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 7

Summary of estimation methods and large sample theory

Regression model

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

Assumptions

1. $E(u_i) = 0$
2. u_i, u_j independent for all $i \neq j$
3. x_i, u_j independent for all i, j (exogeneity)
4. No perfect collinearity
5. $\mathrm{Var}(u_i) = \sigma^2$ for all i
6. u_i has normal distribution

1–5.: Gauss–Markov conditions
1–6.: Conditions of classical linear model

Assumptions differently (for large sample theory – stochastic explanatory variables)

1. Population model: $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$.
2. $\{(x_{1i}, x_{2i}, \ldots, x_{ki}, y_i),\ i = 1, \ldots, n\}$ random independent sample of the model.
3. None of the regressors is constant, no perfect collinearity among the regressors.
4. Exogeneity: $E(u|x_1, \ldots, x_k) = 0$
5. Homoscedasticity: $Var(u|x_1, \ldots, x_k) = \sigma^2$
6. u independent of the regressors, normally distributed.

1–5.: Gauss–Markov conditions
1–6.: Conditions of classical linear model

Multivariate regression model

Estimation: method of moments or OLS (also ML estimation if error term is normal)

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

Matrix

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

Simple regression

$$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$$
$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y}$$
$$S_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2$$

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

$$\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}$$

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} \cdot x_i$$

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\alpha} - \hat{\beta} \cdot x_i$$

Interpretation of multivariate model

Interpretation of coefficients

Partial effect ("ceteris paribus"): effect of a given regressor on the dependent variable, holding the other regressors fixed

Coefficient of determination: R^2

$$RSS = S_{yy}(1 - R^2)$$

Small sample properties of estimation

If assumptions 1–4 hold: OLS unbiased

If assumptions 1–5 (Gauss-Markov) hold: the estimation is BLUE, and the common formula of variance is correct:

$$Var(\hat{\beta}_i) = \frac{\sigma^2}{RSS_i}$$

If assumptions 1–6 (classical linear model) hold:
the t- and F-statistic have t- and F-distribution, respectively (any sample size).

Multivariate regression, t-test

Two sided test: pl. H_0: $\beta_i = 0$, H_1: $\beta_i \neq 0$
One sided test: pl. H_0: $\beta_i = 0$, H_1: $\beta_i > 0$

In case of normal error term: $$\frac{\hat{\beta}_i - \beta_i}{SE(\hat{\beta}_i)} \sim t_{n-k-1}$$

$$Var(\hat{\beta}_i) = \frac{\sigma^2}{RSS_i} \qquad \hat{\sigma}^2 = \frac{RSS}{n-k-1}$$

Simple regression

$$\frac{\hat{\beta} - \beta}{\sqrt{\hat{\sigma}^2 / S_{xx}}} \sim t_{n-2}$$

$$\frac{\hat{\alpha} - \alpha}{\sqrt{\hat{\sigma}^2 \left(1/n + \bar{x}^2 / S_{xx}\right)}} \sim t_{n-2}$$

Multivariate regression, F-test

Testing nested hypotheses
Testing multiple restrictions

$$F = \frac{(RRSS - URSS)/r}{URSS/(n-k-1)} = \frac{(R_U^{\,2} - R_R^{\,2})/r}{(1-R_U^{\,2})/(n-k-1)} \sim F_{r,(n-k-1)}$$

$$RRSS = S_{yy}(1-R_R^{\,2}) \quad URSS = S_{yy}(1-R_U^{\,2})$$

Regression cannot be used:

$$\mathrm{H}_0 : \beta_1 = \beta_2 = \ldots \beta_k = 0$$

$$R_R^{\,2} = 0$$

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

Analysis of variance

Source of variance	Sum of squares	Degree of freedom	Mean sum of squares	F
Regr.	R^2S_{yy}	k	$R^2S_{yy}/k = MS_1$	F = $= MS_1/MS_2$
Residual	$(1 – R^2)S_{yy}$	n – k – 1	$(1 – R^2)S_{yy}/(n – k – 1) =$ $= MS_2$	
Total	S_{yy}	n – 1		

Large sample properties I: consistency

If assumptions 1–4 hold: OLS is consistent. Proof for simple regression

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

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

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

Large sample properties II: asymptotic normality

If assumptions 1–5 (Gauss–Markov) hold: OLS estimator is asymptotically normal:

$$\sqrt{n}\left(\hat{\beta}_i - \beta_i\right) \sim_{asympt} N\left(0, c\right)$$

$$c = n \cdot Var(\hat{\beta}_i) = n \cdot \frac{\sigma^2}{TSS_i(1 - R_i^2)} = \frac{\sigma^2}{\sigma_{x_i}^2 (1 - R_i^2)}$$

Thus the standard deviation goes to zero in order $n^{1/2}$.
The common estimator of σ^2 is consistent, therefore the common t-test is asymptotically valid (even if assumption 6 (normality) does not hold)!

Large sample properties III: F-test and others

If assumptions 1-5 hold (assumption 6 (normality) not needed): F-test is asymptotically valid.

Other large sample tests (only asymptotically valid):

Wald-test: n(RRSS-URSS)/URSS ~ χ_r^2
    regression cannot be used: $nR^2/(1-R^2) \sim \chi_k^2$

Lagrange-multiplicator (LM) test: n(RRSS-URSS)/RRSS ~ χ_r^2
    regression cannot be used: $nR^2 \sim \chi_k^2$

Model selection

Adjusted R^2

$$1-\bar{R}^2 = \frac{n-1}{n-k-1}(1-R^2)$$

Nested hypotheses: t- and F-test

Non-nested hypotheses, same dependent variable: adjusted R^2, information criteria (AIC, BIC – based on log-likelihood)

Omitting relevant variables

If omitted variable is correlated with included regressors: biased estimation (endogeneity)

Simple regression
    True model: $y = \beta_1 x_1 + \beta_2 x_2 + u$
    Estimated model: $y = \gamma_1 x_1 + u$

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

Including irrelevant variables

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 (no endogeneity)

Variance increases

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

Other topics

Forecasting
Outliers
Alternative functional forms
Tests of stability
Dummy regressors
Quadratic terms, interactions
Heteroscedasticity, etc.

Seminar
First exam