ECON 6620 **Econometrics 1** **Class Assignment 2**

Chow Test

The Chow test is an interesting example of an F-test, used to see if the estimated coefficients are significantly different when estimated over different subsets of the data. In a time series, you might have good reason to believe that relationships underwent a fundamental change at some date (In macroeconomics, the oil shock of 1973-1974 appears to mark such a structural break). Likewise, a Chow test can be used with cross-sectional data in cases where you suspect the model parameters may be significantly different in different subsets of your observations. For example you may believe the model to work differently in rural areas than in urban. Dummies can be used to test and correct for this kind of structural change, but a Chow test is often easier than creating dummies for intercepts and slopes. The null hypothesis in a Chow test is that the model is the same for both subsets. If the F-statistic is high enough, you reject the null hypothesis. Here are the steps of a Chow test:

1. Estimate the model over the observations in subset 1, saving the sum of squared errors (ESS1).
2. Estimate the model over the observations in subset 2, saving the sum of squared errors (ESS2).
3. Set ESSUR=ESS1+ESS2. This is *unrestricted* since you are allowing the coefficients to be different in the two subsets. The degrees of freedom for the unrestricted model is equal to the sum of the observations in the two models minus the sum of the number of estimated coefficients in the two models.
4. Estimate the model over all observations, saving the sum of squared errors. This is your ESS *restricted*, since your coefficients are restricted to be the same in both subset 1 and subset 2. The number of restrictions is equal to *two* times the number of estimated parameters, since you are restricting the coefficients estimated over the *two* subsets to be equal to the coefficients estimated over the entire set of observations.
5. Now you construct your usual F-statistic:

$$F_{\#restrictions,dfUR}=((ESSR-ESSUR)/\#restrictions)/(ESSUR/dfUR)$$

J-Test

The J-test is used to compare two different models (Model_1 and Model_2), each attempting to explain exactly the same dependent variable, but with at least one variable in each model not found in the other. Based on a t-statistic, the J-test can help us decide if one model is superior to the other, or if the best option is to combine the two models in some way. The procedure is as follows:

1. Estimate Model_1 using OLS, and obtain the fitted value of the dependent variable.
2. Employ this fitted value as an additional independent variable in the estimation of Model_2.
3. Examine the t-statistic for this fitted value. If significant, then Model_1 contains information not included in Model_2. If insignificant, then Model_1 adds nothing in terms of explanatory power.
4. Repeat steps 1) through 3), this time using fitted values of Model_2 in the estimation of Model_1.

Results of J-test

	Fitted value of Model_1 in Model_2	
Fitted value of Model_2 in Model_1	p-value<0.10	p-value>0.10
p-value<0.10	Combined model appropriate	Model_2 appropriate
p-value>0.10	Model_1 appropriate	Either model appropriate

Formatting of Tables. Published tables have no vertical lines, and only three horizontal lines: one above and one below the column headings, and one at the base of the table. Notes go below the table, the title goes above. The fastest way to produce a table is output the basic structure from R as a csv format file. If working on beast, use FTP to bring the csv file to your PC. Open the csv file in EXCEL, and set up the table as you would like it: set the decimal places, the font, etc., but don't insert the lines. Copy the table from EXCEL, and paste it into a MSWord document, using "paste special," "RTF format." In MSWord, highlight your table, click on the DESIGN tab on the menu bar, and adjust the style there to put in the lines you need. Then right-click on the highlighted table, and click on AUTOFIT, fitting to contents. Here's an example of how a table should look:

TABLE 1. DENSITY OF SIX DATA MATRICES

Data Matrix	Density
Journey to Work 1980	22.8%
Journey to Work 1990	25.7%
Migration 1990, age 25-69	99.9%
Migration 1990, college+ 1990 age 25-59	99.2%
Goods 1993 weight	76.0%
Goods 1993 value	86.5%

Notes: Density is the number of interregional flows, divided by the number of pairs of nodes (89*88).

=======================Homework=======================

We'll use data from the 1997 Census of manufactures to estimate an aggregate production function. Each observation in the data is a county, output is value-added, and three different inputs are considered: capital, production labor hours, and number of non-production employees. A production worker is one who is directly engaged in the fabrication of a product and who is paid an hourly wage. A non-production worker is typically employed in technical, purchasing, marketing or transportation positions in a manufacturing plant, and is paid an annual salary. Capital is the summed payments to all non-labor inputs, and can be considered as the value of non-labor inputs.

We will specify the production function in the Cobb-Douglas functional form:

$$V_i = A_i K_i^{\alpha_K} H_i^{\alpha_H} N_i^{\alpha_N} \qquad \text{(Eq. 1)}$$

In the above equation, the data provide information for *V* (value added in manufacturing), *K* (value of capital services), *H* (production labor), and *N* (non-production labor). The *i* subscript indicates that the value for each of these variables changes for each county. The variable *A* indicates the technological level of each county: the higher the value of *A* for a county, the more output it is able to produce for a given amount of input. But we don't know the value of *A*, nor do we know the function's parameters (the exponents α_K,α_H, and α_N). The parameters are scalars—unlike the data values, they do not vary from one county to another. We can estimate the parameters by linearizing the Cobb-Douglas production function:

$$\ln(V_i) = \hat{\alpha}_0 + \hat{\alpha}_K \ln(K_i) + \hat{\alpha}_H \ln(H_i) + \hat{\alpha}_N \ln(N_i) + \varepsilon_i \qquad \text{(Eq. 2)}$$

The carats over the estimated parameters indicate that they are estimates; $\hat{\alpha}_0$ is the intercept, ε_i is the residual.

Productivity is the ratio of output over an input. Total factor productivity is the ratio of output over some function of all inputs. The technological variable in Equation 1 provides a measure of the total factor productivity of a county:

$$A_i = \frac{V_i}{K_i^{\alpha_K} H_i^{\alpha_H} N_i^{\alpha_N}} \qquad \text{(Eq. 3)}$$

We can use our regression results in Equation 2 to estimate total factor productivity:

$$A_i = \exp(\hat{\alpha}_0 + \varepsilon_i) \qquad \text{(Eq. 4)}$$

The file *manuf1997.dbf* contains observations for 1984 U.S. counties. The data are from the Census of Manufactures for 1997 and the Population Census of 1990.

state	The state FIPS number
COUNTY	The county FIPS number
NAME	Name of county

Some county-level manufacturing variables:

EST	The number of manufacturing establishments in the county in 1997
EST20	The number of manufacturing establishments with 20 or more employees
EMP	Number of manufacturing employees in the county in 1997
pem	Number of manufacturing production workers in the county in 1997
npe	Number of manufacturing *non*-production workers in the county in 1997
HRS	Hours worked by manufacturing production workers in the county in 1997
MAT	Cost of materials for manufacturers in the county in 1997 ($)
VS	Value of shipments (=total revenue) for manufacturers in the county in 1997 ($)
vla	Value-added for manufacturers in the county in 1997 ($)
kap	Expenditures on capital by manufacturers in the county in 1997 ($)
PAY	Total payroll for manufacturing employees in the county in 1997 ($)
wag	Payroll for manufacturing production workers in the county in 1997 ($)
sal	Payroll for manufacturing *non*-production workers in the county in 1997 ($)

Some county-level infrastructure-related variables:

MEDYRBLT	Median year built for housing units in the county in 1990
psewer	Percent of housing units with sewer connection in the county in 1990
pwater	Percent of housing units with connection to public water system in the county in 1990
hwypop	Number of highway lane-miles per person in the county in 1990
hwylnd	Number of highway lane-miles per acre in the county in 1990
popacr	Population per acre in the county in 1990

Some variables giving information about the specific location of each observation:

MSA — This is the MSA number. If PMSA>0 then the county is part of a Metropolitan Statistical Area. Otherwise, the county is a rural county.

CD93 — This is a county-type code, as follows:

0	Central counties of metropolitan areas of 1 million population or more
1	Fringe counties of metropolitan areas of 1 million population or more
2	Counties in metropolitan areas of 250,000 - 1,000,000 population
3	Counties in metropolitan areas of less than 250,000 population
4	Urban population of 20,000 or more, adjacent to a metropolitan area
5	Urban population of 20,000 or more, not adjacent to a metropolitan area
6	Urban population of 2,500-19,999, not adjacent to a metropolitan area
7	Urban population of 2,500-19,999, adjacent to a metropolitan area
8	Completely rural (no places with a population of 2,500 or more) adjacent to a metropolitan area
9	Completely rural (no places with a population of 2,500 or more) not adjacent to a metropolitan area

CORE — This variable is a dummy, which equals one when the county is the nodal county for its Economic Area. The nodal county is the center, biggest county, towards which surrounding counties look for big-city services. For example, Davidson county is the nodal county in the

Nashville Economic Area.

REG This is a regional code, as follows

1	New England	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
2	Mideast	Delaware, District of Columbia, Maryland, New Jersey, New York, Pennsylvania
3	Great Lakes	Illinois, Indiana, Michigan, Ohio, Wisconsin
4	Plains	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota
5	Southeast	Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia
6	Southwest	Arizona, New Mexico, Oklahoma, Texas
7	Rocky Mountain	Colorado, Idaho, Montana, Utah, Wyoming
8	Far West	Alaska, California, Hawaii, Nevada, Oregon, Washington

The following classifications were developed by the USDA Economic Research Service:

```
FM         Farming-dependent, 1989
            0 - Nonmetro other county            1,720
            1 - Nonmetro farming county            556
            8 - Metro county                       813
MI         Mining-dependent, 1989
            0 - Nonmetro other county            2,130
            1 - Nonmetro mining county             146
            8 - Metro county                       813
MF         Manufacturing-dependent, 1989
            0 - Nonmetro other county            1,770
            1 - Nonmetro manufacturing county      506
            8 - Metro county                       813
GV         Government-dependent, 1989
            0 - Nonmetro other county            2,032
            1 - Nonmetro government county         244
            8 - Metro county                       813
TS         Services-dependent, 1989
            0 - Nonmetro other county            1,953
            1 - Nonmetro Services county           323
            8 - Metro county                       813
NS         Nonspecialized, 1989
            0 - Nonmetro other county            1,792
            1 - Nonmetro nonspecialized county     484
            8 - Metro county                       813
RT         Retirement Destination, 1990
            0 - Nonmetro other county            2,086
            1 - Nonmetro retirement county         190
            8 - Metro county                       813
FL         Federal Lands, 1987
            0 - Nonmetro other county            2,006
            1 - Nonmetro Federal lands county      270
            8 - Metro county                       813
CM         Commuting, 1990
            0 - Nonmetro other county            1,895
            1 - Nonmetro commuting county          381
            8 - Metro county                       813
PV         Persistent Poverty, 1990
            0 - Nonmetro other county            1,741
            1 - Nonmetro poverty county            535
            8 - Metro county                       813
TP         Transfers-dependent, 1989
            0 - Nonmetro other county            1,895
            1 - Nonmetro transfers county          381
            8 - Metro county                       813
```

Here's your homework (due before noon, next Thursday).

1) Construct a Cobb-Douglass production function, using value-added as a measure of output, and number of non-production workers, production hours, and non-labor expenditures as inputs. Look at your regression output and, using elementary microeconomic theory, determine the output elasticities for each input.
2) Conduct a Chow test to see if Region 5 (the southeast) has different parameters from the rest of the country.
3) Calculate total factor productivity for each county. Build a model explaining TFP, using any combination of variables. You need to include a paragraph of *ex ante* discussion justifying the independent variables you select for your unrestricted model. Use a J-test to determine whether both *hwylnd* and *hwypop* should be in your model. Use an F-test to justify your final restricted model. Present your results in the format described above. Conclude with *ex post* discussion interpreting your results.
4) Use the syntax in the lines of the program headed "`Intercept and slope for capital`" to see how the intercept and the output elasticity of capital varies across the eight Census regions (REG).
5) Interpret the results from the section headed "`Returns to scale`".
6) Comment every line of the program on the next page, using the "#" symbol (this part, and only this part, is an individual assignment).

R program: S:\TEFF\662\R\r02.R

```
#--1997 Manufacturing, by County; Cobb Douglass production function--
rm(list=ls(all=TRUE))
setwd("d:/class/662/R/")
library(foreign)
library(car)

#--read in the 1997 county-level manufacturing data--
gg<-read.dbf("manuf1997.dbf")
#--use a semicolon to keep two or more commands on the same line--
dim(gg); names(gg); summary(gg); head(gg)
row.names(gg)<-gg$NAME

#--J-test to see if HRS or pem is the better independent variable--
ndf<-data.frame(Q=gg$vla,K=gg$kap,H=gg$HRS,N=gg$npe,P=gg$pem)
row.names(ndf)<-gg$NAME
bb<-log(ndf)
head(bb)
z1<-lm(Q~K+H+N,data=bb)
f1<-z1$fitted.values
z2<-lm(Q~K+P+N,data=bb)
f2<-z2$fitted.values
summary(lm(Q~K+H+N+f2,data=bb))$coefficients
summary(lm(Q~K+P+N+f1,data=bb))$coefficients

#--Chow test comparing MSA/non-MSA counties--
zz<-lm(log(vla)~log(kap)+log(HRS)+log(npe),data=gg)
ESSR<-sum(zz$residuals^2)
z<-which(gg$MSA==0);length(z)
z1<-lm(log(vla)~log(kap)+log(HRS)+log(npe),data=gg[z,])
ESS1<-sum(z1$residuals^2)
z<-which(gg$MSA>0);length(z)
z2<-lm(log(vla)~log(kap)+log(HRS)+log(npe),data=gg[z,])
ESS2<-sum(z2$residuals^2)
ESSUR<-ESS1+ESS2
dfUR<-NROW(gg)-2*zz$rank
numres<-2*zz$rank
Fstat=((ESSR-ESSUR)/numres)/(ESSUR/dfUR)
pval=1-pf(Fstat,numres,dfUR)
cbind(Fstat,pval)

#--Total Factor Productivity--
zz<-lm(log(vla)~log(kap)+log(HRS)+log(npe),data=gg)
TFP<-round(exp(zz$coefficients[1]+zz$residuals),3)
#--look at top and lowest 15 counties--
zc<-order(TFP)
TFP[zc][1:15]
TFP[zc][(length(TFP)-15):length(TFP)]

#--Build model explaining TFP--
TFP.lm<-lm(TFP~EST+EST20+I(PAY/EMP),data=gg)
summary(TFP.lm)
coefs<-names(coef(TFP.lm))
linear.hypothesis(TFP.lm,coefs[2:3])
linear.hypothesis(TFP.lm,c("EST","EST20"))

#---Cobb Douglass with dummies for CORE--
#--Intercept only--
zz<-lm(log(vla)~factor(CORE)+log(kap)+log(HRS)+log(npe),data=gg)
summary(zz)
#--Intercept and slope for capital--
zz<-lm(log(vla)~factor(CORE)*log(kap)+log(HRS)+log(npe),data=gg)
summary(zz)

#--Returns to scale--
zz<-lm(log(vla)~log(kap)+log(HRS)+log(npe),data=gg)
#--negative subscript removes an element--
sum(zz$coefficients[-1])
linear.hypothesis(zz,"log(kap)+log(HRS)+log(npe)=1")
```