Introduction

I) Definition of Econometrics

The application of statistics and probability to the measurement, estimation, and prediction of economic relationships

II) Role of Econometrics

- a) measurement of economic relationships
  - 1) shape of the demand function
  - 2) shape of the supply function
  - 3) the marginal propensity to consume or invest
- b) prediction
  - 1) What will be the effect of raising the sales tax on sales of some item
  - 2) Under the given government budget what will be the GNP next year

III) Role of Economic Theory

1) suggests variables which are relevant
2) may but rarely does suggest a functional form of the relationship
3) provides the hypotheses to be tested

IV) Some Terminology

1) independent variables, regressors, explanatory variables, exogenous variables
2) dependent variables, endogenous variables, explained variable

V) Errors

A) In the economic relationship to be estimated

1) statistical vs deterministic relation

$Q = a - bP$
is a deterministic or functional relationship

$Q = a - bP + \mu$, where $\mu \sim (0, \sigma^2)$ is a statistical relationship.

Where does the μ come from?

eg. Suppose that we want to determine the relationship between price and quantity demanded, ie, to know the shape of the demand curve

Demand depends upon prices of all goods, income, and tastes

We obtain data on quantities purchased, prices and incomes by surveying consumers. But it is not easy to get information on tastes; the best we can do is find out about the consumer as much as we can. We ask age, sex, race, marital status, number of kids, education and other such information that may provide a clue about tastes. But we still do not know "tastes" - therefore there exists some portion of the demand which we can not explain; this is an error in measurement.

But also when we ask the consumer price, quantity, and income information there may be mistakes made. The consumer may not recall quantity very well or only give a round number for income when income is not a round number. All of this is error in our demand equation.

Sources of error

1) omitted variables - either because there is no data or because something is relevant but not specified by the theory, or just omitted
2) measurement error of the dependent variable

measurement error in the independent variables leads to complex problems which we will only lightly discuss late in the semester

3) randomness in behavior

B) In the Inferences Made

1) regression and causation

Causality must be argued by an appeal to some theoretical model - it cannot be concluded based upon a statistical relationship

ex. crop yield and rainfall

2) Regression and correlation

Correlation analysis: the relationship between two variables is being measured where each is assumed to be random

regression analysis: the relationship between variables is being measured but only the dependent variable (Q from above) is assumed to be random. The independent variables are assumed to be constant/fixed.

Ordinary Least Squares Regression

1) Model

a) the relationship is linear in the parameters or can be made so

ex. $q = a + bP + cI + \mu$

$q = a + b(1/P) + cIP + \mu$

not linear in Parameters

$q = a + abP + bP^2 + Iexp(c) + \mu$

b) the model is correctly specified

give example

c) all the independent variables cannot be the same

plot relationship for identical X for all cases

2) Errors

REMIND CLASS OF EXPECTATIONS NOTATION: $E(X)$, $E(X-\mu)^2$, $E(X|Y)$

Let the model be: $Y = \alpha + \beta X + \mu$

Digression about interpretation of α and β

a) mean zero, $E(\mu) = 0$
b) constant variance, $E(\mu^2) = \sigma^2$, homoskedasticity
c) no autocorrelation, $E(\mu_i\mu_j) = 0$, covariance is 0
d) uncorrelated with the independent variables, $E(\mu X)=0$

d) is automatically true if the independent variables are fixed/constant and a) holds.

point out that d means $E(\mu|X) = 0$

3) Regression model as a Conditional Expectation

$E(Y|X) = E(\alpha + \beta X + \mu) = \alpha + \beta X + E(\mu|X) = \alpha + \beta X$

or

$\mu = Y - E(Y|X) = Y - \alpha - \beta X$

4) Derivation of coefficient estimators

Define: An estimator is simply a rule or formula that tells how to estimate the parameter of interest.

Since α and β are parameters to be estimated the values that are obtained from any sample of data are only estimates of the true parameters. For the given data set we are getting estimates of the conditional mean of Y E(Y|X).

a) OLS means chose that set of parameters which minimizes the the sum of the squared deviations from the estimated conditional expectations

DIGRESSION ON SUMMATION NOTATION. SEE PAGES 623-624 IN TEXT.

Let $\epsilon_i = Y_i - a - bX_i$ where ϵ is the estimated error and i denotes the individual observation and Y and X are values found for that person/firm/observation i.

DRAW A DIAGRAM WHICH SHOWS THE RELATIONSHIP BEWTEEN THE ERROR AND THE TRUE LINE WITH THE ERRORS DISTRIBUTED ALONG THE LINE

EXPLAIN THE REASON FOR SQUARING THE DEVIATIONS

OLS says minimize by choosing a and b

$$\sum_{1}^{N} ,_i^2 = \sum_{1}^{N} (Y_i - a - bX_i)^2$$

First-order-conditions:

$$-2\sum_{1}^{N}(Y_i - a - bX_i)X_i = 0$$

or

$$\sum_{1}^{N} X_i e_i = 0$$

Note that this means that the estimated error and the independent variables are uncorrelated.

$$-\sum_{1}^{N}(Y_i - a - bX_i) = 0$$

These equations can be solved for the estimators a and b.

b) An alternative method

$Y_i = a + bX_i$

Sum this equation across i - this is the first order condition for a.

$\Sigma Y_i = Na + b\Sigma X_i$
Divide by N to get $\bar{y} = a + b\bar{x}$

Multiply the orignal model by X_i then sum across i - this is the first order condition for b

$\Sigma Y_i X_i = a\Sigma X_i + b\Sigma X_i^2$

solve for

$a = \bar{y} - b\bar{x}$

substitute for a and rearrange to get

$\Sigma Y_i X_i - \bar{y}\Sigma X_i = b(\Sigma X_i^2 - \bar{x}\Sigma X_i)$

$\Sigma(X_i(Y_i - \bar{y})) = b\Sigma(X_i(X_i - \bar{x}))$

$b = \Sigma(X_i(Y_i - \bar{y}))/\Sigma(X_i(X_i - \bar{x}))$

REMIND CLASS OF ASSUMPTION THAT ALL X'S NOT BE THE SAME AND POINT OUT THAT DENOMINATOR OF b WOULD BE ZERO IN THIS CASE.

5) IMPORTANT POINTS TO NOTE

a) point estimators

b) the estimators a and b depend only on the sample information

c) regression line passes through the point $(\bar{x}\bar{y})$

d) mean of estimated Y, $\hat{y}_m$ equals $\bar{y}$

e) mean of ϵ is zero

Our second method of deriving the least squares estimators relied upon this fact. $\Sigma Y_i = Na + b\Sigma X_i + \Sigma\epsilon_{i \text{ and } \Sigma\epsilon} i = 0$

$Y_i - \bar{y} = -a - b\bar{x} + a + bX_{i+\epsilon} i$
$y_i = bx_i + \epsilon_i$

and multiplying by x_i and summing $\Sigma y_i x_i = b\Sigma x_i^2 + \Sigma x_i \epsilon_i$
using uncorrelatedness of ϵ_i and X we get the estimator b

$b = \Sigma y_i x_i / \Sigma x_i^2$

Show that this is same as before. Point out $\Sigma x_i^2 = \Sigma X_i (X_i - \bar{x})$

Also, we can predict the deviation of Y from its mean

$\hat{y}_i = bx_i$

f) ϵ_i uncorrelated with predicted/fitted Y, $\hat{y}_i$

This result is important later when we discuss systems of equations and is easily demonstrated. using deviation form

$\Sigma \hat{y}_i \epsilon_i = b\Sigma \epsilon_i x_i$ but recall that $\Sigma \epsilon_i x_i = 0$ from uncorrelatedness of X and ϵ

Therefore: ϵ and $\hat{y}_i$ are uncorrelated.

6) precision of estimates

a) variance of the error term

We assumed that the error μ has constant variance σ^2

But this is not a known quantity - it too must be estimated

$\epsilon_{i=Y}i - a - bX_i$

We can estimate the σ^2 by using the estimated errors ϵ

$\hat{\sigma}^2 = (1/(N-2))\Sigma\epsilon_i^2$

where N-2 is known as the number of degrees of freedom and is the sample size minus the number of estimated parameters used in calculating the ϵ_i.

Note also that σ^2 is the variance of Y

b) unbiasedness and variance of b

recall that $b = \Sigma(X_i - \bar{x})(Y_i - \bar{y}))/\Sigma(X_i - \bar{x})(X_i - \bar{x})$

And let me write $(Y_i - \bar{y}) = y_{i,(Xi} - \bar{x}) = x_i$. Let $1/(\Sigma x_i^2) \equiv W$.

Now $b = W(x1y_1 + x_2y_2 + ... + x_Ny_N)$

$= W(x_1(\alpha + \beta X_1 + \mu_{1 - \alpha - \beta\bar{x}}) + x_2(\alpha + \beta X_2 + \mu_2 - \alpha - \beta\bar{x}) + ...)$

Unbiasedness of b:

b is unbiased if the $E(b) = \beta$

$E(b) = E[W(x_1(\beta X_1 + \mu_{1 - \beta\bar{x}}) + x_2(\beta X_2 + \mu_2 - \beta\bar{x}) + ...)$

Recalling that E(constant) = constant and collecting terms

$= W[\beta(x_1(X_1 - \bar{x}) + \beta x_2(X_2 - \bar{x}) + ... + \beta x_N(X_N - \bar{x})) + E(\Sigma x_i\mu_i)]$

$= \beta W\Sigma(x_i(X_i - \bar{x})) + W\Sigma x_iE(\mu_i)$
$= \beta$

Variance of b:

$E(b-\beta)^2 = E\{\beta + W\Sigma x_i\mu_i - \beta\}^2$

$= W^2E[(\Sigma x_i\mu_i)^2]$

$= W^2E[\Sigma x_i^2\mu_i^2 + \Sigma\Sigma_{i\neq j} x_ix_j\mu_i\mu_j]$

$= W^2\Sigma x_i^2E\mu_i^2 + W^2\Sigma\Sigma_{i\neq j} x_ix_jE\mu_i\mu_j$

$E(b-\beta)^2 = (W^2/W)\sigma^2$

$= W\sigma^2$

Since in general we do not know σ^2 we use its estimator to estimate the variance of b.

Note influences on variance of b

1) variance of ϵ
2) spread of independent variables - the more dispersion amongst your independent variables the more precision with which the slope is estimated

7) properties of the OLS estimators - BLUE

a) unbiased
b) linear - the dependent variable enters into formula in a linear way
c) smallest variance of all linear unbiased estimators

explain what this means

d) consistent - as the sample size gets very large the variance (and, in general, the bias if any exists) goes to zero

8) R squared - coefficient of determination

Want an idea of how well our model does in explaining the variation in the dependent variable

$y_i = \hat{y}_i + \epsilon_i$

now square the deviations and sum them

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

$= b^2 \Sigma x_i^2 + \Sigma \epsilon_i^2$

total sum of squares = explained sum of squares + residual sum of squares

TSS = ESS + RSS

or 1 = ESS/TSS + RSS/TSS

$r^2 = ESS/TSS = \Sigma \hat{y}_i^2 / \Sigma y_i^2$

Note that this quantity is :

a) non-negative
b) greater than or equal to zero but less than or equal to 1

In words, r^2 measures the proportion of the variation in the dependent variable that is explained by the model. Note that the r^2 will typically go down as the number of observations goes up - hence high values are not to be trusted when the sample size is small.

9) classical model - normality of μ

When μ is assumed to be normally distributed the assumptions of zero mean, constant variance, and zero covariance are more compactly written as $\mu \sim N(0,\sigma^2)$

a) Why Normality -

1) zero covariance and independence are identical
2) iid implies by central limit theorem that sum of variables approaches normal distribution as sample size gets large
3) in some conditions even without the large sample size or independence the sum of the variables is normally distributed
4) any linear function of normally distributed random variables is also normally distributed
5) normal distribution depends only on mean and variance

b) properties of estimators under normality

1) unbiased
2) minimum variance
3) consistent
4) estimates are normally distributed

Hence, $a \sim N(\alpha,\sigma_\alpha^2)$
$b \sim N(\beta,\sigma_\beta^2)$

5) $(N-2)\hat{\sigma}^2/\sigma^2 \sim$ chi squared with N-2 degrees of freedom
6) coefficient estimates are independent of variance estimate
7) coefficient estimates are have minimum variance of any unbiased estimator - not just linear estimators

c) Maximum likelihood estimation

Start with an assumption on the distribution of the error terms - not necessarily normal. Want to maximize the probability that the given data was generated by the assumed distribution

$Y_i = a + bX_i + \epsilon_i$

$\epsilon \sim N(\eta,\sigma^2)$

Let $f(\epsilon_i)$ be the value of the normal probability density function evaluated at ϵ_i. Or $f(Y_i - a - bX_i)$. Since the ϵ_i are independently distributed the product of the values of f evaluated at each ϵ is the likelihood that these data are generated by the assumed distribution.

$$L = \prod_{i=1}^{N} f(Y_i - a - bX_i)$$

Now maximize by choosing a and b. If f is the normal density function then a and b are the OLS estimators.

This approach is especially useful when the errors are not normally distributed.

Hypothesis testing and interval estimation

Recall that one of the goals of econometrics is to test economic relationships - for example is the price elasticity different from 0 or is the marginal propensity to consume .7

Such questions are hypotheses and can be tested. But to do so we must decide what values of the estimates of the price elasticity we will consider bigger than 0 or which values of the MPC are sufficiently close to .7 that we will consider them to be .7. This is the process of interval estimation.

ex. suppose that we estimate a demand equation and find b = - .35. Is this sufficiently far from zero to conclude that β is not zero?

First, from statistics you know that this can only be answered probabalistically -

$$Pr\{b - \delta \leq \beta \leq b + \delta\} = 1-\alpha$$

that is, the probability that the true parameter value lies in an interval calculated from the estimates equals $1-\alpha$. The interval is called a confidence interval and α is the significance level.

The question we are asking is, "Is the evidence we have compatible (sufficiently close) with our hypothesis?"

The stated hypothesis is known as the Null Hypothesis.

$H_0 : b = 1$ is how we would write the null hypothesis from above.
$H_1 : b \neq 1$ is the Alternative hypothesis

Given our null hypothesis we must form a decision rule - this rule leads us to the construction of confidence intervals or of acceptance regions.

a) Constructing confidence intervals - variance known

Recall that under the assumption of normally distributed errors that the coefficient estimates are also normally distributed

$$b \sim N(\beta, \sigma_\beta^2)$$

This means that we can make b into a standard normal variable by

$$Z = (b-\beta)/\sigma_\beta$$

We can write the probability that Z is between $-\delta$ and δ as

$$Pr(-n_{\alpha/2} \leq (b-\beta)/\sigma_b \leq n_{\alpha/2}) = 1-\alpha$$

where $n_{\alpha/2}$ is some value from the standard normal distribution table.

$$Pr(-n_{\alpha/2}\sigma_b \leq b-\beta \leq n_{\alpha/2}\sigma_b) = 1-\alpha$$

$$Pr(-b - n_{\alpha/2}\sigma_b \leq -\beta \leq -b + n_{\alpha/2}\sigma_b) = 1-\alpha$$

$Pr(b - n_{\alpha/2}\sigma_b \leq \beta \leq b + n_{\alpha/2}\sigma_b) = 1-\alpha$

Example: suppose that we think $\beta = 1$ and that $b = .87$ and $\sigma_b = .34$
and 16 degrees of freedom
construct the 95% confidence interval

$Pr(.87 - (1.96)*.34 \leq \beta \leq .87 + (1.96)*.34) = .95$

Note that $n_{\alpha/2} = 1.96$.

One could also construct an acceptance region.
$H_0 : \beta = k$, $H_1 : \beta \neq k$

$Pr(-n_{\alpha/2} \leq (b-k)/\sigma_b \leq n_{\alpha/2}) = 1-\alpha$

$Pr(-n_{\alpha/2}\sigma_b \leq b-k \leq n_{\alpha/2}\sigma_b) = 1-\alpha$

$Pr(k - n_{\alpha/2}\sigma_b \leq b \leq k + n_{\alpha/2}\sigma_b) = 1-\alpha$

$Pr(k - n_{\alpha/2}\sigma_b \leq b \leq k + n_{\alpha/2}\sigma_b) = 1-\alpha$

b) confidence interval for b - variance unkown

Since it is rare that we know the true variance of b we must use the estimate of that variance to construct the confidence interval. When doing so we replace $n_{\alpha/2}$ by $t_{\alpha/2}$ at the given degrees of freedom.

In the example above suppose that .34 is the estimated standard deviation of b.

$Pr(b - t_{\alpha/2}\hat{\sigma}_b \leq \beta \leq b + t_{\alpha/2}\hat{\sigma}_b) = 1-\alpha$

$Pr(.87 - (2.12)*.34 \leq \beta \leq .87 + (2.12)*.34) = .95$

Note that the interval is larger when we use the estimated variance rather than the true variance.

This makes sense - the estimated variance raises the uncertainty about the accuraccy of the estimate b so to be as confident about getting the true β in our interval we need a bigger interval.

Note that by using the acceptance region approach with the t value is also appropriate.

The acceptance region is constructed analogously.

c) hypothesis testing

1) two-tail tests

suppose that economic theory suggests that the effect of an increase in some variable induces an change in some other variable - for example, suppose our theory suggests that the price elasticity of demand is 1 and our b is an

estimate of the elasticity

Our null hypothesis would be then H_0: $\beta = 1$
H_1: $\beta \neq 1$

We choose a confidence level or a significance level and set up the corresponding confidence interval.

*****
Decision Rule: If the hypothesized value of β is in the confidence interval then accept the null; else reject the null hypothesis.

In our example above would we accept or reject the null hypothesis?

3) one tail tests

Often in economics theory suggests the direction of a relationship but not a magnitude - for example, the price elasticity of demand is negative, the income elasticity positive.

In this case we want to test H_0: $\beta \leq k$, H_1: $\beta > k$

What constitutes the decision rule? We will accept the null hypothesis if the estimate of β is sufficiently less than k (if the hypothesis is that $\beta \geq k$ we accept the null if the estimate is sufficiently greater than k).

$Pr((b-k)/\hat{\sigma}_b \leq -t_\alpha) = 1-\alpha$

$Pr(b \leq k - \hat{\sigma}_b t_\alpha) = 1-\alpha$

4) 2-t rule of thumb

5) Type I errors - probability of rejecting a true hypothesis

Type II error - probabilityof accepting a false hypothesis

Prediction/forecasting

Recall the true model is

$Y_0 = \alpha + \beta X_0 + \mu_0$

where μ_0 is not forecastable from any existing information
So we use
$\hat{Y}_0 = a + bX_0$
to forecast the mean of Y_0

$E(\hat{Y}_0) = E(a + bX_0) = E(a) + E(b)X_0 = \alpha + \beta X_0 = E(Y_0)$

Now since $a + bX_0$ is an estimator it is a random variable with mean $\alpha + \beta X_0$ and a variance

$$\text{var}(\hat{Y}_0) = E(a + bX_0 - \alpha - \beta X_0)^2$$

$$= E\{(a-\alpha) + (b-\beta)X_0\}^2$$

$$= E(a-\alpha)^2 + X_0^2 E(b-\beta)^2 + 2X_0 E(a-\alpha)(b-\beta)$$

$$F^2\left[\frac{\sum_{i=1}^{N} X_i^2}{N\sum_{i=1}^{N}(X_i-\bar{X})^2} + \frac{X_0^2}{\sum_{i=1}^{N}(X_i-\bar{X})^2} - \frac{2X_0\bar{X}}{\sum_{i=1}^{N}(X_i-\bar{X})^2}\right]$$

$$F^2\left[\frac{\sum_{i=1}^{N} X_i^2}{N\sum_{i=1}^{N}(X_i-\bar{X})^2} + \frac{X_0^2}{\sum_{i=1}^{N}(X_i-\bar{X})^2} - \frac{2X_0\bar{X}}{\sum_{i=1}^{N}(X_i-\bar{X})^2} + \frac{N\bar{X}^2 - N\bar{X}^2}{N\sum_{i=1}^{N}(X_i-\bar{X})^2}\right]$$

$$F^2\left[\frac{1}{N} + \frac{(X_0-\bar{X})^2}{\sum_{i=1}^{N}(X_i-\bar{X})^2}\right]$$

Where use has been made of the properties of sums:

$$\sum (X_i-\bar{X})^2 = \sum (X_i^2 - 2X_i\bar{X} + \bar{X}^2) = \sum X_i^2 - 2N\bar{X} + N\bar{X} = \sum X_i^2 - N\bar{X}^2$$

Now in predicting the true value of Y_0 by use its mean $\hat{Y}_0$ we have introduced error because the parameter estimates differ from the true values, and this is reflected in the variance of the meanforecast. But an additional error is also introduced - called a forecast error - because we have ommited consideration of the true error μ in forecasting the actual Y_0. In particular,
$Y_0 - \hat{Y}_0 = e_0$ or true Y_0 differs from the mean forecast by some unobserved and unpredictable amount - this is the forecast error.

$\text{var}(e_0) = \text{var}(Y_0) + \text{var}(\hat{Y}_0)$ as long as Y_0 and $\hat{Y}_0$ are independent

which they are by construction.

since $var(Y_0) = \sigma^2$ then $var(e_0) =$

$$F^2\left[1 + \frac{1}{N} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2}\right]$$

Regression through the origin:

What does the constant term do?

1) picks up systematic effects which in aggregate are measurable but which individually might not be -

2) alternatively, suppose that the error term has not got mean zero - then the constant term picks up the effect of this non-zero mean so that the assumption of a zero mean is not violated

ex. suppose $E(\mu) = 6$, $\mu = u + 6$, where $E(u) = 0$, and etc.

Then the model is $Y = \alpha + \beta X + \mu$ and the mean of the equation is $E(Y) = \alpha + bX + 6$

Alternatively, $Y = \alpha + \beta X + 6 + u = (\alpha + 6) + \beta X + u = \alpha' + \beta X$
whose expectation is $E(Y) = (\alpha + 6) + \beta X = \alpha' + \beta X$

Omission of the error term can lead to biased slope coefficients and inflated t-statistics.

Get a copy of diagram on page 164 of Studenmund and Cassidy to demonstrate bias

Scaling and Units of Measurement

Assume the model is $Y = \alpha + \beta X + \mu$

We know that $b = \Sigma y_i x_i / \Sigma x_i^2$

Now let $X=kW$ where k is a positive constant

so write $Y = \alpha + \beta kW + \mu = \alpha + \delta W + \mu$

Estimate δ by $d = \Sigma y_i w_i / \Sigma w_i^2$

$b = d/k$

Suppose that $Y = cZ$:

$b = c\Sigma z_i w_i / (k\Sigma w_i^2)$

$b = cd/k$

Implication : units of measurement will not affect significance of the results

Must be careful in interpreting the coefficient - does it measure the effect of $1 increase in income on expenditure, say, or a $1000 increase in income

Functional Forms:

Recall that we have assumed models which are linear in the parameters -

Fortunately this does not restrict us to linear relationships between the variables. However, we must be careful with how we specify the model - especially the error μ

ex. Suppose a model $Y = \alpha X^{\beta} \mu$ this cannot be estimated as it is because it is not linear in the parameters

But $\ln Y = \ln \alpha + \beta \ln X + \ln \mu$ can be estimated by OLS.

Except notice that the error term is strange -
in the untransformed model if $\mu \leq 0$ then $Y \leq 0$ this is bothersome, especially in economic models where variables are rarely negative
However, in the transformed model if μ is ≤ 0 then $\ln \mu$ is undefined. This suggests that $\mu > 0$ is necessary - but then $E(\mu) \neq 0$

how about $Y = \alpha X^{\beta} \exp(\mu)$ or $\ln Y = \ln \alpha + \beta \ln X + \mu$

Now μ 0 causes no problems

How do we interpret β? it is simply the elasticity of Y with respect to X. Moreover, the slope of the relationship between Y and X varies with Y and X, something which is not true in the usual model.

Semi-log models:

$\ln Y = \alpha + \beta X + \mu$

or

$Y = \alpha + \beta \ln X + \mu$

What are the elasticity and slope in each of these formulations?

βX and βY for the first and β/Y and β/X for the latter -

related to interpretation -

In the first model a 1 unit change in X generates a $\beta*100$ percent change in Y - so suppose X is a variable which takes on a value of 1 if the person belongs to a union and takes a value of zero otherwise and Y is the hourly wage; then $100*\beta$ percent is the percentage by which a union members' wage exceeds that of a non-union member

In the second model a 1 percent change in X generates a change of β units in Y

Reciprocal models:

$Y = \alpha + \beta(1/X) + \mu$ note that as X rises Y falls if $\beta > 0$

elasticity: $-\beta(1/XY)$ slope: $-\beta(1/X^2)$

Multiple Regression:

Generalize the model to contain several explanatory variables:

$Y = \alpha + \Sigma\beta_k X_k + \mu$ where k = 1,...,K indicates that there are K explanatory variables

Assume that this model is linear in the parameters, correctly specified, and that the X's are fixed. Assume that the error term has the properties assumed before - mean zero, constant variance, zero covariance, independent of the X's.

Assume also that there is no exact linear relationship between the X's.

Suppose the model is $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \mu$, and that $X_2 = d + kX_1$ where d and k are constants. Substitute into the model to get

$Y = (\alpha + d\beta_2) + (\beta_1 + k\beta_2)X_{1 + \mu = \gamma + \delta X}1 + \mu$

If there is a perfect linear relationship between any subset of the X's then the independent influence of those X's on Y cannot be determined.

Derivation and interpretation of coefficients in the multiple regression model:

$Y = a + \Sigma b_k X_k + \epsilon$

Minimize the sum of the squared deviations from the mean:

$\Sigma(Y - a - \Sigma b_k X_k)^2 = \Sigma\epsilon^2$

$$\sum_{i=1}^{N}\left(Y_i - a - \sum_{k=1}^{K} b_k X_{ik}\right)X_{ij} = 0 \quad forall j=1,...,K$$

$$\sum_{i=1}^{N}\left(Y_i - a - \sum_{k=1}^{K} b_k X_{ik}\right) = 0$$

This gives us K+1 equations to solve for K+1 parameters.

Let K=2.

$$\sum_{i=1}^{N}\left(Y_i - a - \sum_{k=1}^{K} b_k X_{ik}\right) X_{i1} = 0$$

$$\sum_{i=1}^{N}\left(Y_i - a - \sum_{k=1}^{K} b_k X_{ik}\right) X_{i2} = 0$$

$$\sum_{i=1}^{N}\left(Y_i - a - \sum_{k=1}^{K} b_k X_{ik}\right) = 0$$

If we divide this last equation by N we get:

$$\bar{y} = a + \sum_{k=1}^{K} b_k \bar{x}_k$$

And if we multiply this equation by each of the sums of the other independent variables and subtract this from our other equations we have:

$$\sum_{i=1}^{N} Y_i X_{i1} - \bar{y}\sum_{i=1}^{N} X_{i1} = b_1\left(\sum_{i=1}^{N} X_{i1}^2 - \bar{x}_1\sum_{i=1}^{N} X_{i1}\right) + b_2\left(\sum_{i=1}^{N} X_{i1}X_{i2} - \bar{x}_2\sum_{i=1}^{N} X_{i2}\right)$$

$$\sum_{i=1}^{N} Y_i X_{i2} - \bar{y}\sum_{i=1}^{N} X_{i2} = b_1\left(\sum_{i=1}^{N} X_{i1}X_{i2} - \bar{x}_1\sum_{i=1}^{N} X_{i2}\right) + b_2\left(\sum_{i=1}^{N} X_{i2}^2 - \bar{x}_2\sum_{i=1}^{N} X_{i2}\right)$$

Letting $x_{ij} = X_{ij} - \bar{x}_j$ and similarly for y_i

these two equations can be written as

$$\sum_{i=1}^{N} y_i x_{i1} = b_1\sum_{i=1}^{N} x_{i1}^2 + b_2\sum_{i=1}^{N} x_{i1}x_{i2}$$

$$\sum_{i=1}^{N} y_i x_{i2} = b_1\sum_{i=1}^{N} x_{i1}x_{i2} + b_2\sum_{i=1}^{N} x_{i2}^2$$

Note that these can be written as

$Q = Zb_1 + Wb_2$

$S = Wb_1 + Vb_2$

so we can find

$_b1 = (QV-WS)/(ZV-W^2)$

$b_2 = (SZ-WQ)/(ZV-W^2)$

$$b_1 = \frac{\left(\sum y_i x_{i1}\sum x_{i2}^2\right) - \left(\sum y_i x_{i2}\sum x_{i1}x_{i2}\right)}{\sum x_{i1}^2\sum x_{i2}^2 - \left(\sum x_{i1}x_{i2}\right)^2}$$

$$b_2 = \frac{\left(\sum y_i x_{i2}\sum x_{i1}^2\right) - \left(\sum y_i x_{i1}\sum x_{i1}x_{i2}\right)}{\sum x_{i1}^2\sum x_{i2}^2 - \left(\sum x_{i1}x_{i2}\right)^2}$$

Now suppose that we were to find the correlation coefficient between the X's:

$r_{12} = \Sigma x_{i1}x_{i2}/[(\Sigma x_{i1}^2)^{.5}(\Sigma x_{i2}^2)^{.5}]$

and if this is squared we have

$r_{12}^2 = (\Sigma x_{i1}x_{i2})^2/(\Sigma x_{i1}^2)(\Sigma x_{i2}^2)$

compare this to the denominator of the coefficient estimators.

$$b_2 = \frac{\left(\sum y_i x_{i2} \sum x_{i1}^2\right) - \left(\sum y_i x_{i1} \sum x_{i1} x_{i2}\right)}{\sum x_{i1}^2 \sum x_{i2}^2 \left[1 - \frac{(\sum x_{i1} x_{i2})^2}{\sum x_{i1}^2 \sum x_{i2}^2}\right]}$$

$$b_2 = \frac{\left(\sum y_i x_{i2} \sum x_{i1}^2\right) - \left(\sum y_i x_{i1} \sum x_{i1} x_{i2}\right)}{\sum x_{i1}^2 \sum x_{i2}^2 \left[1 - r_{12}^2\right]}$$

note that if this correlation coefficient is 1 that the denominator is zero. The correlation coefficient is 1 when there is a perfect linear relationship between the X's

Multiple coefficient of determination:

$Y = a + b_1X_1 + b_2X_2 + \epsilon$
$= \hat{Y} + \epsilon$

$\Sigma Y_i^2 = \Sigma \hat{Y}_i^2 + \Sigma \epsilon_i^2$

recall that $\Sigma \hat{Y}_i \epsilon_i = 0$

subtract $n\bar{y}^2$ from both sides - recall that $\bar{y} = \bar{\hat{y}}$ i.e. mean of actual y's equals mean of the predicted y's, and use the result on summations

$$\sum (X_i - \bar{X})^2 = \sum (X_i^2 - 2X_i\bar{X} + \bar{X}^2) = \sum X_i^2 - 2N\bar{X} + N\bar{X} = \sum X_i^2 - N\bar{X}^2$$

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

which is TSS = ESS + RSS : total sum of squared deviations = explained sum of squared deviations + unexplained sum of squared deviations of Y from its mean

$R^2 = ESS/TSS$

and this measures the degree of association between Y and all the X's.

Specification error (omitted variables bias):

Suppose that the true model is
$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \mu$

but that $Y = \alpha + \delta X_1 + v$ is estimated as $Y = a' + dX_1$. What can we say about the estimates a and d? Will d be an unbiased estimator of β_1?

Suppose that $X_2 = cX_1 + z$ where z is a random disturbance term uncorrelated with μ

substitute

$$Y = \alpha + \beta_1 X_1 + \beta_2(cX_1 + z) + \mu$$
$$= \alpha + (\beta_1 + \beta_2 c)X_1 + (z\beta_2 + \mu)$$

Now $E(d) = (\beta_1 + \beta_2 c) \neq \beta_1$, and hence since $a' = \bar{y}-d\bar{x}$
$E(a') \neq \alpha$

So both coefficient estimates are biased. Moreover, they are also inconsistent - meaning that in very large samples the bias persists.

If we know the both the expected sign of the omitted variable and the sign of the correlation between the omitted variable and the other regressors we can predict the direction but not the size of the bias.

Adjusted R^2

R^2 will never fall when an additional regressor is added because that additional variable cannot reduce the explanatory power of the other variables. Hence, we want a measure of the explanatory power which controls for the number of variables included in the equation.

This is the adjusted R^2.

Recall $R^2 = ESS/TSS = 1 - RSS/TSS$

where $RSS = \Sigma \hat{\epsilon}_i^2$ and $TSS = \Sigma \hat{y}_i^2$

Let the $\bar{R}^2 = 1 - [RSS/(N-k)]/[TSS/(N-1)]$
$= 1 - (1-R^2)(N-1)/(N-k)$

Notice that as k goes up [RSS/(N-k)] goes up if RSS is constant - that is, if the unexplained portion of Y does not change. But RSS can go down with the addition of another variable - if this decrease is sufficiently big as to offset the effect of adding the addition variable on the numerator then the adjusted R^2 can rise, if not the

measure can fall.

Notice also that the adjusted R^2 can be negative (in which case it is reported as zero). As k gets large (N-1)/(N-k) gets very large, and if the R^2 is small then $(1-R^2)$ is relatively large so the negative term in the adjusted R^2 can become greater than 1 in absolute value.

Comparing R^2's is not appropriate when the dependent variables differ - for example, if in one case it is Y and in another ln Y - because it measures the extent to which the variation in the dependent variable is explained by the model.

Multiple regression and inference:

tests on individual coefficients can be done as before.

It is not correct, however, to test several hypotheses from the same regression - once one test is performed then the subsequent tests are conditional on the first test.

Can test the overall significance of the regression: that is, whether or not the dependent variable is related to the dependent variables. Note that some variables may be individually insignificant from zero but pass this overall test - this would be the case if there was a high degree of correlation between the X's.

The F test

Suppose the model is $Y = \alpha + \Sigma\beta_k X_k + \mu$. (Where k runs from 1 to K-1) Let the null hypothesis be that each of the β_k's is simultaneously equal to zero, $H_0: \beta_1=\beta_2=...=\beta_K=0$, the alternative hypothesis is that they are not simultaneously equal to zero.

We can estimate the model above and get an estimate of the residual sum of squares, call this the RSS unrestricted. Next estimate the model $Y = \alpha + \mu$ and calculate the RSS - call this the residual sum of squares restricted. The restriction is that the β's are all constrained to be zero.

If the null hypothesis is true then the difference RSSR - RSSU should be small.

F = ((RSSR - RSSU)/G)/RSSU/(N-K) is distributed F with G, and N-K degrees of freedom. If ((RSSR - RSSU)/G)/RSSU/(N-K) > $F_{.95}$(G,N-K) then reject the null hypothesis that all β's are simultaneously zero. G is the number of restrictions which have been imposed; in this example, K-1.

Most regression programs perform this test as part of the standard output.

An alternative example

suppose that you have a production function
$Q = \alpha K^{\beta} L^{\delta} e^{\mu}$ which you can estimate as $\ln Q = a + b\ln K + d\ln L + \epsilon$.

If you hypothesize that $\beta+\delta=1$ you can test this hypothesis in a straightforward manner -

Estimate the unrestricted model $\ln Q = a + b\ln K + d\ln L + \epsilon$.
and find the RSSU

Then, impose the restriction $\ln Q = a + b\ln K + (1-b)\ln L + \epsilon$

or ln Q = a + b(lnK -lnL) + lnL + ϵ and estimate and calculate the RSSR.

Form F=((RSSR - RSSU)/1)/RSSU/(N-3) and compare to $F_{.95}(1,27)$.

Example: N=30, RSSU=1, RSSR=1.2.

F=(.2/1)/(1/27)=5.4; $F_{.95}(1,27)$=4.25. Since 5.4>4.25 reject the null hypothesis that $\beta+\delta=1$.

The relationship between the F test and the R^2

F = ((RSSR - RSSU)/G)/RSSU/(N-K).

Now multiply and divide by TSS, note that TSS is the same for both the restricted and the unrestricted models. And recall that R^2= 1 - RSS/TSS.

F = ((RSSR/TSS - RSSU/TSS)/G)/((RSSU/TSS)/(N-K))

= (((1-R^2r) - (1-R^2u))/G)/((1-R^2u)/N-K) = (R^2u-R^2r/G)/((1-R^2u)/N-K)

Application of Dummy explanatory variables:

Some variables which might be used to explain the variation in our dependent variable only reflect "qualitative" aspects of an individual economic agent time period; there may be no quantitative variation. For example, we might have a variable which indicates the presence of children in a household but does not indicate how many, or we might know whether or not an individual is married; or we might know that during some time there was a war, or some law was in force which was not at other times. Each of these is an example of a variable which we might believe will explain some of the variation in our dependent variable but which is only "qualitative" in nature.

Take a model such as the following:
$Y = \alpha + \beta X + \delta D + \mu$
where μ satisfies all the usual assumptions and D = 1 if the individual is male and equals zero otherwise.

This is like estimating a model for males and a model for females but forcing each sex to respond the same to changes in X. That is, only the intercept differs by sex, we have two parallel lines. The estimate of α is the intercept for the group for whom D=0, and $\alpha+\delta$ is the intercept for the group for whom D=1. The group for whom D=0 is called the base group because it is against them that the other group is compared.

Suppose that the dummy variable has more than two classifications or groups- for example - race: white, black, other
Then D1 = 0 if not white =1 if white, D2 = 0 if not black = 1 if black, D3 = 0 if white or black, = 1 if other.

$Y = \alpha + \beta X + \delta D3 + \gamma D2 + \mu$
In this case we must omit one of the dummy variables to avoid perfect multicollinearity; the estimate of α is the estimate of the intercept for the base/omitted group.

Alternatively, we could create a variable which is XD=XxD to test to see if the slope varies by sex:
$Y = \alpha + \beta X + \delta XD + \mu$
Note that as X changes Y changes by β if D=0 and by $\beta+\delta$ if D=1

the statistical significance of the estimate of δ then is a test of whether or not the two groups respond in the same fashion to a change in X.

The Chow Test

An alternative, but more involved method of testing for differences is to estimate the model imposing the same behavior on each group and to estimate the same model for only those observations which belong to each group. Assume group 1 has N1 and group 2 N2 members. Estimate the model for all N1 + N2 together and get a RSS_r. This model has N1+N2 - k degrees of freedom where k is the number of coefficients estimated. Next estimate the model for each group and get an RSS1 and an RSS2; sum these to get the RSS_u. The unrestricted model has N1 + N2 -2k degrees of freedom.

Calculate F=((RSSR - RSSU)/k)/RSSU/(N1+N2-2k) and compare to $F_{.95}$(k,N1+N2-2k).

The two approaches are equivalent but the dummy variable approach is easier and likely to provide more information - which coefficients are different and which are the same. Moreover, pooling, and using dummy variables, is likely to give relatively more precise estimates of the coefficients because of the additional data points.

Specification Tests:

attributes of a good model

1) parsimony - the model must focus on the most important aspects of the situation being modeled - don't throw in everything but the kitchen sink

2) the parameters must be identified - we will talk about this more later; in essence, we want to be able to seperate effects from a given cause into identifiable amounts with no uncertainty/indeterminacy

3) the model should fit well

4) Consistent with the theory

5) have predictive power

Types of Specification errors:

1) omitted variable

at least the intercept, and possibly the slopes are biased and the bias does not go away in large samples, error variance and coefficient variances are incorrectly estimated, hypothesis tests may give misleading results

2) irrelevant variable

coefficients are unbiased and consistent, but not minimum variance

testing for irrelevant variables

Use an F or a t test on those variables which are suspect - but cannot do so iteratively, the test must be done on all at once.

After conclusion is drawn it is not appropriate to go back and test other variables in the model.

3) incorrect functional form - coefficients are biased and inconsistent - demonstrate this with a true model $Y = \alpha + \beta \ln X + \mu$
and a think model $Y = \alpha + \beta X + \mu$

Test for functional form and omitted variables:

examine residuals - heteroskedasticity and autocorrelation can be
evidence of incorrect form or omitted variables

4) measurement error

In the dependent variable - OLS is unbiased however variances are not minimum for the error from the measurement problem must be included in the error of the true residuals - this is at least a little circular since we defined measurement error in the dependent variable as a source of the regression error

In an independent variable - suppose $X_i = X_i^* + e_i$

$Y_i = \alpha + \beta(X_i - e_i) + \mu_i = \alpha + \beta X_i + (\mu_i - \beta e_i) = \alpha + \beta X_i + v_i$

$E[(v_i - E(v_i))(X_i - E(X_i))] = E[(\mu_i - \beta e_i)e_i] = -\beta\sigma_e^2 \neq 0$

We will see how to address this problem later in the semester when we talk about instrumental variables and simultaneous equations estimation.

More on Multicollinearity:

Perfect multicollinearity we saw as:

$0 = c_0 + \Sigma c_k X_k$

and we saw that in such a case the model $Y = \alpha + \Sigma\beta_k X_k + \mu$ can not be estimated

If collinearity is not perfect then the model may be estimated.

In the presence of a high degree of collinearity one is likely to find:

a) a high R^2 but b) low t-ratios, c) wide confidence intervals and d) sensitivity of estimates and standard errors to small changes in the data.

Low t-ratios are the product of high correlation between the variables which makes for a large variance in the estimate of the coefficient -

$var(b_1) =$

$$var(b_1) = F^2 \frac{\sum x_{i2}^2}{\sum x_{i1}^2 \sum x_{i2}^2 (1 - \frac{(\sum x_{i1} x_{i2})^2}{\sum x_{i1}^2 \sum X_{i2}^2})}$$

$$var(b_1) = F^2 \frac{\sum x_{i2}^2}{\sum x_{i1}^2 \sum x_{i2}^2 (1 - r_{12}^2)}$$

and notice that the greater is the correlation between the X's the greater is the variance in the estimate of b_1, similarly for b_2.

Heteroskedasticity

when the variance of the error term is not constant across the sample.

Coefficient estimates are still unbiased but are no longer minimum variance estimators. In fact, variance estimated under the formula when μ has constant variance is really a weighted average of the variances when μ is heteroskedastic:

$$var(b) = E\left(\$ + \frac{\sum_1^n x_i \mu_i}{(\sum_1^n x_i^2)^2} - \$ \right)^2$$

$$var(b) = \left(\frac{1}{(\sum_1^n x_i^2)^2} \right) E\left(\sum_1^n x_i^2 \mu_i^2 \right)$$

$$var(b) = \frac{\sum_1^n x_i^2 F_i^2}{\left(\sum_1^n x_i^2 \right)^2}$$

Note that if σ_i^2 is the same for all i that this is precisely the variance formula we found before.

Sources of heteroskedasticity:

1. learning from ones mistakes - errors inbehavior become smaller over time

2. the greater is discretionary income the greater is the possible variation in spending - as less of income is "subsistence" and more is capable of being spent or saved. Similar stories can be told for firms with large profits and etc

3. data collection techniques may be more precise in large firms than in small firms; wealthy people have accountants who keep better records than poor people who keep their own records

Tests for heteroskedasticity:

1) model specification - may be clear that the variance in the dependent variable is likely to differ across size - for example, profits are likely to vary more for firms with large profits than with small profits.

2) formal tests

formal tests and corrections hypothesize that the heteroskedasticity is related to some set of the independent variables and make use of this fact

a) Park test - estimate the model and use the log of the estimated residuals squared as the dependent variable in a regression on a constant and the log of the independent variable you hypothesize might cause the heteroskedasticity - if this variable is found to be significant then you can reject the null hypothesis of no heteroskedasticity

b) Goldfeld-Quandt test - reorder the data based on the hypothesized independent variable ; split into two samples dropping a few observations in the middle of the data. Estimate the model on both the small and the large data sets and form the ratio of their residual sums of squares $(RSS_l/RSS_s)(df_s/df_l)$ and reject the null hypothesis of no heteroskedasticity if greater than $F_{.95}(dfl,dfs)$.

c) estimate original model and get estimates of the errors
square these estimates and regress them on a polynomial in the X's. Test the hypothesis that all the slope coefficients are simultaneously zero. If this hypothesis is rejected you have heteroskedasticity.

Examples using mystat.

Corrections for heteroskedasticity

In general suppose that $E\mu^2 = \sigma^2 f(X_i)$

Note that if we divide μ_i by $(f(X_i))^{.5}$ that now the variance is constant. Therefore, our model by $(f(X_i))^{.5}$

Must be careful in interpreting the coefficients.

Examples in mystat.

depends on X, $X^{2,}$ and polynomial in X

Autocorrelation

Usual assumption is that $E(\mu_i\mu_j) = 0$ for $i \neq j$. But suppose that this is not true - OLS still unbiased but as in heteroskedasticity is no longer minimum variance

$$var(b) = E\left(\$ + \frac{\sum_1^n x_i \mu_i}{(\sum_1^n x_i^2)^2} - \$ \right)^2$$

$$var(b) = \left(\frac{1}{(\sum_1^n x_i^2)^2} \right) E\left(\sum_1^n x_i^2 \mu_i^2 + \sum_{i=1}^n \sum_{j=1, j \neq i}^n x_i x_j \mu_i \mu_j \right)$$

$$var(b) = \frac{\sigma^2}{\sum_1^n x_i^2} + \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n x_i x_j \sigma_{ij}}{\left(\sum_1^n x_i^2 \right)^2}$$

Notice that in this case we may under or overestimate the variance of b since covariances may be positive or negative.

Sources of autocorrelation:

1. inertia
2. specification bias - either due to omitted variables or incorrect functional form

omitted variables - error in estimated equation $v = \beta_o X_o + \mu$ - hence estimated residuals appear related because of the systematic effect of X_o

incorrect functional form - suppose that we fit a line to a non-linear relationship. Then estimated errors will tend to be related
draw diagram of relationship

3. lags - today is affected by yesterday (inertia again)
4. data manipulation - averaging or aggregating tends to smooth data making patterns more likely

Assume that the autocorrelation takes some pattern:

$\mu_t = \rho\mu_{t-1} + \epsilon_t$ where ϵ has the properties of the usual OLS error term. ρ tells us the relation between errors in each period - note that we are assuming that the relationship between period 1 and 0 is the same as the relationship between period 100 and 99. This model is called first order autocorrelation and μ is AR(1).

If instead we had $\mu_t = \rho\mu_{t-1} + \delta\mu_{t-2} + \epsilon_t$ then we have an AR(2) model - second order autocorrelation.

What we are assuming in any case is that movement in the error is composed of a systematic and a random part. In correcting for autocorrelation in our estimation we make use of this systematic part to improve our estimates.

Testing for autocorrelation:

1. estimate using OLS and plot the residuals -eyeball it

2. runs test - use the estimated residuals to see if there are fewer or more sequences of errors of the same sign than would happen by chance

Let N = # of observations - N = N1 + N2
N1 = # of observations with positive estimated error
N2 = # of observations with negative estimated error
n = number of runs - uninterrrupted sequence of + or - errors; that is, times the sign changes + 1

$E(n) = (2N1N2 + N)/N$
$\sigma^2 = 2N1N2(2N1N2 - N)/(N^2(N - 1))$

form a confidence interval using E(n) and σ. If n is in this CI then do not reject the null hypothesis of no autocorrelation; else reject.

3. chi square test of independence - compare frequencies of positive followed by positive, negative by negative, etc. with frequencies if errors are independent - use chi square distribution

4. Durbin-Watson - calculate the ratio of the sum of squared differences in successive residuals to the RSS of the model and compare to the d statistic - difficulty is that this only applies to models with constant terms, no lagged values of the dependent variable as regressors, no missing observations, and first order autocorrelation.

And there are regions where one cannot say with certainty if autocorrelation is or is not present.

$$d = \frac{\sum_{i=2}^{n} (,_t - ,_{t-1})^2}{\sum_{i=1}^{n} ,_t^2}$$

$$d=\frac{\sum_{t=2}^{n} e_t^2+\sum_{t=2}^{n} e_{t-1}^2-2\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}$$

$$d \quad \frac{2(\sum_{t=2}^{n} e_t^2-\sum_{t=2}^{n} e_t e_{t-1})}{\sum_{t=1}^{n} e_t^2}$$

$$d \quad 2(1-\rho)$$

where ρ is the estimate of the first degree autocorrelation.

and since -1 ≤ ρ ≤ 1 it follows that 0 ≤ d ≤ 4.

Moreover, if ρ = 0 then d = 2, so values of d close to 2 are consistent with the null hypothesis of no first order autocorrelation; the closer is d to 0 the greater is the evidence of positive auto correlation; the closer is d to 4 the stronger is the evidence of negative auto correlation.

Use the d calculated by the program (or calculate it yourself) and compare to the values in the table.

From our MYSTAT regression we had a DW = .776 (I think), with 25 observations and two explanatory variables and a constant - go to Table
dl = .981, du = 1.305

.776 ≤ .981 so reject null hypothesis of no autocorrelation.

Supppose DW = 2.986 - then 4-.981 = 3.019, and 4-1.305 = 2.695
and 4-du ≤ 2.986 ≤ 4 -dl which implies uncertain/ indeterminate
Methods of solving the autocorrelation problem:

suppose that ρ is known

$$Y_t=\beta+\beta X_t+\mu_t \quad where \ \mu_t=\rho\mu_{t-1}+v_t$$

$$D Y_{t-1} = D\alpha + D\beta X_{t-1} + D\mu_{t-1}$$

$$Y_t - D Y_{t-1} = (1-D)\alpha + \beta (X_t - D X_{t-1}) + \mu_t - D\mu_{t-1}$$

$$Y_t - D Y_{t-1} = (1-D)\alpha + \beta (X_t - D X_{t-1}) + v_t$$

Must drop the first observation since there is no prior period to use in transforming it - this leads to OLS on this model not quite being correct if the sample is small, if the sample is large the effect of the lost observation need not be large.

One can transform the first period data by multiplying each variable in the first period by $(1-\rho^2)^{.5}$ This procedure works because the transformation of the Y and X variables affects the error in this period.

$$var(\mu_t) = D^2 var(\mu_{t-1}) + var(v_t)$$

$$var(v_t) = (1-D^2)\, var(\mu)$$

$$(1-D^2)^{.5} Y_1 = (1-D^2)^{.5}\alpha + \beta (1-D^2)^{.5} X_1 + (1-D^2)^{.5}\mu_1$$

Finding the variance of $Y_1(1-\rho^2)^{.5}$ we get var (v), and of course the mean of $(1-\rho)^2\mu^5$ is zero, and the transformation of the other data points resolves the autocorrelation.

Use estimate of ρ from original OLS estimation and DW statistic in making the transformation. One can perform this procedure repeatedly, using estimates of ρ to redefine the data and re-estimate the model, until the value of ρ between estimates changes by some very small amount, say .01 or .005. This is known as the Cochrane-Orcutt iterative procedure.

Autoregressive and Distributed Lag Models:
An autoregressive model uses prior values of the dependent variable as explanatory variables

$$Y_t = \alpha + \sum_{k=1}^{K-1} \beta_k X_{tk} + (Y_{t-1} + \mu_t$$

a Distributed lag model uses current and past values of the explanatory variables as regressors

$$Y_t = " + \sum_{k=0}^{K-1} \$_k X_{t-k} + \mu_t$$

Of course, other regressors are possible in the distributed lag model which may or may not be lagged.

Lags in distributed lag models may appear because:

1) psychological reasons - people's behavior adjusts only slowly to new situations

2) Technological reasons - it may be difficult to rearrange productive techniques in a short period of time, or people may be waiting to see how events play out

3) institutional reasons - contracts may be in force for several years, the decision making process and command and control structure may take a long time to filter decisions down from the top to those at the bottom

In a distributed lag model one can interpret coefficients as short, intermediate and long run responses

$$Y_t = " + \$_1 X_t + \$_2 X_{t-1} + \$_3 X_{t-2} + \mu_t$$

β_1 may be interpreted as the short run response, $(\beta_1 + \beta_2)$ as the intermediate run response and $(\beta_1 + \beta_2 + \beta_3)$ as the long run response. Note that the longer is the lag structure the potentially larger is the long run effect - it may not be finite.

Estimation of the distributed lag model:
It is clear that the choice of the number of past values of the variable to include is arbitrary, being limited only by the availability of data. Theory rarely, if ever, says anything about the length of the lag structure. It may suggest however that the farther in the past is an event the smaller the effect it is likely to have, or it might suggest that immediate effects are weak that intermediate effects are strong and then more distant effects are weak again. We want to be able to estimate these varying patterns of effects.

Koyck Lag Model:

Let $\beta_k = \delta^k \beta_o$ k=0,1,2,3,...
where $0<\delta<1$. This implies that all the β's are of the same sign, that the effect of the independent variable declines as time passes, and that the long run effect is finite.

In this formulation δ is known as the rate of decline and $1-\delta$ as the speed of adjustment.

$$Y_t = " + \$_0 \sum_{k=0} * ^k X_{t-k} + \mu_t$$

But of course we cannot estimate this model since we do not have an infinite amount of data. Hence

$$*Y_{t-1} = *\alpha + *\beta_0 \sum_{k=0} *^k X_{t-k-1} + *\mu_{t-1}$$

is subtracted from the original model to get

$$Y_t - *Y_{t-1} = \alpha(1-*) + \beta_0 X_t + \mu_t - *\mu_{t-1}$$

$$Y_t = \alpha(1-*) + \beta_0 X_t + *Y_{t-1} + \mu_t - *\mu_{t-1}$$

Note that now the model is autoregressive in that past values of the dependent variable are used as explanatory variable. This may cause problems because the Y's are stochastic and may be correlated with the error. In addition, we must be concerned about the behavior of our new residuals - for if we assumed theat the μ are uncorrelated the $\mu_t - \delta\mu_{t-1} = v_t$ will be autocorrelated. The presence of the lagged Y violates an assumption of the durbin watson test so we must use some other technique to test for autocorrelation.

Rationalizing the Koyck Lag model:

Adaptive expectations

suppose that people's beliefs about some variable affect their decision making - for example, beliefs about what government behavior is likely to be affect decisions about saving and spending

$$Y_t = \alpha + \beta X_t^e + \mu_t$$

The question is, how are expectations formed, and how do they relate to observed data?

$$X_t^e - X_{t-1}^e = (1-\gamma)(X_t - X_{t-1}^e)$$

That is, suppose I adjust my expectations by some fraction of the gap between the current value and the previous expeced value. Substitute this expression for my current expectation and lag the model, multiply by (1-γ) and subtract from the original to get

$$Y_t = \alpha\gamma + \beta\gamma X_t + (1-\gamma) Y_{t-1} + \mu_t + (\mu_t - (1-\gamma)\mu_{t-1}$$

Partial Adjustment Model:

$$Y_t^* = \alpha + \beta X_t + \mu_t$$

where Y* is desired level of some activity or behavior or level inventories or capital stock. Movement is made toward the desired level through a gradual process

$$Y_t - Y_{t-1} = \lambda (Y_t^* - Y_{t-1})$$

$$Y_t = \lambda (\alpha + \beta X_t + \mu_t) + (1-\lambda) Y_{t-1}$$

$$Y_t = \lambda\alpha + \lambda\beta X_t + (1-\lambda) Y_{t-1} + \lambda\mu_t$$

This model is autoregressive but is less likely than the adaptive expectations or Koyck models to suffer from correlation between the error and the lagged value of the dependent variable because the lagged error is not part of the current error.

Instrumental Variables:

rather than use the lagged value of the dependent variable as a regressor use something that is highly correlated with it but which we know is uncorrelated with the error

Consider our original model

$$Y_t = \alpha (1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + v_t$$

which will generate first order conditions

$$\sum_{t=1}^{N} \left(Y_t - a - bX_t - cY_{t-1}\right) X_t = 0$$

$$\sum_{t=1}^{N} \left(Y_t - a - bX_t - cY_{t-1}\right) Y_{t-1} = 0$$

Show how this equation will not be true except by biasing the coefficients.

$$\sum_{t=1}^{N} \left(Y_t - a - bX_t - cY_{t-1}\right) = 0$$

Replace the second of these equations with

$$\sum_{t=1}^{N}\left(Y_t - a - bX_t - cY_{t-1}\right)X_{t-1} \quad = \quad 0$$

Show how this is true because X is uncorrelated with the errors.

The problem is finding variables which are highly correlated with the lagged dependent variable and uncorrelated with the errors - there is not a unique set of instruments - the estimators are not likely to be minimum variance estimators.

Detecting Autocorrelation in models with lagged dependent variables as regressors: The Durbin h test

In large samples

$$h = \Box\left(\frac{N}{1 - N(var(c))}\right)^{.5}$$

is distributed normally with mean zero and variance 1, where c is the estimate of the coefficient on the lagged value of the dependent variable. Recall that if there is no auto correlation ρ will be close to zero - hence if h falls between -1.96 and +1.96 then cannot reject the null hypothesis of no autocorrelation.

Almon Lag Model:

Recall our original distributed lag model

$$Y_t = " + \$_1 X_t + \$_2 X_{t-1} + \$_3 X_{t-2} + \mu_t$$

each of the β_i's can be approximated by a polynomial in the length of the lag

$$\$_i = p_0 + p_1 i + p_2 i^2 + p_3 i^3$$

now substitute for each of the β_i's using this expression

$$Y_t = " + \sum_{i=1}^{k} (p_0 + p_1 i + p_2 i^2 + p_3 i^3) X_{t-i} + \mu_t$$

$$Y_t = " + p_0 \sum_{i=0}^{k} X_{t-i} + p_1 \sum_{i=0}^{k} i X_{t-i} + p_2 \sum_{i=0}^{k} i^2 X_{t-i} + p_3 \sum_{i=0}^{k} i^3 X_{t-i} + \mu_t$$

estimate this relationship and use the estimated p's to estimate the β's from the original model.

Note problems: must decide on the length of the lag k
must decide on the order of the polynomial - this should be dictated by the number of turning points (one more degree than turning points) with an eye toward a small order polynomial
the variables $Z_j = \Sigma_i i^j X_{t-i}$ will likely be highly collinear and may make hypothesis tests of the p's very imprecise - but this would not necessarily mean that the β's were insignificant

Advantages: no serial correlation or regressors correlated with errors problems,
the shape of the lag is determined by the data - effects may rise then fall then rise again or be small, then rise and ultimately fall - need not fall throughout
may reduce substantially the number of parameters to be estimated -a model with lag of k periods can be estimated with only 3 or four parameters if β's are quadratic or cubic in the lag.

Simultaneous Equations:

Consider the supply and demand model:
$Q_d = \alpha_0 + \alpha_1 P + \mu_d$
$Q_s = \beta_0 + \beta_1 P + \mu_s$
$Q_d = Q_s = Q$

Solve for Q and P:
$Q = \alpha_0 + \alpha_1(\alpha_0-\beta_0)/(\beta_1-\alpha_1) + (\beta_1\mu_d - \alpha_1\mu_s)/(\beta_1-\alpha_1)$
$P = (\alpha_0-\beta_0)/(\beta_1-\alpha_1) + (\mu_d - \mu_s)/(\beta_1-\alpha_1)$

Note that P (and Q) in equilibrium depends upon the error in each equation - in other words, in estimating either the demand or the supply curve the the independent variable is correlated with the equation error -

Therefore, OLS is biased and inconsistent

Definitions:

endogenous variables
predetermined variables

Macro example:
$C = \alpha + \beta Y + \mu$ - consumption as a function of income
$Y = C + I$ - accounting identity income equals consumption + investment

Take I as a predetermined variable, C and Y are both endogenous, or jointly determined

Substitute for C to get a reduced form equation for Y

$$\begin{aligned} Y &= \alpha + \beta Y + I + \mu \\ &= \alpha/(1-\beta) + I/(1-\beta) + \mu/(1-\beta) \\ &= \pi_1 + \pi_2 I + w \end{aligned}$$

and using this inthe consumption equation

$$\begin{aligned} C &= \alpha + \beta(\alpha/(1-\beta) + I/(1-\beta) + \mu/(1-\beta)) + \mu \\ &= \alpha/(1-\beta) + \beta I/(1-\beta) + \mu/(1-\beta) \\ &= \pi_3 + \pi_4 I + w \end{aligned}$$

And from the π's we can recover the α and β. This model is said to be identified.

Consider again the supply and demand example:

Note that in that case we cannot recover the parameters of the supply and demand curves. This model is not identified.

The macro model above is called just or exactly identified.

Take the supply and demand model and add to the demand function an income term

$$Q_d = \alpha_0 + \alpha_1 P + \alpha_2 M + \mu_d$$
$$Q_s = \beta_0 + \beta_1 P + \mu_s$$
$$Q_d = Q_s = Q$$

Use the equilibrium condition to solve for P

$$P = (\alpha_0-\beta_0)/(\beta_1-\alpha_1) + \alpha_2 M/(\beta_1-\alpha_1) + (\mu_d - \mu_s)/(\beta_1-\alpha_1)$$
$$= \pi_1 + \pi_2 M + w$$

and Q

$$Q = (\alpha_0\beta_1 - \alpha_1\beta_0)/(\beta_1 - \alpha_1) + \alpha_2\beta_1 M/(\beta_1 - \alpha_1) + (\beta_1\mu_d - \alpha_1\mu_s)/(\beta_1 - \alpha_1)$$
$$= \pi_3 + \pi_4 M + z$$

Now the question is can we identify any of the underlying parameters? Note that there are 4 reduced form parameters but 5 structural parameters.

comparing π_2 and π_4 reveals that $\beta_1 = \pi_4/\pi_2$,

and $\beta 0 = \pi_3 - \beta_1\pi_1$

So the parameters of the supply function can be recovered. Unfortunately, the parameters of the demand cannot. There is not enough information to identify 5 parameters from 4 equations.

Now add one more variable to the demand equation and one to the supply equation

$$Q_d = \alpha_0 + \alpha_1 P + \alpha_2 M + \alpha_3 A + \mu_d$$
$$Q_s = \beta_0 + \beta_1 P + \beta_2 W + \mu_s$$
$$Q_d = Q_s = Q$$

where A is wealth and W is the wage rate

The reduced form equations are
$$Q = \pi + \delta M + \gamma A + \eta W + z$$
$$P = \psi + \lambda M + \sigma A + \chi W + u$$

In this system there are 8 reduced form paramenters and 7 structural parameters. It will be possible to find unique

values of the demand function parameters; it is just identified. HOwever, solkving for the parameters of the supply function may lead to contradictions since there are 3 parameters but four equations. The supply function is over-identified.

Conditions for identification:

Order condition - in a model of M simultaneous equations for any equation to be identified it must exclude at least M-1 variables appearing in the model (either predetermined or endogenous). if it excludes exactly M-1 it is said to be just identified; if it excludes more than M-1 it is over identified. This is necessary but not sufficient.

Equivalently, to be identified any equation must exclude at least as many predetermined variables as it includes endogenous variables on the right hand side.

So, suppose there are K predetermined variables in the system, k of which are in the equation of interest. There are then K-k predetermined variables excluded from the equation of interest. And suppose there are m endogenous variables on the right hand side of the equation. The equation is identified if $K-k \geq m-1$. If the statement holds as an equality then the equation is just identified; if as an inequality the equation is over identified.

Estimation of identified equations:

If the equations are just identified then we saw that one can use OLS to estimate the reduced form equation and recover the structural coefficients. These estimates are biased but consistent.

Two Stage Least Squares:

Recall that simultaneous equations bias arose because the endogenous variable on the right hand side was correlated with the equation error. What we want to do is find a proxy for this RHS endogenous variable which is not correlated with the error. This is the instrumental variables approach we briefly mentioned while talking about distributed lag and autoregressive models.

The TSLS method is to use as a proxy for the endogenous variable its value predicted from regressing it against all the predetermined variables in the system. The coefficients are biased but consistent. The bias comes from the influence of the error on the coefficient estimates in the first stage which genereates the predicted value used in the second stage.

Take again the supply and demand model:

$Q_d = \alpha_0 + \alpha_1 P + \alpha_2 M + \alpha_3 A + \mu_d$
$Q_s = \beta_0 + \beta_1 P + \beta_2 W + \mu_s$
$Q_d = Q_s = Q$

Recall that the demand equation was just identified but the supply equation was over identified.

Estimate $P = a + bM + cA + dW + e$

and calculate $P' = a' + b'M + c'A + d'W$

estimate

$Q_d = \alpha_0 + \alpha_1 P' + \alpha_2 M + \alpha_3 A + (\alpha_1 e + \mu_d)$
$Q_s = \beta_0 + \beta_1 P' + \beta_2 W + (\beta_1 e + \mu_{s)}$

Using TSLS the parameters of the over identified equation can be estimated uniquely.

Note that the standard error of the second stage is not the standard error of the original model because of the use of the predicted rather than the true P. This means that t-statistics printed by the output are wrong.

Find the correct standard error by using the coefficient estimates from the second stage regression and the true values of the RHS variables to calculate predicted values of μ_d and μ_s and use these values to calculate the variance (standard error) of the regression as usual.

Dummy and limited Dependent Variable Models

Suppose that you have a model such as we usually do, but where Y = 1 if something is true, and equals zero otherwise.

Examples are: Y = 1 if individual voted for Jesse Helms, Y = 0 else; Y = 1 if individual purchased season basketball tickets, Y = 0 otherwise; Y = 1 if individual lives in Orange county, Y=0 else.

In these examples the dependent variable takes on a limited number of values.

Or it may be that we observe the dependent variable only if it, or some other variable, is sufficiently large. An example of this latter would be that we observe contributions to political campaigns only if they are bigger than 0, but something influences the decision of whether to make such a contribution at all. These models are very important and useful but are beyond the scope of this class.

Linear Probability Model

Write the usual model with the assumption that Y=1 only if something is true, such as the individual voted for Jesse; and Y=0 otherwise.

Now recall that given the model $Y = \alpha + \beta X + \mu$ that
$E(Y/X) = \alpha + \beta X$.

Now, let P = prob(Y=1) and (1-P) = prob (y=0).
$E(Y/X) = P*1 + (1-P)(0) = \alpha + \beta X$
but this implies that $P = \alpha + \beta X$.

Recall that Probabilities must be between 0 and 1; that the sum of the probabilities over a set of independent events must equal 1.

Problem is that there is nothing inherent in the model guaranteeing that $\alpha + \beta X$ be between 0 and 1.

In addition, think about the error terms. $\mu = 1 - \alpha - \beta X$,
or $\mu = -\alpha - \beta X$. The error term takes on only two values; but for hypothesis tests we assume it is normally distributed.

Finally, it is straightforward to show that the μ is heteroskedastic.

estimate a model using Mystat

Alternative approaches: Probit and Logit

Logit Model:

Let $P = E(Y=1|X) = 1/(1 + \exp(-(\alpha + \beta X)))$

and $1-P = (1 + \exp(-(\alpha+\beta X)) - 1)/(1 + \exp(-(\alpha + \beta X)))$
$= 1/(1 + \exp(\alpha + \beta X))$

$P/(1-P) = \exp(\alpha + \beta X)$

and $\ln[P/(1-P)] = \alpha + \beta X$

For estimation purposes $\ln[P/(1-P)] = \alpha + \beta X + \mu$

But for any given individual the probability that they are in a given catagory is either 0 or 1 since we know in which they are. Therefore, group individuals of a similar type, say income level, or city or neighborhood, and find the relative frequency that individuals of that type satisfy the condition for Y=1 and Y=0

Now, if the groupings have large numbers of members then it can be shown that μ has mean 0 and variance $1/(N_iP_i(1-P_i))$ where N is the number of individuals in the group. Note that this means there is a problem of heteroskedasticity in the model.

Probit Model:
We are interested in knowing what affects the decision to purchase a car, or to attend college, or to vote for a particular candidate, What we observe is that people either did or did not purchase the car, attend college, or vote for the candidate.

We suppose that the individual makes that choice which maximizes her utility, but we cannot measure the utility in either case; and of course since the individual only chooses one of the alternatives all we can say is the alternative chosen must provide greater utility than the alternative foregone.

Let $I^* = \alpha + \beta X + \mu$ be an unobservable index of utility

and let I=1 if the individual purchases, attends, or votes for, and I = 0 otherwise.

If $I^* > 0$ then I = 1, Then $\mu > -\alpha - \beta X$
$I^* \leq 0$ then I=0. Then $\mu \leq -\alpha - \beta X$.

Assume that μ is normally distributed. Then the prob(I=1) = prob($\mu > -\alpha-\beta X$); prob(I=0) = prob($\mu \leq -\alpha-\beta X$).

If the data can be grouped by X, or ranges of X, then we can take the relative frequencies as probabilities just as

we did above.

For a given X we can take the relative frequency of I=1 as an estimate of the probability and use the inverse of the normal CDF to get an estimate of I* which we will call I'. Use I' (or I' + 5) as the dependent variable in the regression

$I' = \alpha + \beta X + \mu$

where μ has mean zero and is uncorrelated with past or future values of itself. μ is heteroskedastic.