5% probability of being wrong when rejecting Ho > Only if Ho is rejected, a is the probability of being wrong

in that decision and (l-a) is the degree of statistical

certainty that one can have in that conclusion - Alternatively if, for example, (t=1.3 one could only reject

Ho with an a=0.20 probability of being wrong If Ho is not rejected, one does not know the probability of being right or wrong on this decision

of distrib/

Statistical Hypothesis Testing

• 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 crror

term is normally distributed

• However, it has been shown that if the sample size is

relatively large (100-200 observations), the t-test can be reliably applied even in if the error term is not normally distributed, i.e. for the purposed of the t-test the normality assumption is only required when one

is working with small sample sizes

• However, the four main OLS assumptions are

essential for the reliability of the t-tests

4OLS must Hold tt

+test valid w/o if error n N okay

exor subst nonth wont small sample size to better w/ higher

sample site

work

moisteney

lange

S*e*

Statistical Hypothesis Testing

• The basic most common use of the t-statistic is

to test whether any given model parameter (Bj) is statistically different from zero, which implies that the corresponding independent variable (Xj) likely affects Y:

• The null hypothesis is Ho: Bj=0 (i.e. X; has no

effect on Y)

• The alternative hypothesis is Ha: Bj#0 (i.e. Xj does

have an effect on Y) ... This is called a two-tailed test because (under Ha) \* Bj can be positive or negative, i.e. if X; has an effect

on Y, it could be positive or negative]

B2=0 x2 no effect on y How to tell ef az should be included

unt reject null then dont know

if & has effect result of test dep. ond

Seont be a coloit]

X2P has neg effect Ha two tail is not too generous

attest

the only wory about Pr(B2 co]

no sense to be

Statistical Hypothesis Testing

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

rutio af

>t = (B1-0)/Se[B;] =B;/Se[bi]; Avhere B; is the value of Bj

estimated using the Ots formulas and Se[Bj] is the standard error estimate associated with Bj The reminder of the test is carried our as explained

ratio of /se Bet 721 tend to

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)

se more than 2x By thin

These indicate three possible levels of statistical

tend to reject

certainty when rejecting Ho (90,95 and 99%, Comenikahta respectively) Brtoni 15*%* l*e*vel

of statistical certanity

reject to

pon objecte at **resear*c*h***y. T*hes*e indi*

w/ & invoked

Meds

Statistical Hypothesis Testing

• Alternatively, most computer programs calculate the

exact/lowest a values (labeled p-values) at which this standard Ho: Bj=0 can be rejected in favor of a two

tailed alternative hypothesis (Ha: Bj#0) (example)

• In some cases researchers are interested in conducting

t-tests for a similar null hypothesis (Ho: Bj=C) but under one-tailed alternative hypotheses (either Bj>C

or Bj<C)

• This is generally done when the knowledge of the

issue at hand or previous rescarch strongly suggests that one of the "tails” in the alternative hypothesis can be ruled out a priory (example)

Prul for two tall alt Sassumptions nous tromae iu) is decide to take Bs out have to recale all interpuct on

Anal model assuw of xo és ensane units)

- scale Jother xs

[oscull of xz A values ]

sh

Statistical Hypothesis Testing

• Since Ho is the same, opting for a one-tailed

alternative does not affect how the t value has to

be calculated It valnut att

• 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 tron-kdistribution has been ruled out

• Alternatively, it the same critical table value is

used, the actual level of significance would be *a/*2 instead of a.

stail alt:

Hai Bizore theory or knowinge to fuel out posit or knowlage to rule out one of wo out - Rule out

**l*aut*** S26-)

15

Joz.I I thil makes it easier to reject null

1⁄2 t want to keep &

Sub=postine

effect only

ado

etable

Final Remarks

• When reporting the results of a regression analysis is

is customary to report either the standard errors or the t-values in parenthesis 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 Ho: Bj=0 for j=1,...,k) A “rule of thumb” is that if |B;>2xSe[Bj] (i.e. tj>2) Bj is statistically different from zero at the 95% level of statistical certainty (a=0.05 level of statistical significance)

used to show sext togter

parineter est (se) thenton

pariter est to denate level

of signs Lidealy include pvalue to denot zor

Start

Sept 14

**P**

**.**

used for prediction on by cifur Predmet ons for y ci for the pananiters & Predictions Yo xo B,\* ìo :: yo

obsen &

Confidence Interval for Yo

• Notation (with running example): A>Xo = a Ixk vector of specisic explanatory variable

values va*lu*es that am para - Yo = the value to be taken by Y given Xo unknown

o = Xoß = the predicted value of Y given Xo

ûo=Yo-Yo = the prediction error at Xo

• It can be shown that:

>E[ûo] = 0 (show) Var[uo] = 02+o2X0(X'X)-'Xo' (show) And, if U is normally distributed, ûo is normally distributed as well (slow)

P

Confidence Interval for Yo

• Thus ûo/Se[uo] is N(0,1) and ûo/Se[ớo] ~ tp-k)

• A confidence interval for Yo is thus given by:

Yo tx=k)S@[ủo]

• Briefly discuss confidence interval for E[Yo]=XoB - hint: What's the distribution of:

> (Ỷo-E[Yo])/Ŝe[Yo]?

• Outline process step by step

3 things

know Elv.] =0 compute SELO vo~N (mog

do wo cin 95%

5 will beautside

1.-14.545.0 501

{suki14731

filcan

tunnemenkamal

| أبر 14

iiliii «

1

cil]

50

131

1:10 INU10 21

0

411

41

FIGURE 5.6

Conftlerce wervals funds) for mean Y and individual

values,

O

cl*s;* wait; format /rdt 8,5*;* @standard initial and format commands

same data

from esecel

chicken

income

the

.

let data [21,6] =

Beer? 118.8 431.1 179.3 381.4 13760 1*27.*5 390.2 159.8 359 14242 124 366.9 151.1 315.2 14691 117.9 417.9 155.4 335.5 15349 105.5 466.9 142.1 302.5 15547 103.4 432 131.1 257.9 15186 104.2 393.5 122*.7* 255.5 15099 104.3 376.5 112*.7 277* 15076 106.4. 358.2 110.9 259.8 15299 106.2 345.4 118.8 23*7.7* 15954 106.9 323.8 108.1 229.5 1642*7* 108 315.4 116.1 248.1 17090 104. 319.8 105.3 252.8 17507 103.1 322.4. 110 236.3

17804 98.4 326.6 113.9 224.*7* 18193 96.1 *327.7* 104.8 *247.*9 1766*7* 95.4 322.6 98.5 *237.*1 17225

309.1 94.4 215.1. 16985

93

309.4. 93.9 208.4 17524 96.4 290.9 93.5 203.*7* 17929 97.1 284.3 91.*7* 194.8 18143 @defines the dataset as the 21x6 matrix abové@

U NA V OHHHH PHPPPPP

O O

O *D*OO

WNU

UT LOH

O W PU HNNNNNNNNNNNNNN

O WA NU NW

HHHHHHHHHHO VOUS WNA

OUT A WNHO 06

AHHO

95

21;

01 (

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

U

Y=data [ 1]*; Y;* wait;

1st colum @defines Y as the first row of the dataset and prints it to the output

1st

patch norit X=ones (n,1) data[.,2 3 4 5 6]; *X;* wait; @defines X as an nx6 matrix with the first column being an nx1 vector o the remaining five being the second to the sixth columns of data@

=

Bhat=inv (X'\*X) \*X*'*\*Y; Bhat; wait; @computes Bhat using the OLS formula and prints it, then temporarilly S

*0*

Yhat=X\* Bhat*;* @computes Yhat@

Y Yhat*; w*ait; @prints Y and Yhat side by side@

SY=sumc (Y); SY*; w*ait; @computes and prints the sum of the values in Y@

Yhat=sumc (Yhat); SYhat; wait; e computes and prints the sum of the values in Yhat@

**(n-1)** Uhat=Y;-X\* Bhat*;* Uhat; wait; \*

@computes and prints Uhat@

(nxi) resid

Page 1 of 2

SUhat=sumc (Uhat); SUhat; wait; la computes and prints the sum of the values in Uhat@ u*n*e

mere

Um

i

SUhat2=sumc (Uhat.^2); SUhat2; wait; omputes and prints the sum of the squared values in Uhat@

I resid som of o

X'Uhat*; w*ait; @computes and prints X'Uhat@

should be co

e un correlated

Yhat'Uhat wait; @computes and prints Yhat'hat@ Shoul

lat'Uhat@ should be *so v*

s2= (Uhat'Uhat) / (n-rows (Bhat)); wait; @computes estimate for error term variance@

compute

-

N

ôz

on using

a

VBhat=s2\*inv (X'X); VBhat*;* var B:

stundant enor est SEEst=sqrt (diag (VBhat)); Bhat~SEEst; wait;

do voor

square of diagelents

i=0; do until i>20; i=i+1; @selects observation@

Xo=X[i,.]; @defines Xo@

Yo=Y[i]; @defines Yo@

Dhat=Xo\* Bhat; @computes Yohat@

Sehuoh=sqrt(s2+s2\*\*o\*inv(X'X)\*Xo'); @computes the standard error of the forecats@

t=2.13; @states t table value for 15 df and alfa=0.05@

5

f

(Yohat-(t\* Sehuoh) ) ~Yo~ (Yohat+(t\* Sehuoh)); @computes and prints bounds of 95% confidence interval with Yo in the m endo;

Page 2 of 2

AAEC 6610 Quantitative Techniques in Agricultural Economics

Chapter 6

Regression Through the Origin

Standardized Regression Lags and First Differences Alternative Functional Forms

Septzo

J

intercept=0

3

Regression Through the Origin

• 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 hypothetical 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

• Such a model can also be estimated using OLS, by simply excluding the first column of ones)

that stands for the intercept in the X matrix

• Computer programs often provide and option

for estimating the niodel without an intercept,

why theory/understand if

allx =0 theny zo w/ matrix dilete columbad is for

intercept

NO R2

TSS

& ESS TRSS

Regression Through the Origin

• Obviously, the slope parameter estimates will be

different if the model does not have an intercept

• The R2 is not appropriate is this case as its value is no

longer bounded between zero and one and its usual

interpretation is not applicable

• A "raw" R(Corr[Y,Y]) which only takes values Barme

between zero and one can be used instead, but its

interpretation is not the same as the regular R?

• Using a regression through the origin is only

reconimended if one has very strong evidence that the true population value of the intercept is actually zero and a proper functional forin can be identified

gets you nothing

74 $ or $4 €

Scaling and Units of Measurement

• A change in the scaling of any explanatory

variable by a fixed multiplicative factor of w, only changes the value of the corresponding parameter and standard error estimates by a

factor of (l/wX) (example) Dont change anything

is A change in the scaling of the dependent Led variable by a fixed multiplicative factor of

w Wy only changes the value of all model

parameters and their corresponding standard error estimates by a factor of wy(exanıple)

get 8 mean

Standardlize xt ys and lo Y\*= Yi-I

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

sy

? evar (allxs)

X;\*.

\*;-\*

Sx

Standardized Regression

• Note that, because all of its variables have a mean of zero, the intercept in this model would

be zero)

• The interpretation of the beta coefficients

follows from the fact that both Y and the Xs 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 regressions are not used because of the beta coefficients having a particularly appealing interpretation

selim unit of measure measure

en sal from mean - Yt=1.5 obsere is 1.5sdan Bi 2.5 Axley Iso ya.ssol used for relithe importance of \* in

xx\л\wwК

• Doesn't account for that mone mone

often

langer var

Standardized Regression

• The main use of standardized regressions 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

corresponding X contributes relatively more to

explaining the variation in Y

• Show that/how the beta coefficients can be

computed from the OLS parameter estimates given the results in slide 4

which var has more effect look @stund regression coef

same els coef larger sd more effect

ols per unit onginal dan \* f I I const val toy intercept

a date

Sv sy

***cu*nstaurt** for alli

X; \*. Eta

F

는 카이

from

tuintercept

added

wx= 1

intercept

(wx= Imation - sy

Dashably

I asx

wx

Bio Serie B I

KADW

**Lim**

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: - Yi=B1 + B2X1-1+ ui(i=2,...,T) (T=most recent year) For example, a fariner'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

legg effect sometimus value afx of previous time period that a currenty

time periods avent sepetute time series data T = obseraton

Yel int xt xt1

*X2*

xmana

x

lac value

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

what sects y

value of for xX

First Differences of a Variable

• Another common specification in time series models

involves letting the dependent or the independent

variable, or both, be specified in first differences

• The first difference of a variable is its change in

value from one period to the next l• An example of a first difference model is iskart

| - Yi= B1 + B2(XI-XI-1) + 11 (1=2,...,T), or

at - Yo=B1 + BAXif u(1=2,...,T)

• 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

on you anyx

diffenence between this last t to cornest

Lose I obsene create at

-ta-ti es ist the Den x is important comment s

kenels

a dinotel

*lev*et

a

"Nonul

2...,T)

B\*t

tant use both get Bot B2 (xx-xx-x) \*BoXt

{B2 +133) X2 - B2 X=-1

Bu

a

can

forcest?

Examples of First Difference Models

• Some capital investments, for example, could be affected by changes in long-term interest

rates more than by the rate levels themselves

• 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

address a non-stationarity problem 1. Notice that one observation is always "lost" |

when using first difference models

Gy issue of possibilities

deforestation fory forest over timest

OR

get a forest

coner

ا ل املا الا

perfect me

net multilater

-

emem

Yr&+ Yrti

changes is o.1st dit on y il! condton of some

vara non statonary var that follow stuf process

avary can be por unstable over time Hanetouse

---- ist dof

op Reg. Linear en parameters before ols

But model non in enexp varlordep

nofwell de bor

Alternative Model Specifications r. The regression models studied so far assume that

the relationships between Y and the independent

variables (X2, X3,...,Xk) are linear Limest*e*l

• However, in many cases theory or experience

indicates some of the individual Y-Xj

relationships are likely non-linear

• When there is marked non-linearity in a given Y

Xj relationship, it is not appropriate to assume a

linear relationship between Y-Xj

• Fortunately, some non-linear Y-Xj relations can

also be estimated using the OLS method

muse

LARE X&Y linearly related?"

[Dift won hear models

AKA

Log-log or double Log ignore resid@ moment

The Log-Linear Specification

• A special type of non-linear relations become linear

when they are transformed with logarithms

• Specifically, consider îi = 3!XB2XB3..xßk Note that e=2.7182... is the anti-natural logarithm of| 1 (In(e)=In(2.7182...)=1), therefore eB1 is simply a multiplicative constant For example, if B1=4, B2=0.5, B3=1.5 and X3 =10, the model is Yi=e\* Xiż 10 %. = 1726.55 Xi2.

· This type of model can accommodate a variety of non-linear shapes, as depicted in the following figures

inor dearate

ene @ ever inte vs dur@ enarran

Log-Linear Model Specification

20000

1600

2-os a

12000

8

B2=15

000 +

1

*2*

3

4

5

6

*7*

8

9

10

*X*2

Other Non-Linear Relations that can be Modeled with the

Log-Linear Specification

decrease @ \*\*No dear

decrease inte

@ inor

60.00

50.00

40.00

B2 -0.5

>-30.00

1

2

3

4

5

6

*7*

8

9

10

The Log-Linear Specification

(-) > @trate

• In the Log-Linear model specification: - If Bi<0 as Xj increases decreases at a decreasing rate*?*

*(0*) - novelator

- If 0<B;<1 as Xj increases Y increases at a decreasing rate – If Bj=1 as X; increases û increases at a constant rate – If Bj>l as Xi increases Ỹ increases at an increasing rate

slope const reltux? but ax3

This is also known as the Log-Log or Double-Log

a slope

specification, because it becomes a linear relation when taking the natural logarithm of both sides: Y,= ex e ferror com is assurhed

>Ln(Yi)=Ln(eß1 )+Ln(Xi2 B2)+...+L11(Xik Bx)+Ln(e û ) \* = BiLn(e)+B2Ln(Xiz)+...+BkLn(Xik)+uLn(e) (note that

[Not tû cant est w/ oLs]

Ln(e)={ ifid, thus, the intercept above is Bi and the error term is úi

one case for ous w/ non llear

"Linde Ein(812)

parameters

meer nou have to take natural log of

M

un parremeters

( additne tence

EVERY THING

The Log-Linear Specification

• The former implies that standard linear

regression procedures can be used to estimate the coefficients of a Log-Linear model specification, but they are applied to Ln(Y) and Ln(X2), Ln(X,3), ..., Ln(Xık), instead of Y, and X,2, X,3,..., Xyk (i.e. let Y, = Ln(Y,) and X2=Ln(X;2), X,3=Ln(X,3),..., X,k=Ln(Xık) for all i's and then use the latter as your dependent and independent variables

. use

same forme, linear or Non

linear

The Log-Linear Specification

• A disadvantage of the log-linear specification is

that, since one needs to take the Ln of Y, all of the

LY-Xj relations in the model have to be assumed to

| be non-linear

:. Another 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 Bj directly measures

the elasticity of Y with respect to Xj; i.e. the percentage change in Y when Xj changes by one percent (show)

[dont advise but can have as I eno is undfre add small #to everythuy

insteadlad Hoc not a pealing

punimiters

elasticities

• Also note in this model specification the slope (i.e. the

unit change in Y when Xj changes by one unit) is not constant (it varies for different values of Xj), but the elasticity is constant throughout!

Varying Slope and Constant Elasticity of

Log-Linear Model Specifications

40000 350 00 30000

G1-4. B2=175 *25*000 Elasticity is Constant and Equals 1.75%

> 20000

15000

10000 5000

"Increasing Slope *7* 8 9 10

1

2

3

4

5 X 6

The Log-Linear Specification

• Final notes on the Log-Linear specification: - Since the dependent variable is, Ln(Ythe RP of

the estimated log-linear model is not comparable to that of the lincar 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 original cquation, not directly from the estimated model

usually log or linear

•R? not same; bounded b/wot theasures olovar on any that is exprained ny model 2thys use est purum to predict us compute rho R2 Look of

b/ wotl

compersible to linear model

shechore

ster? depon

jest and corre

et bwl i/yi

eng

data

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 (Bj's) measure the relative (i.e. proportional) change in

Y when the corresponding X (Xi) changes by an

absolute (i.e. one unit) amount

• That is if multiplied by 100. they measure the

percentage change in Y when the corresponding

X (Xj) changes by one unit

• As X increases Y may increase at an increasing

rate (Bj>0) or decrease at a decreasing rate (Bj<0) (see next two graphs)

other use elut of given predictions gives y \. take loyy not logx?

enlŷi) = ßit Bzxiz 2

proportunala to unito » purum est are small %DY when xol unit

\*ef x 100 currature is deffenent

*100*

actualley

proz

pest

2

vizelé, têx Mix)

Yi = e(t

= ene (%i) - B1 + Bar

Graph of Log-Lin Model Graph of Y=exp (1+0.5X) (Model is LA1Y)=1+0.5X)

5000

4000

3009

→ also how compute R2

pos? → n @ a Reete

no q @ t Rate CUMIT) ty cant be neg x can beneg)

*10*00

1

2

3

4

5

6

7

§

9

10 11 12 13 14 15

or a

Graph of Log-Lin Model

Dear@dear eate

Graph of Y = e xp 15.0.5X) (Model is In (Y)=5.0.5X1

1

2

3

4

5

6

7

8

9

10

11 12 13 14 15

Semi-Log Models: The Lin-Log Model

• 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 (Bj’s) measure the absolute (i.e. unit) change in Y when the corresponding X (Xj) 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 (Xj)

changes by one percent

• As X increases Y may increase at a decreasing

rate (Bj>0) or decrease at a decreasing rate (Bj<0) (see next two graphs)

can do op logy & normal,

So unit ay for %ry I uerge Bs 1100 get unitay when xD1%

Graph of Lin-Log Model

a@t rete

Å so

Graph of Y=1+1.51n (X)

\*no r@ prate

your beneg xcunt beney

1

*2*

3

4

5

6

*7*

8

9

10

11 12 13 14 15

Graph of Lin-Log Model

Graph of Y: 5.1.510 (X)

@ & rete na -B @ to rate y can be ney scant beneg

No & Careete

I

?

6

*7*

8

9

10 11 12 13 14 15

The Reciprocal Specification

• The reciprocal model specification is:

- Y = B + B2(1/X72) + BiXi3 ...+ BkXik

Reciprocal Model Specification

No natt log but do Ž

• Ê pos de beate but lowest

oval can take esh

as an BX bapprochentercept

• § as xaya @to nete intercept is max value

Yi24+2,5(1*/*Xi2)

Yi24:25(1/XL)

Note:4 - B1+ BoXfact

1 2 3 4

+BeXi

5 6

*7*

8

9

10

slope és op. sign B

g*t*

can have reciprocal trea xs

can have off multiccelin

The Reciprocal Specification

• A reciprocal model specification, for example, fits

Y-Xj relations that look like in the previous graphi

• As Xj (X2 in the graph) increases, Y increases or

decreases, but always at a decreasing rate

• In all cases, as Xj gets large Y approaches a limit

value (which equals 4 in the graphed example)

• The slope of a reciprocal model specification is: - dÝi/dX, = -Bi(1/X,%), which can be positive or negative

depending on the sign of B; - Unlike the linear model specification, the slope is

different depending on the value of Xj

A Note on OLS Estimation

• Note that, as in the case of the Log-Linear model,

the Log-Lin, Lin-Log and Reciprocal models can be

estimated by standard OLS procedures

• In the case of the Log-Lin model one has to take the natural logarithm of the Yi data before applying

OLS (i.e. use Ln(Yi) instead of the Yi data)

• In the case of the Lin-Log model one has to take the natural logarithm of the Xij data before applying

OLS (i.e. use Ln(Xij) instead of the Xij data)

• In the case of the Reciprocal model one has to use

(1*/*Xij) as the explanatory variable(s) (i.e. (1*/*Xij) instead of the Xij data)

Graph 1 of Log-Reciprocal Model

Graph of Y :exp{5.101/X }} {Hodelis 101Y)=5.10/1/X])

eu dia B B 2

gets sheepe a @athen signifpoint

get a @urate

B.(-)

1

2

§

10

12 14

16

18

*20*

*22 24*

*25*

*26*

Graph 2 of Log-Reciprocal Model

parem pos → ab@brate

Graph of Yo e 10(5+2 (1X }) { Hodelis ln (Y)=5+2(1/X }}

1

*2*

*·*

6

10 12 14 18 19 20 *22 24 2*6 28

The Polynomial Specification

• A polynomial model specification (with respect to X2 only) is:

• Y = B + B21X2 + B^2X2^ + B3X13 + ...+ BkXk

Polynomial Specification

combopelynom Squanes keep linear ornet leneur & D = zulo polynom xis pos or/-)

1

2

3

4

5 6 1. X2

*7*

8

9

10

11

Polynomial Specification

0

1

2

3

4

5

6 X2

*7*

8

9

10

11

• In a polynomial specification, as Xj increases, Y

can increase or decrease at an increasing or at a decreasing rate: it is a very flexible non-linear model specification However, one needs to be aware of the fact that might eventually imply a reversal in the direction of

the relationship between Y and X; (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

and the reciprocal specifications

use for test af nonlinearty 7 = B1 + B2 X2 u istik 1 Box

test t test toseaef B3 =B4 - 0

The Polynomial Specification

• Another advantage of the polynomial specification

is that it can be easily used to test non-linearity in the relationship between Y and any given

independent variable Xj

• If the parameter corresponding to Xj2 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 create a multicollinearity problem

can do any exponent \*"

The Polynomial Specification

• A polynomial model can be estimated by OLS,

viewing Xj' as any other independent variable in

the multiple regression

• In the example before where a polynomial

specification with respect to X2 is desired both X2 and X2' would be included as independent

variables in the data set used for OLS estimation

• Polynomial models with respect to more than one of the independent variables can be similarly estimated by OLS

12

Thurs

econom thery or undely Knoway

thely

Choice of Functional Forin

• The underlying theory and/or practical knowledge

of the problemi at hand can often help decide on

the choice(s) of functional form

• Other aspects to consider include: >The signs of the parameterestimates being consistent

with theory and/or practical knowledge of the problem → The level of statistical significance of the parameters in

the alternative models (i.e. functional formis) > The RP or, better yet, the adjusted R?jas long as the

dependent variable is the sanie across model choices

• Some use a residual analysis technique to ensure

selection of an appropriate functional form (do)

[Right signs]

nyher stut signif the better diftit parum Adj12? lay cant use R/Adjizz

need now rz

to make

tune

test to make sure it forma and you don't exclude and they mayo

Thurs Sept 29

AAEC 6610 Quantitative Techniques in Agricultural Economics

Chapter 8 F Tests

F-Test

• 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 not explained by

the model

• The greater the RSS relative to the ESS, thes,

larger the proportion of the variation in Y not explained by the model

Review: TSS: E(y-7)

ESS: £Ï-

RSS: E(Y*,-*4*)* Sgreater Rss than ESS Larger (proporton of very net explained

F Test

• Also recall that

►TSS = RSS + ESS, where > The TSS has n-1 degrees of freedom

► The RSS has n-k degrees of freedom > The ESS has k-1 degrees of freedom Notice that the degrees of freedom (DF) are Ladditive, i.e: DF(TSS) = DF(ESS) + DF(RSS)

Whats the deal? - probability

theory need to remember of somethy to do w/ #of indiv random var envolve*d* in *c*alluketon

niyis → one y → n-1 forts 4 – 4 + 4-1 Cr 6S nyi-kili →n-k for eso

(n-1(K-1) (htrsni

#Bused to call

F Test

• The F-test is used to evaluate the null hypothesis that all of the model's parameters, with thesexception of the intercept are zero:

Ho: B2=B3=... =Bk=0 \*\*

• Ho means that none of the independent

variables included in the model can be

statistically shown to have an effect on Y

• Thus, if Ho is not rejected, the model as a

whole can not be said to be useful to explain Y

• The alternative hypothesis Ha is that at least\*\*ly

one of the independent variables included in the model affects Y (back to residual analysis)

U want not to rejecto

ble can't explein o

can jointly have effect

rout not endin

multical r5e, hard to (w/srcell)

to testfir signif INDIV

• var that have effecton y out

dont know which ones.

cone (back multiccelin

etnotes

\* small sampus

no x

• The F\* statistic is calculated as:

\*F\* = (ESS/k-1)/(RSS/n-k)]

• Under Ho, if all four major OLS assumptions are

meet and ui is normally distributed, F\* follows an F distribution with (k-1) and (n-k) degrees of

freedom

• 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 Ho

equirofpval

"Signiff"

Other Types of F Tests

3other types

• There are other types of F tests that are

useful in multiple regression analysis; specifically:

> Joint tests on several regression coefficients > Tests involving linear functions of the

regression coefficients > Tests involving the equality of coefficients of

different regression equali0115

va

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

unrestricted (UR) model, since no assumptions have been made about any of its coefficients:

► Y = B1+B2X2+B3X3+...+BkXx+u

• Suppose that we want to test if a subset of q.

(q<k) of the regression coefficients is jointly equal to zero

detyong of var sachariete

large #obsere & karyetteyp var to catgory of var locaton, demogup

Socialchar

*5*, ҳ+1. —

look@sane stuff not signit

INDIV w/ ttest

UR = original model

**NY**

Joint Tests on Several Coefficients

• To do so, lets rewrite the unrestricted model as:

> Y = B1+B2X2+...+Bk-qXk-a-Bk-q+1Xk-q+1+...+BkXk+u

• If the last q coefficients are all equal to zero, the correct model will be the restricted (R) model:

2Y = Bi+B2X2+...+Bt-qXk-qếu

• The null hypothesis is:

> Ho: Bk+9+=Bk+9+2=... =B2=0

• The alternative hypothesis is that at least one of

those coefficients does not equal zero

ktutel; ges #parum entest – #afrestrictions ; have *q*eless

pariters

always Ho forf.

Joint Tests on Several Coefficients

fui Rest 2. Vi unrest]

ef UIR=Ulunr all parameters

aneo

• To test this Ho we need to estimate both the

unrestricted (UR) and the restricted (R) models

• Notice that the RSS of the restricted model

(RSSR) must be equal to or greater than than

the RSS of the unrestricted model (RSSUR)

• If Ho is correct, dropping the q variables

would have little effect on the explanatory power of the model and the RSSR should be only slightly higher than the RSSUR

Joint Tests on Several Coefficients

= # of Restrictions

ef RSSR a alot then rel to Rssuria pass es large test in a iness is signifigant

• The test statistic for Ho: Bk-9+1=Bk-gwit... =Bk=0,

is (then note what happens when q=k-1);

| »F\* = {(RSSR-RSSUR)/q}/{(RSSUR)/(n-k)}

• If Ho 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

nun

RSSI (0-(K-q)) RSBUR (n-k) = 0-$+q=1 tt

**10**

Tataum=Q Mechanics are the

same

n

ttest not signif Ftest & tfest multical inase at

not so enftest

Joint Tests on Several Coefficients

• 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-test result insignificant while the joint F test allows for the rejection of Ho (i.e. it results significant) - F-test is impervious to multicollinearity

testes / useful /

Octobery

لا ک

ے

x

قنع

| F Tests Involving Linear Functions of the

Regression Coefficients

• Recall that in a Cobb-Douglas production function

the sum of the (exponential) 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 a 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)

( 3 + B = ( с иим сұTS (CP) ness eu ý, = BitBq luxz \* B3 enxz

Restricton: B2 +B3 =12 HOOR B2=1-B3

impose

en ý = 8 + 1 - B en x2 + B3 en xz saneas,

en fra Bitluxiz. B.3 luxe & ensis restricti engineuxiz= Bi + Bgfentis lu xiz).

[compute be by difference]

surve es Br

instead

Issr > Ressur

to same as saying a pummane =

• Another example of an F-test involving linear

functions of the model coefficients involves the restriction that B2-B3=0ki.e., that B2=B3L

• The unrestricted and restricted models would be:

►Y=B1+B2X2+B3X3+...+BkXh+u (UR) vs.

Sunthenon

► Y = B:+B2(X2+X3)+...+BkXk+u (R)

• Alternatively, consider:

> Y=B1+B2X2+B3X3+...+BxXx+u (UR) VS. >Y=B1+B2(X2-X3)+...+BXn+u (R)

which implies that B2+B3=0 (i.e. that B:=-B2N

• Make sure that you know how estimate all these

restricted models!

az

attention

Pay

1.

each unter an euch linear function is one restriction only use one

parum

• The F\* statistic in all of the previous cases is the

same as before:

F\* = {(RSSR-RSSUR)/q}/{(RSSur)/(n-k)} Where q is the number of restrictions, i.e., the difference

in the number of parameters between the two models

• The concepts can be easily extended to

restrictions involving linear functions of more than two coefficients (B2 B3 B4=R; such as the

Cobb-Douglas example discussed before)

• Notice, however, that each linear function implies

one restriction only, regardless of the number of coefficients involved in the function

F Tests for the Equality of Coefficients of

Different Regression Equations Sometimes one wants to test if the same model applies to two somewhat different data sets or if two separate models are nceded (one for each)

• For this (Chow) test, consider these two models:

> Yi = B1+B2Xi2+B3X13+...+BkXik+ui (i=1,...,n)

► Yj = Qi+azXj2+a3X13+...+axXjk+uj (j=1,...,m)) These two models might have a different number of observations, their dependent variables have to be different but closely related, and their independent variables must be the same, but not necessarily take the same values (examples) -

two sep model or can use the same tan hame 2 models

depvars should be closey relt

indep have to betesane ex: data overtime into a

tins series + parinetes shift over time, by a ut? dif furms use sance product function

• Here Ho 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

• Ha is that at least one pair of coefficients are not

equal

• The unrestricted model consists of the two

separate equations, with RSSURERSSI+RSS2 (the sum of the RSS of the two separately estimated

equations)

• The degrees of freedom of RSSUR is the sum of

the degrees of freedom of RSS1 and RSS2, i.e. (n-k)+(m-k) = min.2k monozk

ys is also when before

- we sprit data

**IM*M***

• Notice that the null hypothesis is Ho: Bıral,

B2=(2, ..., Bk=Qk; which means that the two models can be written as a single equation:

► Ys = Bi+B2Xs2+B3X33+...+BkXsk+us (s=1,...,n+m) g where the subscript for the observations is changed

to s to denote the stacking of the n+m observations

• In our F testing framework the RSS of this

"merged” model is considered the RSSR

• The previously advanced F\* statistic is also

applied for this test with the appropriate degrees of freedom (k and n+m-2k):

• f\*= {(RSSR-RSSUR)/k}/{(RSSUR)/(n+m-2k)} in

Rest

difb/w z isk

nom-2k vs nem-k

• Notice that the degrees of freedom (df) for the

numerator equals df(RSSR)-df(RSSUR) = (n+m-k) - {(n-k)+(m-k)} = n+m-k-n-m+2k = k

= the number of restrictions imposed

• If Ho is rejected the data can not be pooled

and two separate regressions must be

estimated

• Note that, with the use of dummy variables, one could still estimate a single models where some of the coefficients are not equal pair wise

Reject nulla use sep. -

Want to not reject

intermed step allow somes to diftsome not to use dummy var.

다

니

AAEC 6610 Quantitative Techniques in Agricultural Economics

Chapter 9 Dummy Variable Regression Models

Use of Dummy Variables

• In many models, one or more of the independent

variables is qualitative (i.e., categorical) in nature

• This means that they can only take on few values

*often* with no naturai/logical ordering

• Examples include:

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

Use of Dummy Variables

Solif religion? need & dum var

Sdif column ;or 4 seasons

• 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 inodel, 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?

negri

للعان دہ :241)

lozi=1-D22

for all obsure

--/

. For example let Yi be cotton yields in farm i, X2i

indicates whether yields are from the current or the previous year (two categories), X3i indicate which of three varieties was planted (three categories), and Xdig...,Xki be other variables that affect farm yields Two dummy variables will be created for X2 (D21 and D22) and three for X3 (D31, D32 and D33) In the jih observation, D21=1 if yields were from this year and 0 otherwise, D22=1 if yields were from last year and 0 otherwise; D31=1 if the 1st variety was planted and 1 otherwise, D32=1 if the 2nd variety was planted and 0 otherwise, and D33=1 if the 3rd variety was planted and 0 otherwise (see dataset) i

osuther

1 verty i of other

• verta Dazon other

veurs

another

| 03121-D 32-833

J

have to dop one cat

Use of Dummy Variables

1. The estimated model would be:

Yi = Bi+B2D21i+B3D31i+B4D32i+ B5X4i+B6Xsi Notice that the dummy variables corresponding to one of the categories of X2 and X3 (D22 and D33)\* have been excluded from the model to avoid perfect multicollinearity (any one dummy/category can be excluded, 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 (X4 and X5)

Why

compost entercept [still study this]

• For this year's yields and the {st variety:

Yi = (B1+B2+B3)+BsX4i+B6Xsi . For this year's yields and the 2nd variety:

Yi = (B1+B2+B4)+B5X-si+B6Xsi

• For this year's yields and the 3rd variety:

Yi =(B1+B2)+BsXsi+B6Xsi

• For last year's yields and the 1st variety:

Yi = (B1+B3)+BsXsi+B6Xsi

• For last year's yields and the 2nd variety:

Yi = (B1+B4)+BsX1+B6Xsi

• For last year's yields and the 3rd variety:

Yi = B1+BsX4i+B6Xsi

Beis ist for diff blw

Yellds blwyty B3 def b/wverit thert I dites yesles *b*lusvent 143

subt B3-BI

enterc*e*pt,

For excluded

cutia/

1.

out

ce

Use of Dummy Variables

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

• B2 measures the difference in yields (for any variety

and level of X+,...,Xk) between this and last year

• B3 measures the difference in yields (for either this or

last year and for any level of X4.... ,Xk) between

varieties one and three

• B4 measures the difference in yields (for either this or

last year and for any level of X4,...,Xk) between

varieties two and three

• B3-B4 measures the difference in yields (for either this

or last year and for any level of X4,...,Xk) between varieties one and two

Therefore, notice that (see example):

• If B2 is not statistically different from zero (according to

the t test), cotton yields are predicted to be the same

during this and last year

• If B3 is not statistically different from zero (according to

tlie t test), cotton yields are predicted to be the sanc

under variety one and variety three

• If B+ is not statistically different from zero (according to

the t test), cotton yields are predicted to be the same

under variety two and varicty three

• To test whether cotton yields can be declared statistically

different under varieties one and two, one needs to estimate a different (but equivalent) version of the model excluding cither D31 or D32 instead of D33 or use F test

Use of Dummy Variables

• 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:

assume that yrt

verlety dont altect

responsme ness toxulo

This year

Bi+B2

Last year

X+ (or X5),

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” For example, consider the following model: Yo = Bi+B2D21i+B3D311X4+B4D321X4+B5X42+...+BXk+€.

sterept shut

• Here, the year effect is accounted for by an intercept

shifter, as before, but the variety effect is considered through an X4 slope shifter (take derivative) Bs is the X4 slope (it measures the change in yields when X4 -fertilizer use-changes by one unit) when variety three is planted (why*?)*

rallon slepe vel to some var

too enterruction b/w verriety 2+ phos agorication

• B4+Bs is the X4 slope when variety two is planted

B3+Bs is the X4 slope when variety one is planted

• Notice that the same categorical variable can be

used as both a slope and an intercept shifter

This year, Variety 2

Slope = B4+B5

BI+B2

Last year, Variety 3

Slope = B5

Yield Response to Fertilizer Use

X4 (X5 constant)".

Izv2

B3 B5

YZ,u3

B5

V,?

slope

th

October le cont.

AAEC 6610 Quantitative Techniques in Agricultural Economics

Chapter 10 Multicollinearity

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

• Intuitively, 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

ne

aldt afeconomic var

Related

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, everything

else being constant, the standard error associated to a given OLS parameter estimate will be higher if the corresponding independent variable is more highly correlated to the other independent variables in the model

ch other

a serfect? 21, su devide by sinot

possible

R

(x'x)

f

clue

ore

>zvar the r2 is Regressions

A on all of other var

model

Multicollinearity From the formulas for computing the standard error estimates in the two independent variable case: > V[B2] =*@*?l{(1-7\*.)(x2-x2)?? zvav*e* >VIB:]=32/{(1-)\_X3i-X) where r23 is the correlation coefficient between X2

and X3, note that: > The larger 121, the larger the standard errors > The larger the error term variance (02, i.e. the level of

unexplained variation), the larger the standard errors > The larger E(Xji-X;(i.e. the sample size or the amount

of variation in Xji), the smaller the standard errors

of \* 2 \* x3 highcor varf

var will double ej. 5

effect on se ist of effect

of varence undep raron xes needed]

gefrse ef

ger not

Multicollinearity

• This potential problem is known as multicollinearity

• It could be the reason why independent variables

that are believed to be key in determining the value of the dependent variable do not result statistically significant when conducting the basic t-tests It is not a mistake in the model specification or a violation of an OLS assumption, but rather an

undesirable characteristic of the data

• It is more common in time-series data models because time often affects the value taken by many of the independent variables in these type of models causing them to be highly correlated to each other?

3k y you think л цируе bout not if emport var seems not stut signit? check multical

not amistake/vwlaton on any as assump Most models have

some

perfect

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 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 including total annual rainfall as well as rainfall during each of the seasons as independent variables in a

time series model - When an independent variable takes a constant value in

all observations

x2 = d, +22 azt... d ka lok

somof a=0 but one of theses

non o and get perfect mun -f know val of one know other

if all petra valves 14

Perfect Multicollinearity

es a mistake, thcluded

*re*dundant info

• If there is perfect multicollinearity, the OLS method can not produce parameter or standard

error estimates

• 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 information)

Severe Multicollinearity

• A certain degree of correlation between the

independent variables is normal and expected in most cases

• 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

es se low & all est are signit; dont have issue multi ael will lead to besser but happy w/ what got?

w/ small # observ.

Symptoms of Multicollinearity

• The following, when taken together, are

considered symptoms of a multicollinearity problem: – Independent variable(s) considered critical in

explaining the model's dependent variable are not

statistically significant according to the t tests - High R?, highly significant F test, but few or no

statistically significant i tests connetines – Parameter estimates drastically change values and*/*

or become statistically significant when excluding some independent variables from the regression

take var out to ther parameter

o alot & sed alot

could have 2

out of 5 that

are issues

affeet thop those

Parametes

, simptomus?

VIF

Detecting Multicollinearity

• A simple test for multicollinearity is to conduct

“artificial” regressions between each independent variable (as the “dependent” variable) and the

remaining independent variables

· Variance Inflation Factors (VIF;) are calculated as:

• VIF; = 1*/*(1-R;?) where Rj? is the R’ of the artificial

regression with the jth independent variable as a "dependent” variable; which range from 1 to oo

• VIF; calculates the “inflation" on V[Bj] (i.e. (S[B;]))

caused by multicollinearity: If there is no multicollinearity, all VIF's equal one (i.e. there is no variance inflation due to multicollinearity).

one for each expranvar.

VIF x2 = //1-R2 <2 vs all else) est regression w/ xq as dept all other explan var = expran

- VIF=2 vous corres 52 is twice

what wouldhy woens what percento az is alreay accounted

for in their - / -rz factor on var est SEE es of that factor Run multipul checks but ef

van es sume

Detecting Multicollinearity VIF; = 2, for example, means that V[Bj] is twice what it would be if X; was not at all affected by multicollinearity

• A VIF;> 10 is clear evidence that the estimation of

B; is being affected by multicollinearity l• A second way of detecting multicollinearity is by 1 condition number of the X matrix M. As the independent variables (i.e., the columns of X)

M are increasingly correlated to each other the ratio of

the highest to the lowest eigenvalue of X’X becomes larger

3x wheet world be

who

tes expected

In when the sume

otuns condition effects

How competect any furch seint

matrix will have 19 an value

[k colun kiganvald

the more spreced the igan.

" more correlated

Inignest flowestigan

gives onerall read then condition is not

immune to scale a

Detecting Multicollinearity

• The condition number is the square root of this

ratio | • A (maximum) condition number of more than sob 100 is considered evidence of a multicollinearity

problem in the independent variable matrix

• Discuss advantage and disadvantage of CN

• Other methods available to assess the degree of

multicollinearity in a model include Theil's best multicollinearity effect and the determinant of the

correlation matrix of the independent variables (1 if no correlation, zero if all are perfect) 12

**BAD**

cho

**(iwan*a*nt)**

olowest I can dannel /

Roter*dam*m

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) Exclude 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 secondary var

Even if this is not the case, recall thať omitting Arelevant independent variables from the model will P make the OLS estimators biased

nud to be careful an dropiny

A can helep justify why you left var en model of evala en not stat signif? may be skewed can also double check before

getting rid of var

deaguastie

mone observ? use

sense that

brewlhal impuet

that har duesht

affect y real data adj for enf. imythy trendup: So when usenom have more multicul

*w*a

Addressing Multicollinearity

• Modifying the model specification can

sometimes help, for example: - Using real instead of nominal economic data – Using a reciprocal instead of a polynomial

specification on a given independent variable - When justified, using various non-linear functional

forms might reduce the level of overall linear

correlation across the X's

• 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 inferences are not disappointing

"

oct

Ad & type 2 error

& incorrect accept of

null

AAEC 6610 Quantitative Techniques in Agricultural Economics

Chapter 11

Heteroskedasticity and Generalized Least Squares

not lidt

Heteroskedasticity

• Heteroskedasticity occurs when the error term (and

thus the dependent variable Y) does not have a

constant variance across observations (see handout)

• The OLS parameter estimates are still unbiased, but

the OLS standard error estimates are biased (show and discuss correct way to compute OLS\_variance)

usualway

• This means that any statistical test or confidence

interval that uses the OLS standard error estimates

will be incorrect on the average

• Also, the OLS parameter estimates are no longer the

most efficient (i.e., minimum variance), even if the error term is normally distributed *)*

eust

rest

/

homusted t noautowri

use oslont

I

zvur [No longer Bi

diag es const

*c*omputers...

assoneelid /

**X EL**

Stand error est une blas *v*ar = E[Ė – *E(*2)} (@-E*c*s)]

\_\_ B\_ Find out ß-B = (\*'x)"X'u.

var Bc Effx'x7'x"UL'\*\*'\*)"}

BASE —(x'x)X'E [UD] xf'x)"!

tramomusteht nuauto wrr liel)

0%\*'x)

Un Blas

o Gauss

but not min need newst

Marc*ov*e other linear s that

ALWAYS Have white test est model

I better

dcomput resid E squased

Regression residlts are dep - ong IV (un) their squars tanxpoodsinden

semelar to funct form - careful w/ dum vur dont \* lum.

Need R2 on x2 w/ prum #df

entercept excluded dort . Ho: Homossed

key point what kind of a? really want . if to fail to reject incorrectly whates malt Protect agains thinkry free of Hetesk o wrory) consist wo

use highd 25% beritval

1

/ consist dont reject to

nove efficy

Tests for Heteroskedasticity | . The most common test for heteroskedasticity is

known as the White test, which is conducted by:

>Regressing the squared OLS residuals as the dependent variable versus the explanatory variables, their squares, and cross products as the independent variables This "auxiliary regression niust include an intercept and exclude any redundant right-hand-side variables > Under Ho: Homoskedasticity, n times the R? of that

regression is distributed as a chi-square random variable with p degrees of freedom (x()), where p is the number of parameters in the auxiliary regression minus one

\*t

int

lower

uyrunt Ho: Homosed

nr

to

1

Tests for Heteroskedasticity

As usual, Ho is tested by comparing the computed statistic (nR?) with the table value of x?cp) at the desired a level of significance (briefly discuss selection of a -

objective should be not to reject at 0.20 or so) >> Rejecting Ho, indicates likelihood of heteroskedasticity > It is important to note that, in the White test, misspecification of the random component (exclusion of

relevant explanatory variables or use of incorrect functional form) or correlation between the independent variables and the error term, could also cause rejection of Ho (although is should be noted that there are better tests

for these other conditions)

► Thus, it can only be used as a definite test for heteroskedasticity if the researcher is confident that the first two major OLS assumptions hold

Disadvantage can be other reasons why reject

nullon white test , viol Assom 1/2 trigger rejection

on null hypo w/o heterosk

White is openended

Tests for Heteroskedasticity

• The Breusch-Pagan is an older, Langrange

multiplier test >In this test, the *specific alternative hypothesis* of heteroskedasticity is Ha: 0%=h(27) (expand), where h is any possible nonlinear function of Zy, 2 is an (1xS) vector of "suspect” variables including an intercept, and y is an (Sxl) vector of parameters >Possible functional forms for o?=h(27) include:

0,2=(28)?, 0;2=27, 0,2=exp(zy), 0,2=In(2y), etc. >Since an intercept (Y,) is included in 2, the null hypothesis of homoskedasticity is equivalent to Ho: y\*=0, where y\* includes all other parameters in y

not accounted

Prefenince: dobooth tests Hai mor*e* restricted form of these some fl.) of uneart (exp var)

en intercep h = some informen, ? r., normal

of dit... fuas I all & except int are or cant do dia 82 Puzz + 83 /24 = fu zg?

[Restrictive alt]

can have weird form heterosted

I

fone

Tests for Heteroskedasticity

more compr to compute

est Regnession but dv) OLS Resid? & EZ, Rose ô2, RSS expliar ane un zi; justuse suspets in in form

null: Homoskel Aux Regression

- test stat is ESSAR/2

test stat x2 p-1 orsi df

Avanti al mone targeted to beterested (not as reusity thicked)

#suspects

If frow wiyutist */.*

nke

\* 1 • The Breusch-Pagan is an older, Langrange

multiplier test *>If the error term is normally distributed,* under Ho, BPTS=ESSAR/2 follows a x?cs-1), where ESSAR is the explained sum of squares of a regression with a?*:/*62 as the dependent variable (where o2=Eur/n) 6 and the suspect variables in z as the explanatory variables (list test steps) > As usual, Ho is tested by comparing the computed statistic with the table value of x'(S-1) at the desired *a* level of significance not exactly Discuss advantage and disadvantages of BP test

Disadvant: Ha more restrictived

requiner normality enerrors onlike white

oct is

use white test to see

if zal model is better

•param naved truevar

doesut & enterp ofcoef

var Ewi vi] = o2 for all i

heteroskedasticity, on can try to use different functions of

Vi = Bit B2x; + vi var[ui] = Zi (-2

Y: -F t Szo Wi Y;\* int\* xi\* var Cut] =0

Dealing with Heteroskedasticity

want

• If a single “suspect" variable can be identified, the

set wi such that

heteroskedasticity problem might be lessened by using a weighted least squares (WLS) estimator:

• In WLS, before applying OLS, all of the data is

nerror terms&n residuals

multiplied by a set of weights (W;), were the weights are a set of n values that when multiplied by the error term make it have a constant variance (i.e., V[w,u] =02)

• When there is only one variable (Z;) involved in the heteroskedasticity, on can try to use different functions of

it (such as w;= 2;, Z;?, Z:12*, 1/2*;*, 1/2, 1/*2*,*1/2, In(Z), etc.)

• The residuals of the WLS regression must be re-tested before one can consider the problem solved (see example) est for BitB2

- thansfur enerything by the wt.

dont affect mean? → dont

- can be exog var.

include R? not valid: Yes transform but Bane comperable

***n2***

Dealing with Heteroskedasticity

• If heteroskedasticity with respect to more than one variable is discovered one can use the

erorteruvar

predictions from an auxiliary regression to obtain estimates of the error term variances for

each of the observations *p*up

• Since the dependent variable in that regression is

the squared OLS residuals (u2), the predictions

from it are (proportional) estimates for the 0,2's

• Those predictions might need to be transformed - vale

in order to obtain the appropriate weights to correct for heteroskedasticity

s are your wts; proportonal

• Although acceptable, this is not the optimal method to correct for heteroskedasticity

Ûi- depend vara Zofols Ressel converse ya aslesed? en white rest

some can beney, the 2015 does it know sould be pos . i just add lovest (most negral)

will

to signa Square

do r

uxstils

correl orcovur

matra

prob.

ГР° */*

Generalized Least Squares (GLS)

• Generalized least squares is a general method to

estimate regression models when the error term

is not i.i.d., i.e, when Var[U]=o247021,

• It is based on the fact that if 4 is a positive

definite symmetric matrix (as it is the case for all proper covariance matrices), there exist

another (nxn) matrix P such that PYP’=In"

• Therefore, if Var[U]=o24, Var[PUZFPVar[U]P' =oPYP’=oʻl,

Nxnn

• Also note that Y=P-1(P').' and thus 4.1=P'P

iid error ols notiid GLS not iid g24 Pe correl matrix only appy if no heteroskl pos def nonzero pos detemunt

o cun tuke enverse

/

wt matrix

atheet how how

to just Th*a*t e*xi*sts]

transformaton

\*

P

\* Matrix

-

Matrix

varius P.' snan (non)

[nan]

(nans

Us the multipry

"

\*

Generalized Least Squares (GLS)

• The following model then has i.i.d. errors:

►PY=PXB+PU or Y\*=X\*B+U\* *09\**miid

• Thus, if OLS is used to estimate such transformed

model, it will produce efficient parameter

estimates and unbiased standard error estimates

• Application of OLS to this model yields the GLS estimator:

>B=(XʼP’PX)-'X'PPY=(X’Y-'X)"!X’Y-Y,

*\**>V[B]=o?(X'PPX)-1 = (X’Y•!X)1

• Note that the previously discussed WLS method

to correct for heteroskedasticity is in fact a particular application of the GLS estimator

multiply enythy by p=transformaton matrix - need on *c*o*v*arferr matrix

apply

als to transformed

model

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: \*>HCSE=(X'X)''X'SX(X'X)-' where 2 is a matrix with

the square of the OL'S residuals in the diagonal and

zeros elsewhere (where does it come from?)

• This yields consistent estimates for the standard

| errors, meaning that they can be considered

correct and confidently used for hypothesis testing (t-tests) and building confidence intervals if the sample size is relatively “large” (n at least >100)

izo

pes non 'q=(P.p)" error term covar matrix

l*ange* sample → not conc*e*rned about reftleny, mly issue is se care wrong

varf@]- (x \*JK' var (*1*1 ) \*4x'x}"

1 နှစ်

*G* =*<?(*x'xj efird asymtotny correct plug in 2 for

var j = Squar els resid endingouel

need est of ŷ then apply formulas -

- Not unbias; only

Consistant

Wis f(s)

boky w/ noo

ayheteroskal 'predef from white test 1. GLS no longer effics als se are incorrect

iday.

varst lansat

u=vor o

si= /varui

Probem al cross sectonal

deta

Key Points on Heteroskedasticity

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.

In the presence of heteroskedasticity the error terins corresponding to different observations have different variances. This difference in variance is usually associated with one or more of the explanatory variables in the model. enprar c*a*ratteet y cun alto altectva

CU)

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.

A higher error term variance means that the range of the observed errors is also higher. If, for example, the error term variance is o' = 16 (and the error term distribution is close to normal) about 95% of the errors would fall within +2xo of the error term mean, which is zero. So most of the errors would fall within 0+2x4 i.e. within -8 and 8. Alternatively, if the error term variance is oʻ = 4, most of the error terms would fall within 0+2x2 i.e. within -4 and 4.

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

OLS Residuals Plot under Heteroskedasticity

**OLS Residual Values**

10

20

30

40

50

60

70

80

90

100

110

120

Farm Size *(*Acres)

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:

Plot of Coffee Yields versus Farm Size

but oy

not suceptable

toscatt

will o relation

B/w try but not two

*d*atu pomts

**Coffee Yields (per acre)**

more ven soy sool type?

0

10

20

30

40

50 60 70 Farm Size (acres)

80

90

100

110

120

So, as farm size increases average coffee yields increase (the parameter estimate corresponding to 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.

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:

Reality Looks Plot of Coffee Yields versus Farm Size note epe

Coffee Yields (per acre)

0

10

20

30

40

50 60 70 Farm Size (acres)

80

90

100

110

120

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

In a model with more that 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 a particular 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 the plantation in the previous example).

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.

7 No pattern, Regardless 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.

pattern of Helterosk wf one vart

obsene

anorer

Oct is

AAEC 6610 Quantitative Techniques in Agricultural Economics

Chapter 12 Autocorrelation

Autocorrelation: Introduction

Jmosk

test

• Recall that when estimating the parameters of a

regression model using OLS, it is assumed that

the population error term is independently and

ft identically distributed

• An independently distributed error-term means

that when the observations are organized in a particular order, usually time (but it could be space), the sign and magnitude of the error term of any particular observation is not related to (i.e. it is independent of) the sign and size of the previous (and all other) error-terms

**M**

homeste of Autocorr - usually a timeseries issue

organize uvia time; ex eh ..., et } nothory to do w/ each other near time us are sometimes highly

can also hapen for location

Autocorrelation/Introduction

• Autocorrelation occurs when this error-term

independence assumption is violated

• It is most common in (but not unique to) time

series data, for example, commodity prices often exhibit random cyclical behavior over time, i.e. there are several periods*/*years of “good” prices

followed by periods of “depressed” prices

• This usually translates into an autocorrelated

error-term {u=put-1+vi, Vmiid N(0,0%)}

• Discuss Excel example of AR(1) process

Auto regressive I ARLIDE \*Ut=p Utoltve

fragen VENN(0,0%) iid stand nome

ist order autocome fes. so in this ex make up know cale

orintnis

example is

later in *C*lass

creates tendence that

affect corrent

prevernor

as long as the spel

process comes back to

( closser to 1= mor dumps

fis meg: get alternaton

percorrell/notot

Flgure 1: Observed and Expected Coffee Prices (1914-1996)

**Olseved Prices**

Price (1997 U.S.Si

**8 8 8 8 8 8 8 8 8 8 8 8 8 8 BB8 8 8 8 8 B8**

pv for comodlity ef error es î next fewabo. A viceversa

- postlumped together (Shocks can't

be obscorbed in one time periud)

Expected Pnces

**Ini**

1918

**1921**

**STAI**

SO

**Year**

Oct20

Autocorrelation/Introduction

• Derive variance and autocorrelation function for

an AR(1) process, 12 & 4 matrix, Gauss example

• Derive the variance and autocorrelation function

of a MA(1) process, see Excel & Gauss example

• Briefly discuss higher-order AR(p), MA(q) and

ARMA(p,q) processes, see Gauss example

• Discuss random walk, stationary issue (AR vs MA) and Dickey-Fuller and Phillips-Perron tests,

see Gauss example

• Structural vs time series (forecasting) models

• Discuss forecasting using Gauss example

Мати соис: умре Az-

tPCSS it- Putitvt utan 10,00) Elue] =*0*

ijd

var [ue]- t has to be >= + <1

\*\* var ve] -02 - \*\* coral cor fun, ver] - for *161-80%)* \* correl (ut, ut-1] =p

Stop

EXAM

Consequences

Assumpton 3+4

Start NOVI

-

use GLS not ors

• 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) is the minimum-variance estimator 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

Autocorrelation/Diagnostics

• A visual inspection of a plot of the OLS

residuals versus time can be used as a

preliminary diagnostic tool for autocorrelation

• Groups of several positive residuals followed by

several negative residuals indicate possible

positive autocorrelation, which is most common

• Residuals with clearly alternating signs indicate

negative autocorrelation, which is very rare

• The best know formal test for autocorrelation is

the Durbin-Watson statistical

to visual sign neqautocorre doesn't really happen

CO

Formal

test

Autocorrelation/Diagnostics The basic Durbin-Watson statistic tests for first order autocorrelation; i.e. correlation between the previous and the present period residual (i.e. dependent variable) values: It is calculated as:

d\* = Ś(ût-ût-)??ûcz

• If there is no first order autocorrelation the value of

d\* will be close to 2 (briefly discuss why) K

• It 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)

White test for autocont

assumestuctual time series compors stat

El Cur - *-)"| zsūc)?* No autocorv (isk orel*e*r) > call remerz

• More precisely, let Ho: no autocorrelation vs Ha:

positive autocorrelation (a one-tailed alternative)

• Compare the calculated d\* with the "significance

points” in the Durbin-Watson statistic Table at the desired &, for the appropriate sample size (n) and number of independent variables in the model (k)

If d\*>du conclude Ho (d\* <4-du)

• If d\* <dl conclude Ha (d\*>4-dl)

Otherwise the test is inconclusive

Do example of DW test in Excel

• A two-tailed alternative is also possible (discuss)

lag valy as explan

til Structual time series & add layy as

ded, d\* would \*\* x; 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:

• 11\* = {1-{d\*/2)} VT/{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

easiertest? don't need table

dun N(0, 1) one fail 2 test

164720% z

• Durbin has shown that under Ho: No

autocorrelation 1\* is approximately normally distributed with unit variance; thus the test can

be conducted using a standard normal table

• Notice that if the alternative hypothesis is

positive autocorrelation the level of

significance is *a/*2 \*

• A modified version of the h test has to be used In if T(V[B])}>1 (why*?*) will take sote

K. Model selection, Bartlett and Box-Pierce tests

K for higher-order autocorrelations, AIC and

V SBC, go back to Gauss example...ed

Rare

A

Adj R2B

3

stayo

want to min add perum &p=% oded

for #penem

Penelit

more

Perilized

can check of MA needed

Transform then use oLS

- evror es ARCA)

tincseries

*"..,7* Struit

Correcting for Autocorrelation

• The solution for autocorrelation is analogous to that

of heteroskedasticity: Transform 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:

► Yo=B1 + B2X2 + ... + BkXke + Ui, t = 1,..., TS >u: = put-1 + Vi, (05|p|<1) where both Ui and Vi have zero unconditional expected values and constant variances through time, ui is autocorrelated but Vi 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

assoney is known

• If p were known, the following procedure could be

used to correct for autocorrelation

• Notice that the former model implies:

BY1-1 = PB1+pB2X2-1+ ... +pBkXkt-1 + put-1; thus Yı-p Yt-1= (1-P)B1+ B2(X21-pX21-1) +...+ Bk(Xkt-pXkt-1) +(ut-put-1); or: D

Y\*, = (1-P)B1 + B2X2\*, + ... + BkXk\*+ VIUS*C*O*L*S >Notice that the error term (vi) 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

multitory byl -subtruct previous. f from next obs Yt- 14t enterceptes if not all xs are 1st diffish

1. t-pty

put= Pût-1 + Zz

EGLS Estimation

• This involves using an estimate of p (Ô) to

implement the previously discussed procedure

• The simplest estimate is the sample correlation coefficient between the error terms:

>ộ = £ û û//RS$ *residualo*

• This (approximate) GLS estimator is also know

as the Cochrane-Orcutt procedure

• A second alternative is to regress û, on û- Beppo

(without an intercept) and use the parameter estimate from that regression as an estimate for p

estimated Generilized least squares → 0/1 2 steps

we heightly

the

d#=0>pal d\*=27P=0 no autocort

EGLS Estimation 7

• Yet a third alternative is to use ô = 1 - 0.50\* whered\* 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 B > D -> p Compute residuals and obtain a first estimate for p

>Obtain parameter estimates using EGLS > Re-compute the residuals (with EGLS parameter estimates and raw data) and obtain a second estimate for p Continue process until the estimate for p stops changing

raw

zal round Gis

Exact generilized PV: X-PO

(fxn2x) *O*LS >*GL*S → Minimize *l*epus cpus),

EGLS Estimation

• However, it is not clear that iterating helps

obtain more efficient parameter estimates

• 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 {know as Non-Linear Least Squares (NLS)} that can only be solved by numerical search procedures

Z squared transforme

residuals

- Min Luip 'pu) – Min (U'*q\*u*). Not shot Min (Cy-xB) 4 (of xB) ]

all scalar)

( can lointly find Btl enone 1

shot

Sece Notes

NLS Estimation by Simple Search

• 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

• 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)

Horley

11/10*/*2016

Novio

numericaly orderd & continuous

AAEC 6610 Quantitative Techniques in Agricultural Economics

Qualitative Choice Models

is qualitative

resource econ

Introduction to Qualitative Choice Models

• Qualitative choice models are models in which the

dependent variable involves two or more qualitative

choices (as opposed to quantitative values)

• These models are most valuable in the analysis of survey and economic experiment data:

• Whcther or not a (potential) visitor is willing to pay a --

particular fee to enter a park NPS

• Technology adoption studies: whether or not farmer

decides to adopt a recommended production practice(s)

• Which of four different available types of irrigation

systems (tractors, etc) a farmer decides to use

• Which type of vehicle (car, minivan, truck, SUV or

domestic, Japanese, European) a consumer purchases

Bid is how much they will pay; respense

as yes or no

Binary-Choice Models

• Binary choice models are used when the

dependent variable can only take two mutually

exclusive values (i.e. is a dummy 0/1 variable)

• 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

• 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

yes or no; for B prob of choosing one or

other choke Bis a en probability

DILC

11/10/2016

Linear Regression I DONT assune: Normality (yes not normal)

o will be heteroskadestic

The Linear Probability model

• The linear probability model is the most

elementary binary-choice model:

• Y = XB+U; where X is an nxk matrix of

independent variables, which can be of any kind, Y is an nxl vector of qualitative dependent variable realizations (i.e. zero or one values), and U is assumed to be an nxl vector of independently

distributed random variables with zero mean

• The Linear Probability Model above is estimated by

OLS, which yields inefficient parameter estimates and biased standard crror estimates (brietly explain why, will specifically show later)

The Linear Probability Model (LPM)

• If we denote the probability that Yi takes a

value of 1 by Pi; then E[Yi]=1(Pi)+0(1-Pi)= Pi

• Also since E[ui]=0, E[Yi]=XiB=Pi and

Yi=XiB is thus the LPM estimate for Pi (Pi)

• Therefore, another disadvantage of the LPM model is that Yi = Pi = XiB could take values

greater than one or less than zero OLS Pousut

• To be able to interpret Ýi = Pi as the predicted

probability that Yi takes a value of 1 given Xi, the following convention has been adopted:

Burnooli ev is yi est Pr[y]=1 is Pi Prłyi]=0 is (1-pi) discreat Random var: posib.prob d addup. getneg prob used to >l> <oto

*dicte*r non

7

Those

prob est are bras.

So ECÂ) #Pi

Summary

• By convention, since Yi is a Bernoulli (0-1)

random variable with Pr[Yi]=1 being Pi and

Pr[Yi]=0 being (1-Pi), then E[Yi]=Pi

• Now, since in the LPM we define Yi=XiB+ui

where E[ui]=0, it follows that E[Yi]=XiB (in

that particular model)

• Thus, in the case of the LPM, Pi is defined to

be equal to XiB and Xiß = Pi is the LPM prediction for Pi

11*/*10*/*2016

*Din*

var is 25@.5 as appich of Sapphy

The Linear Probability Model

| Xiß when 0 < Xiß < 1.

Pi = 1

when XiB 2 1 lo when Xiß SO

• It can be shown that the error term of the LPM is

heteroskedastic: {012=E[üi?]=Pi(1-Pi)=XiB(1-XiB)}

• It has lower variances when Pi is close to 0 or 1 and

higher variances when Piis closer to 0.54

• What are the consequences of heteroskedasticity?

• A third weakness of the L.P.M is that the model's

predictions (as above) are biased

Rt model

> Pick

The Probit and Logit Models

• These models address the formerly discussed

drawbacks of the linear probability model and insure that the predicted probabilities are unbiased and lie

in the (0,1) interval B

• This requires that E[Yij is expressed as a non-linear

function of XiB, specifically a cumulative probability distribution function (CDF), because a CDF only takes values between 0 and I regardless of the value of its argument (XiB in this case):

• E[Yi] = Pi = CDF(XiB) (draw picture)

• Numerous alternative CDF's could be used to

complete the specification of the model above

VB

pies f

The Probit and Logit Models The probit model is obtained by assuming a standard

normal CDF, while the logit model is obtained by

assuming a logistic CDF

• More specifically, in the probit model:

• Pi = NCDF(XiB) = (1*/*127)) e 2?dz

where Z is a normally distributed random variable with

zero mean and unit variance (draw pictures)

• By construction, Pi will lie on the (0,1) interval

• As XiB increases, the probability that Yi takes a

value of 1 (Pi) also increases

Slide 7

OR1

Substitute for ui and recognize that E[Yi2]=Pi also. Octavio Ramirez, 1*1/1*8*/*2014