ECONOMETRICS

Sponsored by a Grant TÁMOP-4.1.2-08/2/A/KMR-2009-0041

Course Material Developed by Department of Economics,

Faculty of Social Sciences, Eötvös Loránd University Budapest (ELTE)

Department of Economics, Eötvös Loránd University Budapest

Institute of Economics, Hungarian Academy of Sciences

Balassi Kiadó, Budapest

Authors: Péter Elek, Anikó Bíró

Supervised by Péter Elek

June 2010

Week 5

Multivariate regression II.

Forecasting

$$\hat{y}_0 = \hat{\alpha} + \hat{\beta}_1 x_{10} + \hat{\beta}_2 x_{20}$$

Forecast error

$$\hat{y}_0 - y_0 = \hat{\alpha} - \alpha + (\hat{\beta}_1 - \beta_1) x_{10} + (\hat{\beta}_2 - \beta_2) x_{20} - u_0$$

Variance of forecast error (k regressors)

$$\sigma^2 \left(1 + \frac{1}{n}\right) + \sum_{l=1}^{k} \sum_{m=1}^{k} (x_{l0} - \bar{x}_l)(x_{m0} - \bar{x}_m) Cov(\hat{\beta}_l, \hat{\beta}_m)$$

Estimating the variance of forecast error

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

$$\hat{y}_0 = \hat{\alpha} + \hat{\beta}_1 x_{10} + \hat{\beta}_2 x_{20}$$

$$y_0 = E(y \mid x_1 = x_{10}, x_2 = x_{20}) =$$

$$= \alpha + \beta_1 x_{10} + \beta_2 x_{20}$$

$$y = y_0 + \beta_1 (x_1 - x_{10}) + \beta_2 (x_2 - x_{20}) + u$$

Predicted *expected* value and its standard error =
= estimated constant and its standard error of auxiliary regression

Sampling distribution of the coefficient estimates

If the assumptions are satisfied (normality and homoscedasticity, as well)

$$\hat{\beta}_i \sim N\left(\beta, \frac{\sigma^2}{RSS_i}\right) \sim N\left(\beta, \frac{\sigma^2}{TSS_i \cdot \left(1 - R_i^2\right)}\right)$$

where RSS_i is the residual and TSS_i is the total sum of squares in the regression of x_i on the other explanatory variables, and R_i^2 is the coefficient of determination in the same regression (analogy with simple regression!)

Omitting relevant variables I.

Simple regression

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

Estimated model: $y = \gamma_1 x_1 + u$

$$\hat{\gamma}_1 = \frac{\sum_i x_{1i} y_i}{\sum_i x_{1i}^2} = \beta_1 + \beta_2 \frac{\sum_i x_{1i} x_{2i}}{\sum_i x_{1i}^2} + \frac{\sum_i x_{1i} u_i}{\sum_i x_{1i}^2}$$

$$E(\hat{\gamma}_1) = \beta_1 + b_{12}\beta_2$$

Bias:	Corr(x_1,x_2)>0	Corr(x_1,x_2)<0
β_2 >0	+	–
β_2 <0	–	+

Omitting relevant variables II

k explanatory variables, k_1 + 1, …, *k.* omitted

$$E(\hat{\beta}_i) = \beta_i + \sum_{j=k_1+1}^{k} b_{ji}\beta_j, \quad i = 1, \ldots, k_1$$

$$x_j = b_{j1}x_1 + \ldots + b_{jk1}x_{k1} + u$$

Omitting relevant variables, example

Wage tariff (2003)

Weak negative correlation between education and age

Partial effect of age is positive – if omitted, the estimated coefficient of education is slightly downward biased

Estimated equations

LOG(EARN) = 10.46 + 0.1547 EDUC9 + 0.0078 AGE
LOG(EARN) = 10.79 + 0.1544 EDUC9

Irrelevant variables in the regression

True model: $y = \beta_1 x_1 + \beta_2 x_2 + u$
Estimated model: $y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u,\ \ \beta_3 = 0$
Does not affect unbiasedness
Variance increases:

$$Var(\beta_i) = \frac{\sigma^2}{RSS_i}$$

RSS_i: from the regression of x_i on the other explanatory variables (additional regressor:
RSS_i decreases, except for these
are uncorrelated)

t-test

$$\hat{\sigma}^2 = \frac{RSS}{n-k-1} \sim \sigma^2 \frac{\chi^2_{n-k-1}}{n-k-1}$$

“good” estimator of the variance of error term, therefore:

$$SE(\hat{\beta}_i) = \hat{\sigma}^2 / RSS_i \qquad \frac{\hat{\beta}_i - \beta_i}{SE(\hat{\beta}_i)} \sim t_{n-k-1}$$

Two sided test: $H_0: \beta_i = 0$, $H_1: \beta_i \neq 0$

One sided test: e.g. $H_0: \beta_i = 0$, $H_1: \beta_i > 0$

Confidence interval: $\hat{\beta}_i \pm c \cdot SE(\hat{\beta}_i)$

t-test, example

Credit approval, testing discrimination on settlement level:

$approval\_rate_i = \alpha + \beta_1 minority\_rate_i + \beta_2 avg\_inc_i + + \beta_3 avg\_wealth_i + u_i$, i = 1…n

No difference according to minority ratio:

H_0: $\beta_1 = 0$

Negative discrimination against minorities:

H_1: $\beta_1 < 0$

Example, testing significance of a regressor

Do more experienced earn more, given education? (Wage tariff, 2003)

$\log(Earn_i) = \alpha + \beta_1 \text{Educ} + \beta_2 \text{Exp}_i + u_i$

$H_0: \beta_2 = 0 \qquad H_1: \beta_2 > 0$

Dependent Variable: LOG(EARN)
Method: Least Squares
Included observations: 201971

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.556	0.004	2630.523	0.0000
EDUC9	0.164	0.001	320.482	0.0000
EXP	0.008	9.45E-05	**79.859**	0.0000

Example: relationship between earnings and years of education

$\log(Earn_i) = \alpha + \beta_1 \text{Educ\_y}_i + u_i$, Wage tariff 2003

(Univariate case: F-test is the square of t-test)

Dependent Variable: LOG(KER)
Method: Least Squares
Date: 02/17/09 Time: 16:15
Sample: 1 201971
Included observations: 201971

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.11927	0.005027	2013.122	0.0000
ISKEV	0.121868	0.000409	297.8910	0.0000

R-squared	0.305251	Mean dependent var	11.58047
Adjusted R-squared	0.305248	S.D. dependent var	0.592283
S.E. of regression	0.493679	Akaike info criterion	1.426146
Sum squared resid	49223.58	Schwarz criterion	1.426247
Log likelihood	-144018.0	Hannan-Quinn criter.	1.426175
F-statistic	88739.06	Durbin-Watson stat	0.888245
Prob(F-statistic)	0.000000		

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	88739.06	(1, 201969)	0.0000
Chi-square	88739.06	1	0.0000

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2)	0.121868	0.000409

Restrictions are linear in coefficients.

Analysis of variancie

Is the regression model useable?

Source of std. Dev.	Sum of squares	Degrees of freedom	Mean sum of squares	F
Explained (ESS)	$R^2 S_{yy}$	k	$R^2 S_{yy}/k = MS_1$	$F = MS_1/MS_2$
Residual (RSS)	$(1 – R^2)S_{yy}$	n – k – 1	$(1 – R^2)S_{yy}/(n – k – 1) =$ $= MS_2$	
Total (TSS)	S_{yy}	n – 1		

F-test of the usability of the regression

H_0: $\beta_i = 0$ (i = 1,…,k)
If H_0 satisfied

TSS ~ $\sigma^2 \, Chi_{n-1}^2$
RSS ~ $\sigma^2 \, Chi_{n-k-1}^2$, ESS ~ $\sigma^2 \, Chi_k^2$ independent

Therefore:

$$F = \frac{ESS/k}{RSS/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)} \sim F_{k,n-k-1}$$

So we reject H_0 if F > critical value of $F_{k,n-k-1}$ distribution

Seminar
Multivariate regression II

Practicing

- Maddala: 4/1, 4/3, 4/4, 4/5, 4/6, 4/9, 4/10
- Wooldridge: 4.12, 4.14, 4.17, 4.19, 6.15

Discussion

- t- and F-tests
- Forecasting with EViews

Data

- Subsample of Wage tariff (see week 4)
- Wooldridge housing price data (hprice.dta)