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 4

Multivariate regression: estimation, ant its properties

Basics
Estimation
OLS asymptotics
Coefficient interpretation
Forecasting
t-test, F-test

Introduction

Multiple explanatory variables

$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i,$

$i = 1 \ldots n$

Example

$\log(Wage)_i = \alpha + \beta_1 Educ_i + \beta_2 Experience_i + u_i,\ i = 1 \ldots n$

Assumptions

1. $E(u_i) = 0$
2. $V(u_i) = \sigma^2$ for all i
3. u_i, u_j independent for all $i \neq j$
4. x_i, u_j independent for all i, j
5. u_i normally distributed
6. No perfect collinearity (none of the regressors can be expressed as a linear function of the other regressors)

Endogeneity

Key: exogenous explanatory variables:

$E(u| x_1, x_2, \ldots, x_k) = 0$

(from assumptions 1 and 4)

Endogenous explanatory variable, if:

$E(u| x_j) \neq 0$

E.g. omitted explanatory variable which is correlated with x_j – biasedness

Perfect collinearity

Linear functional relationship among the regressor (assumption 6 is not satisfied)

Example: $Grade_i = \alpha + \beta_1 Learn_i + \beta_2 Rest_i + \quad + \beta_3 Other_i + u_i$, $i = 1 \ldots n$

Learn + Rest + Other = 168

Other = (168 – Learn – Rest)

Estimaton, two regressors

3 normal equations (method of moments)

E(u) = 0 $$\sum \hat{u}_i = 0$$

cov(u,x_1) = 0 $$\sum x_{1i}\hat{u}_i = 0$$

cov(u,x_2) = 0 $$\sum x_{2i}\hat{u}_i = 0$$

$$\hat{u}_i = y_i - \hat{\alpha} - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}$$

Estimaton, two regressors

Or: method of optimal least squares

$$\min_{\hat{\alpha},\hat{\beta}_1,\hat{\beta}_2} Q = \sum (y_i - \hat{\alpha} - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})^2$$

3 normal equations (same as before)

$$\frac{\partial Q}{\partial \hat{\alpha}} = 0 \Rightarrow \sum_i 2(y_i - \hat{\alpha} - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})(-1) = 0$$

$$\frac{\partial Q}{\partial \hat{\beta}_1} = 0 \Rightarrow \sum_i 2(y_i - \hat{\alpha} - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})(-x_{1i}) = 0$$

$$\frac{\partial Q}{\partial \hat{\beta}_2} = 0 \Rightarrow \sum_i 2(y_i - \hat{\alpha} - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})(-x_{2i}) = 0$$

Estimation, more regressors

$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i,\ i = 1 \ldots n$

$k + 1$ unknowns, $k + 1$ normal equations

Residual sum of squares: $RSS = S_{yy}(1 - R^2)$

R^2: multiple coefficient of determination

Estimation, matrix

Model: $\boldsymbol{y = X\beta + u}$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

$\boldsymbol{y}$: $n \times 1$, $\boldsymbol{X}$: $n \times k$, $\boldsymbol{\beta}$: $k \times 1$, $\boldsymbol{u}$: $n \times 1$

Estimation, matrix cont.

min $Q = \boldsymbol{u'u} = (\boldsymbol{y} - \boldsymbol{X\beta})' (\boldsymbol{y} - \boldsymbol{X\beta})$

$Q = \boldsymbol{y'y} + \boldsymbol{\beta' X' X \beta} - 2\,\boldsymbol{\beta' X' y}$

$-2\boldsymbol{X'y} + 2\boldsymbol{X'X\beta} = 0$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y}$$

Analogy with univariate regression!

OLS asymptotics, univariate

Usual assumptions, but homoscedasticity and normality not needed:

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

$$\text{plim}\hat{\beta} = \frac{Cov(x,y)}{Var(x)} = \frac{Cov(x, \alpha + \beta x + u)}{Var(x)} =$$

$$= \beta + \frac{Cov(x,u)}{Var(x)} = \beta$$

Gauss–Markov-theorem

OLS is best linear unbiased estimator (BLUE)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y}) \text{ linear}$$

$$\text{E}(\hat{\boldsymbol{\beta}}) = \text{E}[(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}))] =$$

$$= \boldsymbol{\beta} + E[(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{u})] = \boldsymbol{\beta} \text{ unbiased}$$

$$\text{Var}(\hat{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' =$$

$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$$

Gauss–Markov, minimal variance

Alternative unbiased linear estimator

$$\boldsymbol{\beta}^* = \hat{\boldsymbol{\beta}} + \mathbf{Cy} = \boldsymbol{\beta} + \mathbf{CX}\boldsymbol{\beta} + [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{C}]\mathbf{u}$$

$$\mathrm{E}(\boldsymbol{\beta}^*) = \boldsymbol{\beta} + \mathbf{CX}\boldsymbol{\beta}, \text{ unbiased} : \mathbf{CX} = \mathbf{0}$$

$$\mathrm{Var}(\boldsymbol{\beta}^*) = \mathrm{E}(\boldsymbol{\beta}^* - \boldsymbol{\beta})(\boldsymbol{\beta}^* - \boldsymbol{\beta})' =$$

$$= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{C}]\mathrm{E}(\mathbf{uu}')[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{C}]' =$$

$$= (\mathbf{X}'\mathbf{X})^{-1}\sigma^2 + (\mathbf{C}'\mathbf{C})^{-1}\sigma^2 \geq \mathrm{Var}(\hat{\boldsymbol{\beta}})$$

Interpretation of coefficients

Partial effect (ceteris paribus)

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$$

$$\Delta\hat{y}_i = \hat{\beta}_1 \Delta x_{1i} + \hat{\beta}_2 \Delta x_{2i}$$

Filtering a regressor (filtering the effect of x_2)

$$y_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{u}_i$$

$$x_{1i} = \hat{\gamma} + \hat{\delta} x_{2i} + \hat{v}_i$$

$$y_i = \hat{\varphi} + \hat{\theta}\hat{v}_i + \hat{w}_i \qquad \hat{\theta} = \hat{\beta}_1$$

Example, estimation 1

Wage tariff 2003, simple regression

$\log(Earn_i) = \alpha + \beta_1 Educ_i + u_i$

Dependent Variable: LOG(Earn)
Method: Least Squares
Sample: 1 201971

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.788	0.0028	3837.18	0.000
Educ9	0.155	0.0005	305.66	0.000
R-squared	0.316			
Adjusted R-squared	0.316			

Example, estimation 2

Wage tariff 2003, two regressors

$\log(Earn_i) = \alpha + \beta_1 Educ_i + \beta_2 Exp_i + u_i$

Dependent Variable: LOG(Earn)
Method: Least Squares
Sample: 1 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
R-squared	0.337215			
Adjusted R-squared	0.337208			

Example, filtering a regressor

$$\text{Estimated equations}: \log(Earn_i) = 10.556 + 0.164 \cdot Educ_i + 0.008 \cdot Exp_i$$
$$Educ_i = 6.158 - 0.042 \cdot Exp_i \Rightarrow \text{Resid}_\text{i}$$

Dependent Variable: LOG(Earn)
Method: Least Squares
Sample: 1 201971

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.580	0.00107	10791.97	0.0000
RESID	**0.164**	0.00051	320.4442	0.0000
R-squared	0.337053			
Adjusted R-squared	0.337050			

Seminar
Multivariate regression: estimation, and its properties

Practicing examples: Wooldridge: 3.3, 3.7, 3.9, 3.11, 3.13, 3.14, 3.17

Discussion

- Importance of including more regressors in a model (exogeneity)
- Assumptions needed for unbiasedness and efficiency of OLS
- Interpretation of coefficients

Data

- Earnings model based on a subsample of Wage tariff
- Eurostat data: relationships among number of patents, R&D expenditures, number of researchers