Econometrics 2 - Lecture 1

ML Estimation, Diagnostic Tests

Contents

The Linear Model

Y: explained variable
X: explanatory or regressor variable
The model describes the data-generating process of *Y* under the condition *X*

A simple linear regression model

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

β: coefficient of *X*
α: intercept

A multiple linear regression model

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_K X_K$$

Fitting a Model to Data

Choice of values b_1, b_2 for model parameters β_1, β_2 of $Y = \beta_1 + \beta_2 X$, given the observations (y_i, x_i), $i = 1,\ldots,N$

Fitted values: $\hat{y}_i = b_1 + b_2 x_i$, $i = 1,\ldots,N$

Principle of (Ordinary) Least Squares gives the OLS estimators

$$b_i = \arg\min_{\beta 1, \beta 2} S(\beta_1, \beta_2),\ i=1,2$$

Objective function: sum of the squared deviations

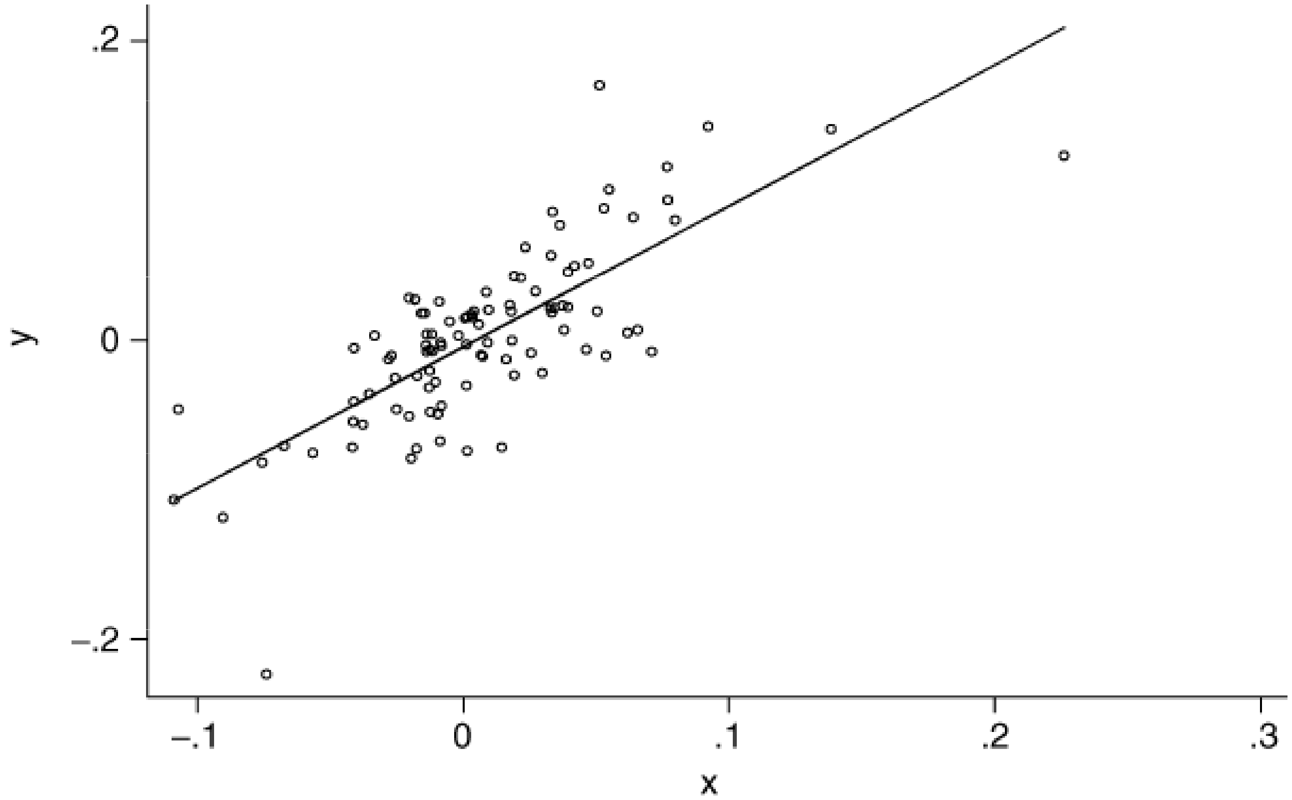
$$S(\beta_1, \beta_2) = \Sigma_i [y_i - \hat{y}_i]^2 = \Sigma_i [y_i - (\beta_1 + \beta_2 x_i)]^2 = \Sigma_i e_i^2$$

Deviations between observation and fitted values, residuals:

$$e_i = y_i - \hat{y}_i = y_i - (\beta_1 + \beta_2 x_i)$$

Observations and Fitted Regression Line

Simple linear regression: Fitted line and observation points (Verbeek, Figure 2.1)



Contents

- Linear Regression: A Review
- Estimation of Regression Parameters
- Estimation Concepts
- ML Estimator: Idea and Illustrations
- ML Estimator: Notation and Properties
- ML Estimator: Two Examples
- Asymptotic Tests
- Some Diagnostic Tests
- Quasi-maximum Likelihood Estimator

OLS Estimators

Equating the partial derivatives of S(β$_1$, β$_2$) to zero: normal equations

$$b_1 + b_2 \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i$$

$$b_1 \sum_{i=1}^{N} x_i + b_2 \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i y_i$$

OLS estimators b_1 und b_2 result in

$$b_2 = \frac{s_{xy}}{s_x^2}$$

$$b_1 = \bar{y} - b_2 \bar{x}$$

with mean values $\bar{x}$ and

and second moments

$$s_{xy} = \frac{1}{N} \sum_i (x_i - \bar{x})(y_i - \bar{y})$$

$$s_x^2 = \frac{1}{N} \sum_i (x_i - \bar{x})^2$$

OLS Estimators: The General Case

Model for Y contains K-1 explanatory variables

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_K X_K = x'\beta$$

with $x = (1, X_2, \ldots, X_K)'$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_K)'$

Observations: $[y_i, x_i] = [y_i, (1, x_{i2}, \ldots, x_{iK})']$, $i = 1, \ldots, N$

OLS-estimates $b = (b_1, b_2, \ldots, b_K)'$ are obtained by minimizing

$$S(\beta) = \sum_{i=1}^{N} (y_i - x_i'\beta)^2$$

this results in the OLS estimators

$$b = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i y_i$$

Matrix Notation

N observations

$(y_1,x_1), \ldots, (y_N,x_N)$

Model: $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, $i = 1, \ldots, N$, or

$y = X\beta + \varepsilon$

with

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix},\ X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

OLS estimators

$b = (X'X)^{-1}X'y$

Gauss-Markov Assumptions

Observation y_i (i = 1, …, N) is a linear function

$$y_i = x_i'\beta + \varepsilon_i$$

of observations x_{ik}, k =1, …, K, of the regressor variables and the error term ε_i

$x_i = (x_{i1}, \ldots, x_{iK})'$; $X = (x_{ik})$

A1	$E\{\varepsilon_i\} = 0$ for all i
A2	all ε_i are independent of all x_i (exogenous x_i)
A3	$V\{\varepsilon_i\} = \sigma^2$ for all i (homoskedasticity)
A4	$Cov\{\varepsilon_i, \varepsilon_j\} = 0$ for all i and j with $i \neq j$ (no autocorrelation)

Normality of Error Terms

A5	ε_i normally distributed for all i

Together with assumptions (A1), (A3), and (A4), (A5) implies

$$\varepsilon_i \sim NID(0,\sigma^2) \text{ for all } i$$

i.e., all ε_i are

- independent drawings
- from the normal distribution $N(0,\sigma^2)$
- with mean 0
- and variance σ^2

Error terms are “normally and independently distributed” (NID, n.i.d.)

Properties of OLS Estimators

OLS estimator $b = (X'X)^{-1}X'y$

1. The OLS estimator b is unbiased: $E\{b\} = \beta$
2. The variance of the OLS estimator is given by

   $$V\{b\} = \sigma^2(\Sigma_i x_i x_i')^{-1}$$

3. The OLS estimator b is a BLUE (best linear unbiased estimator) for β
4. The OLS estimator b is normally distributed with mean β and covariance matrix $V\{b\} = \sigma^2(\Sigma_i x_i x_i')^{-1}$

Properties

- 1., 2., and 3. follow from Gauss-Markov assumptions
- 4. needs in addition the normality assumption (A5)

Distribution of *t*-statistic

t-statistic

$$t_k = \frac{b_k}{se(b_k)}$$

follows

1. the *t*-distribution with *N*-*K* d.f. if the Gauss-Markov assumptions (A1) - (A4) and the normality assumption (A5) hold
2. approximately the *t*-distribution with *N*-*K* d.f. if the Gauss-Markov assumptions (A1) - (A4) hold but not the normality assumption (A5)
3. asymptotically ($N \rightarrow \infty$) the standard normal distribution N(0,1)
4. approximately the standard normal distribution N(0,1)

The approximation errors decrease with increasing sample size *N*

OLS Estimators: Consistency

The OLS estimators b are consistent,

$$\text{plim}_{N \to \infty} b = \beta,$$

- if (A2) from the Gauss-Markov assumptions and the assumption (A6) are fulfilled
- if the assumptions (A7) and (A6) are fulfilled

Assumptions (A6) and (A7):

A6	$1/N\ \Sigma^N_{i=1} x_i x_i'$ converges with growing N to a finite, nonsingular matrix Σ_{xx}
A7	The error terms have zero mean and are uncorrelated with each of the regressors: $E\{x_i \varepsilon_i\} = 0$

Contents

Estimation Concepts

OLS estimator: minimization of objective function S(β) gives

- K first-order conditions $\Sigma_i (y_i - x_i'b) x_i = \Sigma_i e_i x_i = 0$, the normal equations
- Moment conditions

$$E\{(y_i - x_i'\beta) x_i\} = E\{\varepsilon_i x_i\} = 0$$

- OLS estimators are solution of the normal equations

IV estimator: Model allows derivation of moment conditions

$$E\{(y_i - x_i'\beta) z_i\} = E\{\varepsilon_i z_i\} = 0$$

which are functions of

- observable variables y_i, x_i, instrument variables z_i, and unknown parameters β
- Moment conditions are used for deriving IV estimators
- OLS estimators are special case of IV estimators

Estimation Concepts, cont'd

GMM estimator: generalization of the moment conditions

$$E\{f(w_i, z_i, \beta)\} = 0$$

- with observable variables w_i, instrument variables z_i, and unknown parameters β
- Allows for non-linear models
- Under weak regularity conditions, the GMM estimators are
  - consistent
  - asymptotically normal

Maximum likelihood estimation

- Basis is the distribution of y_i conditional on regressors x_i
- Depends on unknown parameters β
- The estimates of the parameters β are chosen so that the distribution corresponds as well as possible to the observations y_i and x_i

Contents

Example: Urn Experiment

Urn experiment:

- The urn contains red and yellow balls
- Proportion of red balls: p (unknown)
- N random draws
- Random draw i: $y_i = 1$ if ball i is red, 0 otherwise; $P\{y_i = 1\} = p$
- Sample: N_1 red balls, $N-N_1$ yellow balls
- Probability for this result:

$$P\{N_1 \text{ red balls, } N-N_1 \text{ yellow balls}\} = p^{N1} (1 - p)^{N-N1}$$

Likelihood function: the probability of the sample result, interpreted as a function of the unknown parameter p

Urn Experiment: Likelihood Function

Likelihood function: the probability of the sample result, interpreted as a function of the unknown parameter *p*

$$L(p) = p^{N1} (1 - p)^{N-N1}$$

Maximum likelihood estimator: that value $\hat{p}$ of *p* which maximizes L(*p*)

$$\hat{p} = \arg\max_p L(p)$$

Calculation of $\hat{p}$: maximization algorithms

- As the log-function is monotonous, extremes of L(*p*) and log L(*p*) coincide
- Use of log-likelihood function is often more convenient

$$\log L(p) = N_1 \log p + (N - N_1) \log (1 - p)$$

Urn Experiment: Likelihood Function, cont'd

Verbeek, Fig.6.1

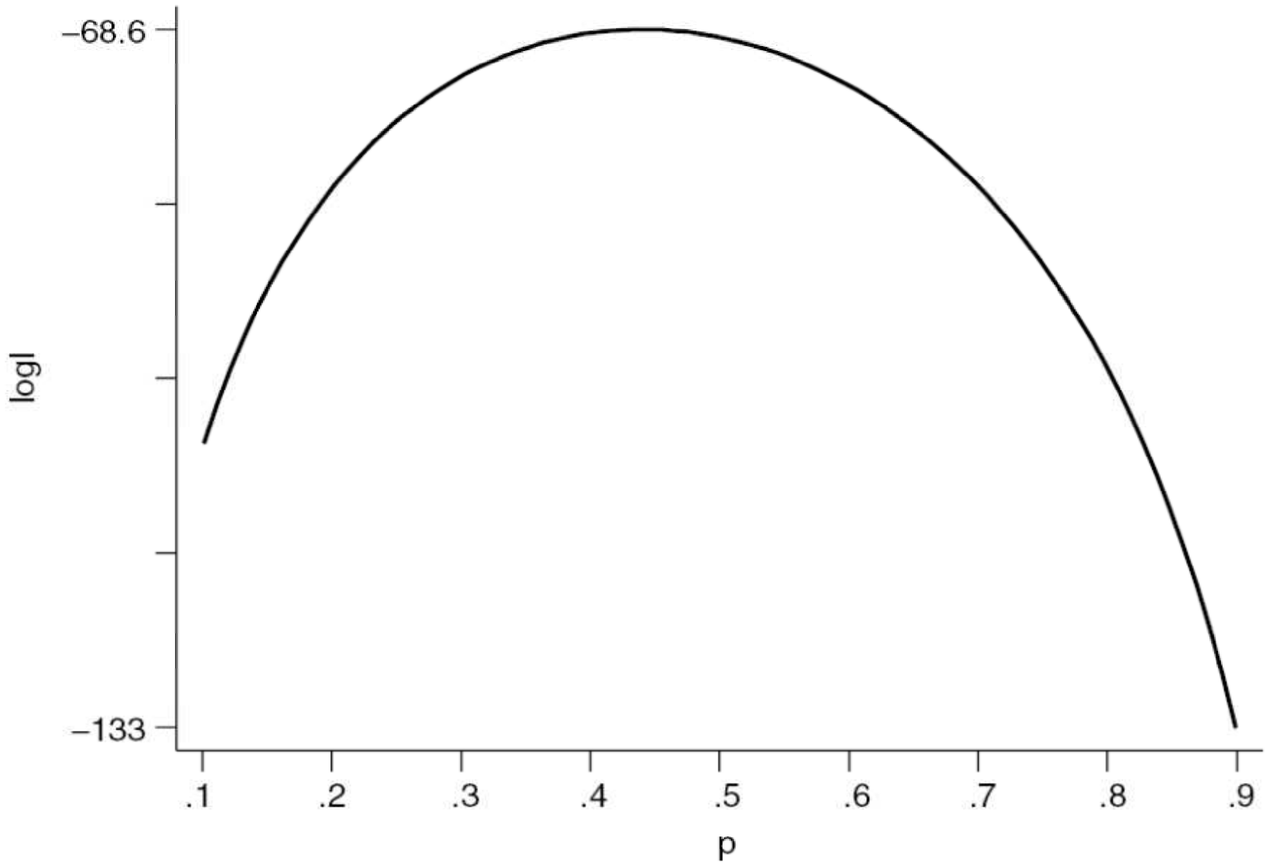


Figure 6.1 Sample loglikelihood function for $N = 100$ and $N_1 = 44$

Urn Experiment: ML Estimator

Maximizing log L*(p)* with respect to *p* gives the first-order condition

$$\frac{d \log L(p)}{dp} = \frac{N_1}{p} - \frac{N - N_1}{1 - p} = 0$$

Solving this equation for *p* gives the maximum likelihood estimator (ML estimator)

$$\hat{p} = \frac{N_1}{N}$$

For N = 100, N_1 = 44, the ML estimator for the proportion of red balls is $\hat{p}$ = 0.44

Maximum Likelihood Estimator: The Idea

- Specify the distribution of the data (of *y* or *y* given *x*)
- Determine the likelihood of observing the available sample as a function of the unknown parameters
- Choose as ML estimates those values for the unknown parameters that give the highest likelihood
- In general, this leads to
  - consistent
  - asymptotically normal
  - efficient estimators

  provided the likelihood function is correctly specified, i.e., distributional assumptions are correct

Example: Normal Linear Regression

Model

$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$

with assumptions (A1) – (A5)

From the normal distribution of ε_i follows: contribution of observation *i* to the likelihood function:

$$f(y_i | x_i; \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2} \frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\sigma^2} \right\}$$

due to independent observations, the log-likelihood function is given by

$$\log L(\beta, \sigma^2) = \log \prod_i f(y_i | x_i; \beta, \sigma^2)$$

$$= -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_i \frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\sigma^2}$$

Normal Linear Regression, cont'd

Maximizing log L w.r.t. β and σ^2 gives the ML estimators

$$\hat{\beta}_2 = Cov\{y, x)\} / V\{x\}$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

which coincide with the OLS estimators, and

$$\hat{\sigma}^2 = \frac{1}{N} \sum_i e_i^2$$

which is biased and underestimates σ^2!

Remarks:

- The results are obtained assuming normally and independently distributed (NID) error terms
- ML estimators are consistent but not necessarily unbiased; see the properties of ML estimators below

Contents

ML Estimator: Notation

Let the density (or probability mass function) of y_i, given x_i, be given by $f(y_i|x_i,\theta)$ with K-dimensional vector θ of unknown parameters

Given independent observations, the likelihood function for the sample of size N is

$$L(\theta \mid y, X) = \prod_i L_i(\theta \mid y_i, x_i) = \prod_i f(y_i \mid x_i; \theta)$$

The ML estimators are the solutions of

$$\max_\theta \log L(\theta) = \max_\theta \Sigma_i \log L_i(\theta)$$

or the solutions of the first-order conditions

$$s(\hat{\theta}) = \frac{\partial \log L(\theta)}{\partial \theta}\Big|_{\hat{\theta}} = \sum_i \frac{\partial \log L_i(\theta)}{\partial \theta}\Big|_{\hat{\theta}} = 0$$

$s(\theta) = \Sigma_i\, s_i(\theta)$, the vector of gradients, also denoted as *score vector*

Solution of $s(\theta) = 0$

- analytically (see examples above) or
- by use of numerical optimization algorithms

Matrix Derivatives

The scalar-valued function

$$L(\theta \mid y, X) = \prod_i L_i(\theta \mid y_i, x_i) = L(\theta_1, \ldots, \theta_K \mid y, X)$$

or – shortly written as log L(θ) – has the *K* arguments $\theta_1, \ldots, \theta_K$

- *K*-vector of partial derivatives or gradient vector or gradient

$$\frac{\partial \log L(\theta)}{\partial \theta} = \left(\frac{\partial \log L(\theta)}{\partial \theta_1}, \ldots, \frac{\partial \log L(\theta)}{\partial \theta_K} \right)'$$

- *K*x*K* matrix of second derivatives or Hessian matrix

$$\frac{\partial^2 \log L(\theta)}{\partial \theta \, \partial \theta'} = \begin{pmatrix} \frac{\partial^2 \log L(\theta)}{\partial \theta_1 \partial \theta_1} & \cdots & \frac{\partial^2 \log L(\theta)}{\partial \theta_1 \partial \theta_K} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \log L(\theta)}{\partial \theta_K \partial \theta_1} & \cdots & \frac{\partial^2 \log L(\theta)}{\partial \theta_K \partial \theta_K} \end{pmatrix}$$

ML Estimator: Properties

The ML estimator

1. is consistent
2. is asymptotically efficient
3. is asymptotically normally distributed:

$$\sqrt{N}(\hat{\theta} - \theta) \rightarrow \mathrm{N}(0, V)$$

V: asymptotic covariance matrix of $\sqrt{N}\hat{\theta}$

The Information Matrix

Information matrix $I(\theta)$

- $I(\theta)$ is the limit (for $N \to \infty$) of

$$\bar{I}(\theta) = -\frac{1}{N} E\left\{\frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'}\right\} = -\frac{1}{N}\sum_i E\left\{\frac{\partial^2 \log L_i(\theta)}{\partial\theta\partial\theta'}\right\} = \frac{1}{N}\sum_i I_i(\theta)$$

- For the asymptotic covariance matrix V can be shown: $V = I(\theta)^{-1}$
- $I(\theta)^{-1}$ is the lower bound of the asymptotic covariance matrix for any consistent, asymptotically normal estimator for θ: Cramèr-Rao lower bound

Calculation of $I_i(\theta)$ can also be based on the outer product of the score vector

$$I_i(\theta) = -E\left\{\frac{\partial^2 \log L_i(\theta)}{\partial\theta\partial\theta'}\right\} = E\left\{s_i(\theta)s_i(\theta)'\right\} = J_i(\theta)$$

for a misspecified likelihood function, $J_i(\theta)$ can deviate from $I_i(\theta)$

Covariance Matrix V: Calculation

Two ways to calculate V:

- A consistent estimate is based on the information matrix $I(\theta)$:

$$\hat{V}_H = \left(-\frac{1}{N} \sum_i \frac{\partial^2 \log L_i(\theta)}{\partial\theta \, \partial\theta'} \Big|_{\hat{\theta}} \right)^{-1} = \bar{I}(\hat{\theta})^{-1}$$

  index "H": the estimate of V is based on the Hessian matrix

- The BHHH (Berndt, Hall, Hall, Hausman) estimator

$$\hat{V}_G = \left(\frac{1}{N} \sum_i s_i(\hat{\theta}) s_i(\hat{\theta})' \right)^{-1}$$

  with score vector $s(\theta)$; index "G": the estimate of V is based on gradients
  - also called: OPG (outer product of gradient) estimator
  - $E\{s_i(\theta)\, s_i(\theta)'\}$ coincides with $I_i(\theta)$ if $f(y_i | x_i, \theta)$ is correctly specified

Contents

Urn Experiment: Once more

Likelihood contribution of the *i*-th observation

$$\log L_i(p) = y_i \log p + (1 - y_i) \log (1 - p)$$

This gives scores

$$\frac{\partial \log L_i(p)}{\partial p} = s_i(p) = \frac{y_i}{p} - \frac{1-y_i}{1-p}$$

and

$$\frac{\partial^2 \log L_i(p)}{\partial p^2} = -\frac{y_i}{p^2} - \frac{1-y_i}{(1-p)^2}$$

With $E\{y_i\} = p$, the expected value turns out to be

$$I_i(p) = E\left\{-\frac{\partial^2 \log L_i(p)}{\partial p^2}\right\} = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}$$

The asymptotic variance of the ML estimator $V = I^{-1} = p(1-p)$

Urn Experiment and Binomial Distribution

The asymptotic distribution is

$$\sqrt{N}(\hat{p} - p) \rightarrow N\big(0, p(1 - p)\big)$$

- Small sample distribution:

  $N\hat{p} \sim B(N, p)$

- Use of the approximate normal distribution for portions $\hat{p}$

  rule of thumb:

  $N\,p\,(1-p) > 9$

Example: Normal Linear Regression

Model

$y_i = x_i'\beta + \varepsilon_i$

with assumptions (A1) – (A5)

Log-likelihood function

$$\log L(\beta,\sigma^2) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i (y_i - x_i'\beta)^2$$

Score contributions:

$$s_i(\beta,\sigma^2) = \begin{pmatrix} \dfrac{\partial \log L_i(\beta,\sigma^2)}{\partial\beta} \\ \dfrac{\partial \log L_i(\beta,\sigma^2)}{\partial\sigma^2} \end{pmatrix} = \begin{pmatrix} \dfrac{y_i - x_i'\beta}{\sigma^2} x_i \\ -\dfrac{1}{2\sigma^2} + \dfrac{1}{2\sigma^4}(y_i - x_i'\beta)^2 \end{pmatrix}$$

The first-order conditions – setting both components of $\Sigma_i s_i(\beta,\sigma^2)$ to zero – give as ML estimators: the OLS estimator for β, the average squared residuals for σ²:

Normal Linear Regression, cont'd

$$\hat{\beta} = \left(\sum_i x_i x_i'\right)^{-1} \sum_i x_i y_i, \ \hat{\sigma}^2 = \frac{1}{N} \sum_i (y_i - x_i'\hat{\beta})^2$$

Asymptotic covariance matrix: Likelihood contribution of the *i*-th observation ($E\{\varepsilon_i\} = E\{\varepsilon_i^3\} = 0$, $E\{\varepsilon_i^2\} = \sigma^2$, $E\{\varepsilon_i^4\} = 3\sigma^4$)

$$I_i(\beta, \sigma^2) = E\{s_i(\beta, \sigma^2) s_i(\beta, \sigma^2)'\} = \text{diag}\left(\frac{1}{\sigma^2} x_i x_i', \frac{1}{2\sigma^4}\right)$$

gives

$$V = I(\beta, \sigma^2)^{-1} = \text{diag}\ (\sigma^2 \Sigma_{xx}^{-1}, 2\sigma^4)$$

with $\Sigma_{xx} = \lim\ (\Sigma_i x_i x_i')/N$

For finite samples: covariance matrix of ML estimators for β

$$\hat{V}(\hat{\beta}) = \hat{\sigma}^2 \left(\sum_i x_i x_i'\right)^{-1}$$

similar to OLS results

Contents

Diagnostic Tests

Diagnostic tests based on ML estimators

Test situation:

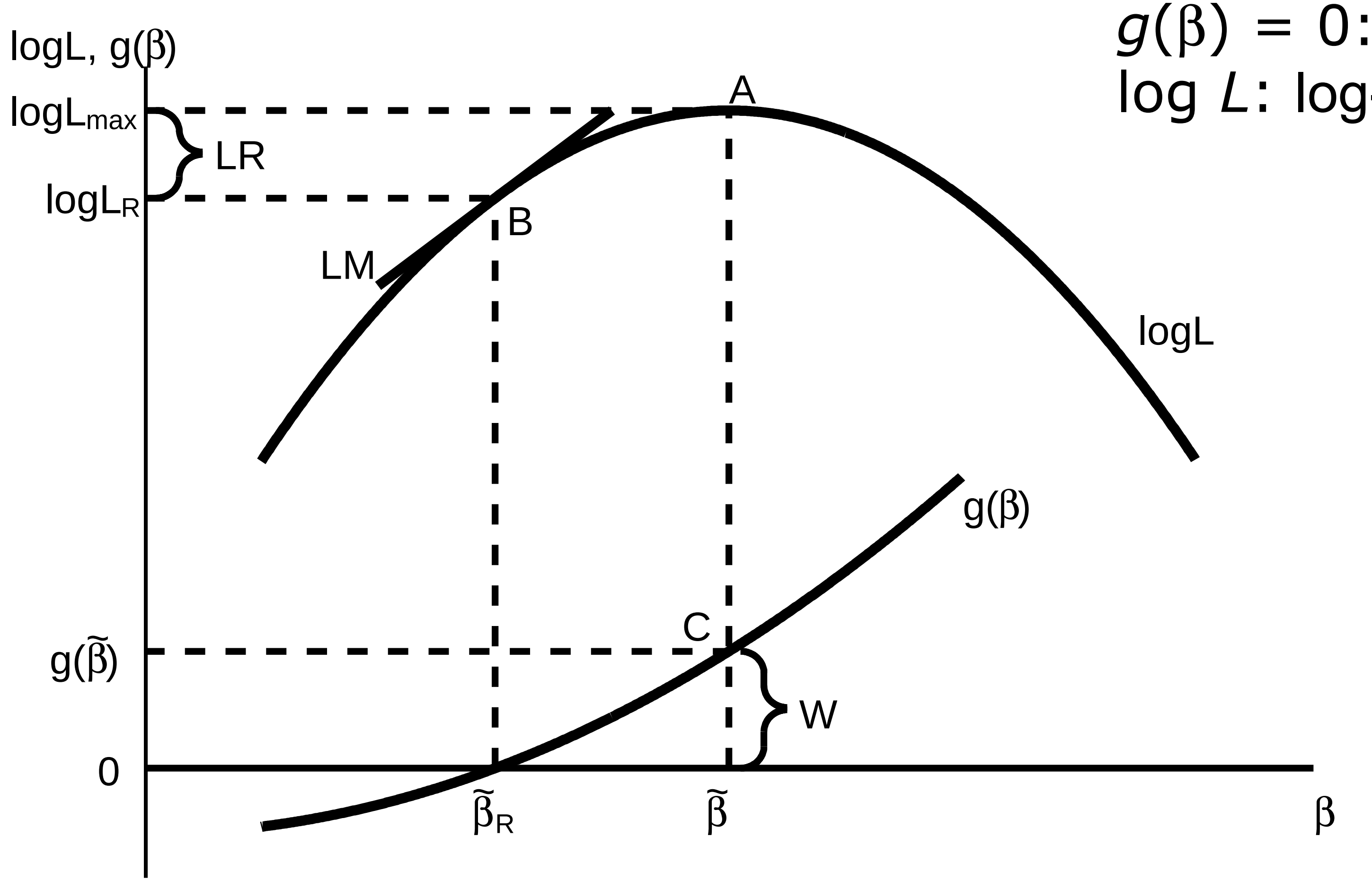
- K-dimensional parameter vector $\theta = (\theta_1, \ldots, \theta_K)'$
- $J \geq 1$ linear restrictions ($K \geq J$)
- H_0: $R\theta = q$ with JxK matrix R, full rank; J-vector q

Test principles based on the likelihood function:

1. Wald test: Checks whether the restrictions are fulfilled for the unrestricted ML estimator for θ; test statistic ξ_W
2. Likelihood ratio test: Checks whether the difference between the log-likelihood values with and without the restriction is close to zero; test statistic ξ_{LR}
3. Lagrange multiplier test (or score test): Checks whether the first-order conditions (of the unrestricted model) are violated by the restricted ML estimators; test statistic ξ_{LM}

Likelihood and Test Statistics

$g(\beta) = 0$: restriction
$\log L$: log-likelihood



The Asymptotic Tests

Under H_0, the test statistics of all three tests

- follow asymptotically, for finite sample size approximately, the Chi-square distribution with *J* df
- The tests are asymptotically (large *N*) equivalent
- Finite sample size: the values of the test statistics obey the relation

$$\xi_W \geq \xi_{LR} \geq \xi_{LM}$$

Choice of the test: criterion is computational effort

1. Wald test: Requires estimation only of the unrestricted model; e.g., testing for omitted regressors: estimate the full model, test whether the coefficients of potentially omitted regressors are different from zero
2. Lagrange multiplier test: Requires estimation only of the restricted model
3. Likelihood ratio test: Requires estimation of both the restricted and the unrestricted model

Wald Test

Checks whether the restrictions are fulfilled for the unrestricted ML estimator for θ

Asymptotic distribution of the unrestricted ML estimator:

$$\sqrt{N}(\hat{\theta}-\theta) \rightarrow N(0,V)$$

Hence, under H_0: $R\,\theta = q$,

$$\sqrt{N}(R\hat{\theta}-R\theta) = \sqrt{N}(R\hat{\theta}-q) \rightarrow N(0,RVR')$$

The test statistic

$$\xi_W = N(R\hat{\theta}-q)'\left[R\hat{V}R'\right]^{-1}(R\hat{\theta}-q)$$

- under H_0, ξ_W is expected to be close to zero
- *p*-value to be read from the Chi-square distribution with *J* df

Wald Test, cont'd

Typical application: tests of linear restrictions for regression coefficients

- Test of H_0: $\beta_i = 0$

  $$\xi_W = b_i^2/[se(b_i)^2]$$

  - ξ_W follows the Chi-square distribution with 1 df
  - ξ_W is the square of the *t*-test statistic
- Test of the null-hypothesis that a subset of *J* of the coefficients β are zeros

  $$\xi_W = (e_R'e_R - e'e)/[e'e/(N-K)]$$

  - *e*: residuals from unrestricted model
  - e_R: residuals from restricted model
  - ξ_W follows the Chi-square distribution with *J* df
  - ξ_W is related to the *F*-test statistic by $\xi_W = FJ$

Likelihood Ratio Test

Checks whether the difference between the ML estimates obtained with and without the restriction is close to zero

for nested models

- Unrestricted ML estimator: $\hat{\theta}$
- Restricted ML estimator: $\tilde{\theta}$; obtained by minimizing the log-likelihood subject to $R\,\theta = q$

Under H_0: $R\,\theta = q$, the test statistic

$$\xi_{LR} = 2\left(\log L(\hat{\theta}) - \log L(\tilde{\theta})\right)$$

- is expected to be close to zero
- *p*-value to be read from the Chi-square distribution with J df

Likelihood Ratio Test, cont'd

Test of linear restrictions for regression coefficients

- Test of the null-hypothesis that *J* linear restrictions of the coefficients β are valid

  $$\xi_{LR} = N \log(e_R' e_R / e'e)$$

  - *e*: residuals from unrestricted model
  - e_R: residuals from restricted model
  - ξ_{LR} follows the Chi-square distribution with *J* df

Lagrange Multiplier Test

Checks whether the derivative of the likelihood for the constrained ML estimator is close to zero

Based on the Lagrange constrained maximization method

Lagrangian, given $\theta = (\theta_1', \theta_2')'$ with restriction $\theta_2 = q$, *J*-vectors θ_2, q

$$H(\theta, \lambda) = \Sigma_i \log L_i(\theta) - \lambda'(\theta - q)$$

First-order conditions give the constrained ML estimators $\tilde{\theta} = (\tilde{\theta}_1', q')'$ and $\tilde{\lambda}$

$$\sum_i \frac{\partial \log L_i(\theta)}{\partial \theta_1} \Big|_{\tilde{\theta}} = \sum_i s_{i1}(\tilde{\theta}) = 0$$

$$\tilde{\lambda} = \sum_i \frac{\partial \log L_i(\theta)}{\partial \theta_2} \Big|_{\tilde{\theta}} = \sum_i s_{i2}(\tilde{\theta})$$

λ measures the extent of violation of the restriction, basis for ξ_{LM}

s_i are the scores; LM test is also called *score test*

Lagrange Multiplier Test, cont'd

Lagrange multiplier test statistic

$$\xi_{LM} = N^{-1}\tilde{\lambda}'\hat{I}^{22}(\tilde{\theta})\tilde{\lambda}$$

has under H_0 an asymptotic Chi-square distribution with J df

$\hat{I}^{22}(\tilde{\theta})$ is the block diagonal of the estimated inverted information matrix, based on the constrained estimators for θ

Calculation of ξ_{LM}

- Outer product gradient (OPG) version of the LM test:

$$\xi_{LM} = \sum_i s_i(\tilde{\theta})'\left(\sum_i s_i(\tilde{\theta})s_i(\tilde{\theta})'\right)^{-1}\sum_i s_i(\tilde{\theta}) = i'S(S'S)^{-1}S'i$$

- Auxiliary regression of a N-vector i = (1, …, 1)' on the scores $s_i(\tilde{\theta})$ with restricted estimates for θ, no intercept; $S' = [s_1(\tilde{\theta}), \ldots, s_N(\tilde{\theta})]$
- Test statistic is $\xi_{LM} = N R^2$ with the uncentered R^2 of the auxiliary regression
- Other ways for computing ξ_{LM}: see below

An Illustration

The urn experiment: test of H_0: $p = p_0$ ($J = 1$, $R = I$)

The likelihood contribution of the i-th observation is

$$\log L_i(p) = y_i \log p + (1 - y_i) \log (1 - p)$$

This gives

$$s_i(p) = y_i/p - (1-y_i)/(1-p) \text{ and } I_i(p) = [p(1-p)]^{-1}$$

Wald test:

$$\xi_W = N(\hat{p} - p_0)\left[\hat{p}(1-\hat{p})\right]^{-1}(\hat{p} - p_0) = N\frac{(\hat{p} - p_0)^2}{\hat{p}(1-\hat{p})} = N\frac{(N_1 - Np_0)^2}{N(N - N_1)}$$

Likelihood ratio test:

$$\xi_{LR} = 2\left(\log L(\hat{p}) - \log L(\tilde{p})\right)$$

with

$$\log L(\hat{p}) = N_1 \log(N_1 / N) + (N - N_1)\log(1 - N_1 / N)$$

$$\log L(\tilde{p}) = N_1 \log(p_0) + (N - N_1)\log(1 - p_0)$$

An Illustration, cont'd

Lagrange multiplier test:

with

$$\tilde{\lambda} = \sum_i s_i(p)|_{p_0} = \frac{N_1}{p_0} - \frac{N - N_1}{1 - p_0} = \frac{N_1 - Np_0}{p_0(1 - p_0)}$$

and the inverted information matrix $[I(p)]^{-1} = p(1-p)$, calculated for the restricted case, the LM test statistic is

$$\xi_{LM} = N^{-1}\tilde{\lambda}[p_0(1 - p_0)]\tilde{\lambda}$$

$$= N(\hat{p} - p_0)[p_0(1 - p_0)]^{-1}(\hat{p} - p_0)$$

Example

- In a sample of $N = 100$ balls, 44 are red
- H_0: $p_0 = 0.5$
- $\xi_W = 1.46$, $\xi_{LR} = 1.44$, $\xi_{LM} = 1.44$
- Corresponding p-values are 0.227, 0.230, and 0.230

Contents

Normal Linear Regression: Scores

Log-likelihood function

$$\log L(\beta,\sigma^2) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i (y_i - x_i'\beta)^2$$

Scores:

$$s_i(\beta,\sigma^2) = \begin{pmatrix} \dfrac{\partial \log L_i(\beta,\sigma^2)}{\partial\beta} \\ \dfrac{\partial \log L_i(\beta,\sigma^2)}{\partial\sigma^2} \end{pmatrix} = \begin{pmatrix} \dfrac{y_i - x_i'\beta}{\sigma^2} x_i \\ -\dfrac{1}{2\sigma^2} + \dfrac{1}{2\sigma^4}(y_i - x_i'\beta)^2 \end{pmatrix}$$

Covariance matrix

$$V = I(\beta,\sigma^2)^{-1} = \text{diag}(\sigma^2\Sigma_{xx}^{-1}, 2\sigma^4)$$

Testing for Omitted Regressors

Model: $y_i = x_i'\beta + z_i'\gamma + \varepsilon_i$, $\varepsilon_i \sim NID(0,\sigma^2)$

Test whether the J regressors z_i are erroneously omitted:

- Fit the restricted model
- Apply the LM test to check H_0: $\gamma = 0$

First-order conditions give the scores

$$\frac{1}{\tilde{\sigma}^2}\sum_i \tilde{\varepsilon}_i x_i = 0, \quad \frac{1}{\tilde{\sigma}^2}\sum_i \tilde{\varepsilon}_i z_i, \quad -\frac{N}{2\tilde{\sigma}^2} + \frac{1}{2}\sum_i \frac{\tilde{\varepsilon}_i^2}{\tilde{\sigma}^4} = 0$$

with constrained ML estimators for β and σ²; ML-residuals $\tilde{\varepsilon}_i = y_i - x_i'\hat{\beta}$

- Auxiliary regression of N-vector i = (1, …, 1)' on the scores $\tilde{\varepsilon}_i x_i, \tilde{\varepsilon}_i z_i$ gives the uncentered R^2
- The LM test statistic is $\xi_{LM} = N R^2$
- An asymptotically equivalent LM test statistic is NR_e^2 with R_e^2 from the regression of the ML residuals on x_i and z_i

Testing for Heteroskedasticity

Model: $y_i = x_i'\beta + \varepsilon_i$, $\varepsilon_i \sim NID$, $V\{\varepsilon_i\} = \sigma^2 h(z_i'\alpha)$, $h(.) > 0$ but unknown,
$h(0) = 1$, $\partial/\partial\alpha\{h(.)\} \neq 0$, J-vector z_i

Test for homoskedasticity: Apply the LM test to check H_0: $\alpha = 0$

First-order conditions with respect to σ² and α give the scores

$$\tilde{\varepsilon}_i^2 - \tilde{\sigma}^2, \quad (\tilde{\varepsilon}_i^2 - \tilde{\sigma}^2) z_i'$$

with constrained ML estimators for β and σ²; ML-residuals $\tilde{\varepsilon}_i$

- Auxiliary regression of N-vector $i = (1, \ldots, 1)'$ on the scores gives the uncentered R^2
- LM test statistic $\xi_{LM} = NR^2$; a version of Breusch-Pagan test
- An asymptotically equivalent version of the Breusch-Pagan test is based on NR_e^2 with R_e^2 from the regression of the squared ML residuals on z_i and an intercept
- Attention: no effect of the functional form of $h(.)$

Testing for Autocorrelation

Model: $y_t = x_t'\beta + \varepsilon_t$, $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$, $v_t \sim NID(0,\sigma^2)$

LM test of H_0: $\rho = 0$

First-order conditions give the scores

$$\tilde{\varepsilon}_t x_t', \quad \tilde{\varepsilon}_t \tilde{\varepsilon}_{t-1}$$

with constrained ML estimators for β and σ^2

- The LM test statistic is $\xi_{LM} = (T-1)\, R^2$ with R^2 from the auxiliary regression of the ML residuals on the lagged residuals; Breusch-Godfrey test
- An asymptotically equivalent version of the Breusch-Godfrey test is based on NR_e^2 with R_e^2 from the regression of the ML residuals on x_t and the lagged residuals

Contents

Quasi ML Estimator

The quasi-maximum likelihood estimator

- refers to moment conditions
- does not refer to the entire distribution
- uses the GMM concept

Derivation of the ML estimator as a GMM estimator

- weaker conditions
- consistency applies

Generalized Method of Moments (GMM)

The model is characterized by R moment conditions

$$E\{f(w_i, z_i, \theta)\} = 0$$

- $f(.)$: R-vector function
- w_i: vector of observable variables, z_i: vector of instrument variables
- θ: K-vector of unknown parameters

Substitution of the moment conditions by sample equivalents:

$$g_N(\theta) = (1/N) \Sigma_i f(w_i, z_i, \theta) = 0$$

Minimization wrt θ of the quadratic form

$$Q_N(\theta) = g_N(\theta)' W_N g_N(\theta)$$

with the symmetric, positive definite weighting matrix W_N gives the GMM estimator

$$\hat{\theta} = \arg\min_\theta Q_N(\theta)$$

Quasi-ML Estimator

The quasi-maximum likelihood estimator

- refers to moment conditions
- does not refer to the entire distribution
- uses the GMM concept

ML estimator can be interpreted as GMM estimator: first-order conditions

$$s(\hat{\theta}) = \frac{\partial \log L(\theta)}{\partial \theta}\Big|_{\hat{\theta}} = \sum_i \frac{\partial \log L_i(\theta)}{\partial \theta}\Big|_{\hat{\theta}} = \sum_i s_i(\theta)\Big|_{\hat{\theta}} = 0$$

correspond to sample averages based on theoretical moment conditions

Starting point is

$$E\{s_i(\theta)\} = 0$$

valid for the *K*-vector θ if the likelihood is correctly specified

$E\{s_i(\theta)\} = 0$

From $\int f(y_i|x_i;\theta)\, dy_i = 1$ follows

$$\int \frac{\partial f(y_i \mid x_i;\theta)}{\partial \theta} dy_i = 0$$

Transformation

$$\frac{\partial f(y_i \mid x_i;\theta)}{\partial \theta} = \frac{\partial \log f(y_i \mid x_i;\theta)}{\partial \theta} f(y_i \mid x_i;\theta) = s_i(\theta) f(y_i \mid x_i;\theta)$$

gives

$$\int s_i(\theta) f(y_i \mid x_i;\theta)\, dy_i = E\{s_i(\theta)\} = 0$$

This theoretical moment for the scores is valid for any density *f*(.)

Quasi-ML Estimator, cont'd

Use of the GMM idea – substitution of moment conditions by sample equivalents – suggests to transform E{$s_i(\theta)$} = 0 into its sample equivalent and solve the first-order conditions

$$\frac{1}{N}\sum_i s_i(\theta) = 0$$

This reproduces the ML estimator

Example: For the linear regression $y_i = x_i'\beta + \varepsilon_i$, application of the Quasi-ML concept starts from the sample equivalents of

$$E\{(y_i - x_i'\beta)\, x_i\} = 0$$

this corresponds to the moment conditions of the OLS and the first-order condition of the ML estimators

- does not depend of the normality assumption of ε_i!

Quasi-ML Estimator, cont'd

- Can be based on a wrong likelihood assumption
- Consistency is due to starting out from E{s_i(θ)} = 0
- Hence, "quasi-ML" (or "pseudo ML") estimator

Asymptotic distribution:

- May differ from that of the ML estimator:

$$\sqrt{N}(\hat{\theta}-\theta) \rightarrow N(0,V)$$

- Using the asymptotic distribution of the GMM estimator gives

$$\sqrt{N}(\hat{\theta}-\theta) \rightarrow N\left(0, I(\theta)^{-1} J(\theta) I(\theta)^{-1}\right)$$

  with $J(\theta) = \lim (1/N)\Sigma_i E\{s_i(\theta)\, s_i(\theta)'\}$

  and $I(\theta) = \lim (1/N)\Sigma_i E\{-\partial s_i(\theta)/\partial\theta'\}$
- For linear regression: heteroskedasticity-consistent covariance matrix

Your Homework

1. Open the Greene sample file "greene7_8, Gasoline price and consumption", offered within the Gretl system. The variables to be used in the following are: G = total U.S. gasoline consumption, computed as total expenditure divided by price index; Pg = price index for gasoline; Y = per capita disposable income; Pnc = price index for new cars; Puc = price index for used cars; Pop = U.S. total population in millions. Perform the following analyses and interpret the results:
   a. Produce and interpret the scatter plot of the per capita (p.c.) gasoline consumption (Gpc) over the p.c. disposable income.
   b. Fit the linear regression for log(Gpc) with regressors log(Y), Pg, Pnc and Puc to the data and give an interpretation of the outcome.

Your Homework, cont'd

c. Test for autocorrelation of the error terms using the LM test statistic $\xi_{LM} = (T\text{-}1)\, R^2$ with R^2 from the auxiliary regression of the ML residuals on the lagged residuals with appropriately chosen lags.
d. Test for autocorrelation using NR_e^2 with R_e^2 from the regression of the ML residuals on x_t and the lagged residuals.

2. Assume that the errors ε_t of the linear regression $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$ are NID(0, σ^2) distributed. (a) Determine the log-likelihood function of the sample for $t = 1, \ldots, T$; (b) show that the first-order conditions for the ML estimators have expectations zero for the true parameter values; (c) derive the asymptotic covariance matrix on the basis (i) of the information matrix and (ii) of the score vector; (d) derive the matrix S of scores for the omitted variable LM test [cf. eq. (6.38) in Veebeek].