Advanced Econometrics - Lecture 4

ML Estimation, Diagnostic Tests

Advanced Econometrics - Lecture 4

- ML Estimator: Concept and Illustrations
- ML Estimator: Notation and Properties
- ML Estimator: Two Examples
- Asymptotic Tests
- Some Diagnostic Tests
- Quasi-maximum Likelihood Estimator

Estimation Concepts

IV estimator: Model allows derivation of moment conditions which are functions of

- Observable variables (endogenous and exogenous)
- Instrument variables
- Unknown parameters θ

Moment conditions are used for deriving estimators of the parameters θ

Special case is the

- OLS estimation

An alternative concept is basis of maximum likelihood estimation

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

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

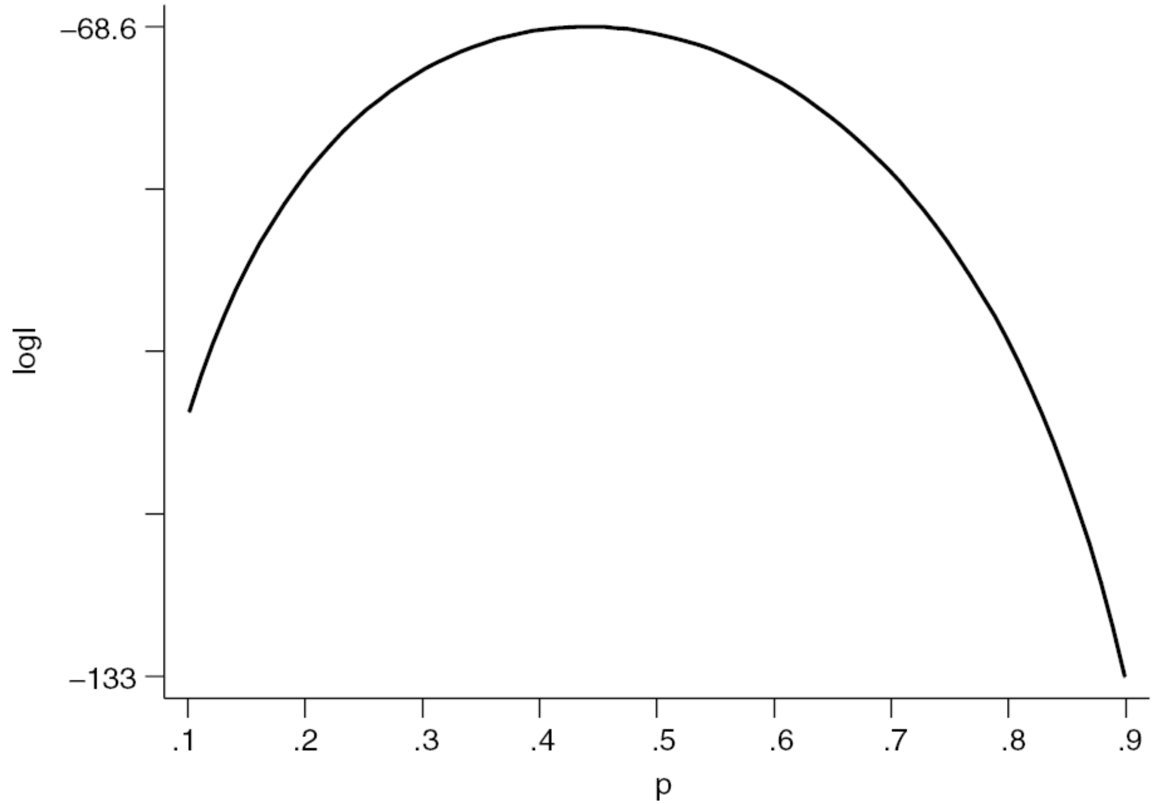


Figure 6.1 Sample loglikelihood function for $N = 100$ and $N_1 = 44$

Urn Experiment: ML Estimator

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

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

Solving the 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: the likelihood contribution of observation *i* is

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

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

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

Normal Linear Regression, cont'd

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

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

$$\hat{\beta} = -$$

which coincide with the OLS estimators:

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

which underestimates σ^2!

Remarks:

- The results are obtained assuming the *IIN*-distributed error terms
- ML estimators might or might not have nice properties like unbiasedness, BLUE

Advanced Econometrics - Lecture 4

- ML Estimator: Concept and Illustrations
- ML Estimator: Notation and Properties
- ML Estimator: Two Examples
- Asymptotic Tests
- Some Diagnostic Tests
- Quasi-maximum Likelihood Estimator

ML Estimator: Notation

Let the density (or probability mass function) 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

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

The ML estimators are the solutions of

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

or of the first-order conditions

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

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

Solution of $s(\theta) = 0$

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

ML Estimator: Properties

The ML estimator

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

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

V: asymptotic covariance matrix

Information Matrix

Information matrix $I(\theta)$

- Is the limit (for N →∞) of

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

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

The matrix of second derivatives is also called the Hessian matrix

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

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

Covariance Matrix *V*: Calculation

Two ways to calculate *V*:

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

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

  index "H": the estimate of *V* is based on the Hessian matrix, the matrix of second derivatives

- 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": based on gradients, the first derivatives

  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

Advanced Econometrics - Lecture 4

- ML Estimator: Concept and Illustrations
- ML Estimator: Notation and Properties
- ML Estimator: Two Examples
- Asymptotic Tests
- Some Diagnostic Tests
- Quasi-maximum Likelihood Estimator

Example: Urn Experiment

Likelihood contribution of the *i*-th observation

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

This gives

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

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: Binomial Distribution

The asymptotic distribution is

$$\sqrt{N}(\hat{p} - \quad \rightarrow \quad p(1 -$$

- Small sample distribution:

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

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

  $Np(1-p) > 9$

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:

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

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

Normal Linear Regression, cont'd

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

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

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

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:

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

similar to OLS results

Advanced Econometrics - Lecture 4

- ML Estimator: Concept and Illustrations
- ML Estimator: Notation and Properties
- ML Estimator: Two Examples
- Asymptotic Tests
- Some Diagnostic Tests
- Quasi-maximum Likelihood Estimator

Diagnostic Tests

Diagnostic tests based on ML estimators

Test situation:

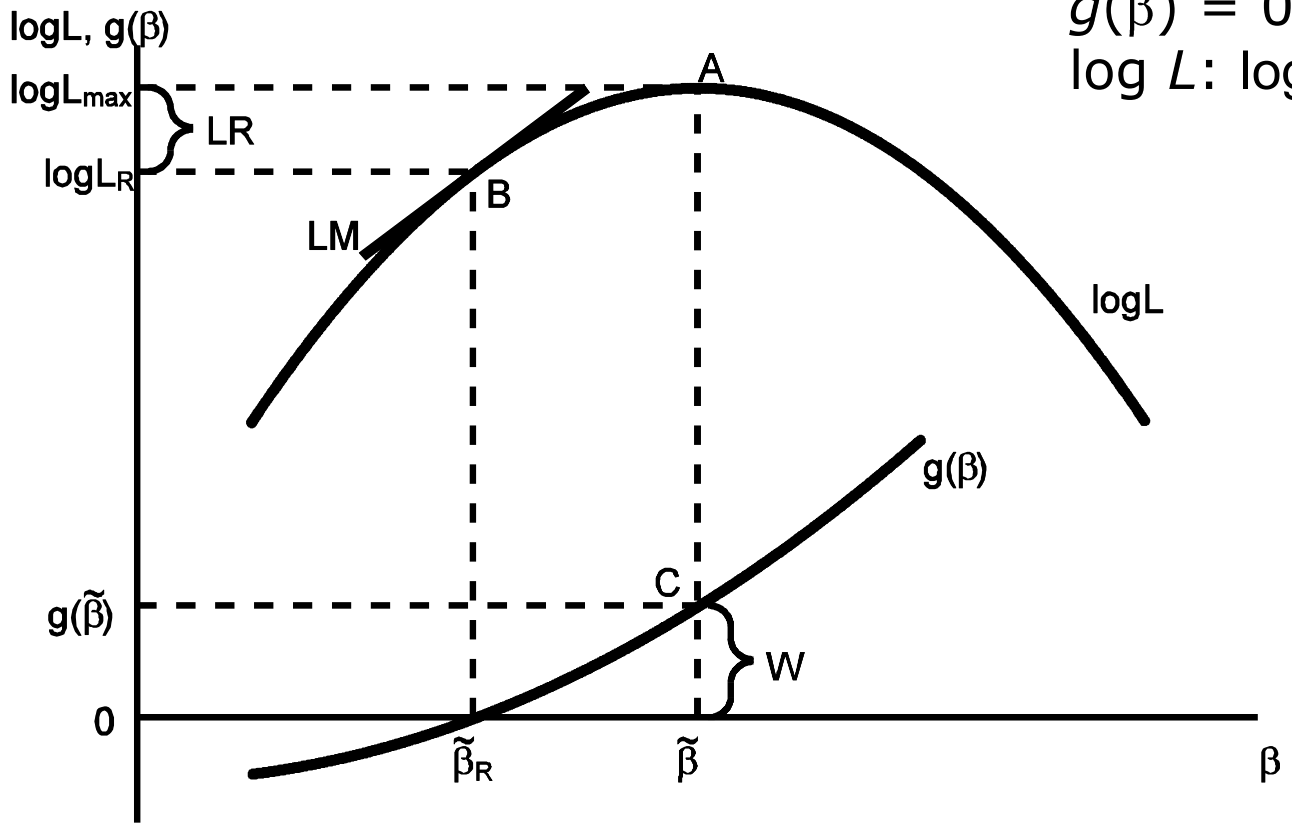
- *K*-dimensional parameter vector $\theta = (\theta_1, \ldots, \theta_K)'$
- $J \geq 1$ linear restrictions
- H_0: $R\theta = q$ with JxK matrix R, full rank; J-vector q

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 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: tests require different computational efforts

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

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

is - under H_0 - expected to be close to zero

p-value to be read from the Chi-square distribution with *J* df

Algebraically equivalent with *e*: residuals from unrestricted model, and e_R: residuals from restricted model, to

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

Likelihood Ratio Test

Checks whether the difference between the log-likelihood values 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

Calculation: with e (e_R): residuals from unrestricted (restricted) model

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

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-vector 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 $\tilde{\lambda}$

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

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

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

Lagrange Multiplier Test, cont'd

Lagrange multiplier test statistic

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

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

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

Calculation of ξ_{LM}

- Outer product gradient (OPG) version
- Auxiliary regression of a N-vector i = (1, …, 1)' on the scores $s(\theta)$ with restricted estimates for θ, no intercept
- Test statistic is

$$\xi_{LM} = N R^2$$

with the uncentered R^2 of the auxiliary regression

Advanced Econometrics - Lecture 4

- ML Estimator: Concept and Illustrations
- ML Estimator: Notation and Properties
- ML Estimator: Two Examples
- Asymptotic Tests
- Some Diagnostic Tests
- Quasi-maximum Likelihood Estimator

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:

$$\xi = \hat{p} - \quad \hat{p}(1 - \hat{p})$$

Likelihood ration test:

$$\xi = \quad \log L(\hat{p}) - \log L(\tilde{p})$$

with

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

$$\log L(\tilde{p}) = \log(p_0) + \quad - \quad)\log - $$

An Illustration, cont'd

Lagrange multiplier test:

with

$$\tilde{\lambda} = \sum_i \frac{\partial}{\partial p} \log L_i(p)\Big|_{p_0} = N\frac{\hat{p}}{p_0} - N\frac{1-\hat{p}}{1-p_0} = N\frac{\hat{p}-p_0}{p_0(1-p_0)}$$

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

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

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

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

$$\frac{1}{\tilde{\sigma}^2}\sum_i \tilde{\varepsilon}_i, \quad \frac{1}{\tilde{\sigma}^2}\sum_i \tilde{\varepsilon}_i$$

with constrained ML estimators for β and σ^2

- Auxiliary regression of N-vector i = (1, …, 1)' on the scores gives the uncentered R^2
- The LM test statistic is $\xi_{LM} = N R^2$
- An asymptotically equivalent LM test statistic is $N R_e^2$ 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, $h(0) = 1$, $h'(.) \neq 0$, J-vector z_i

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

First-order conditions give the scores

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

with constrained ML estimators for β and σ^2

- Auxiliary regression of N-vector $i = (1, \ldots, 1)'$ on the scores gives the uncentered R^2
- LM test statistic $\xi_{LM} = NR^2$; 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 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

$$\varepsilon \qquad \varepsilon$$

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

Advanced Econometrics - Lecture 4

- ML Estimator: Concept and Illustrations
- ML Estimator: Notation and Properties
- ML Estimator: Two Examples
- Asymptotic Tests
- Some Diagnostic Tests
- Quasi-maximum Likelihood Estimator

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 suffice
- consistency applies

Generalized Method of Moments (GMM)

GMM is a general estimation method, which encompasses all parametric estimation techniques

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, exogenous or endogenous
- z_i: vector of instrument variables
- θ: K-vector of unknown parameters
- Basis of the moment conditions are theoretical arguments

Example: For linear model $y_i = x_i'\beta + \varepsilon_i$, $w_i' = (y_i, x_i')$

GMM Estimator

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

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

Properties: GMM estimator

- consistent (for any choice of W_N)
- the most efficient estimator if for W_N the optimal weighting matrix is used:

$$W_N^{opt} = \left(E\{f(w_i, z_i, \hat{\theta})\, f(w_i, z_i, \hat{\theta})'\}\right)^{-1}$$

- asymptotic distribution:

$$\sqrt{N}(\hat{\theta} - \theta) \rightarrow N(0, (D W_N^{opt} D')^{-1})$$

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

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

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(θ)} = 0

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

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

Transformation

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

gives

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

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

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

This reproduces the ML estimator

Example: For the linear regression $y_i = x_i'\beta + \varepsilon_i$, application of the 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{$s_i(\theta)$} = 0
- Hence, "quasi-ML" (or "pseudo ML") estimator

Asymptotic distribution:

- May differ from that of the ML estimator:

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

- Using the asymptotic distribution of the GMM estimator gives

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

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

Exercise

1. The dataset DatS03 Investment, USA 1968-1982 (Economic Report of the President: 1983, also dataset F3.1 of W.H. Greene) contains the variables GNP (nominal GNP), INVEST (nominal investment), PC (consumer price index) and R (interest rate, measured as the average of the discount rate on the New York Federal Reserve Bank). Based on these variables the investment function
   $$IR_t = \beta_1 + \beta_2 (t-1967) + \beta_3 GNPR_t + \beta_4 R_t + \beta_5 PI_t + u_t$$
   is defined, where IR and GNPR are to real investment or real GNP converted variables, respectively, and the inflation rate PI is calculated from the consumer price index PC.
   a. Some economists believe that the investments are only determined by the real interest rate (R - PI), and that the inflation rate has no effect otherwise. Specify a suitable linear constraint (in matrix notation) for the investment function that makes it possible to verify this claim.
   b. Estimate the coefficients of the investment function with and without consideration of the restriction $\beta_4 + \beta_5 = 0$ and test by means of the *F*-, the Wald, the Lagrange multiplier and the likelihood ratio test whether $\beta_4 + \beta_5 = 0$ can be regarded as a true restriction. Explain the conditions that must be met for the use of the different tests.

Exercise, cont'd

2. Test the null hypothesis that in the investment function from problem 1 the following three linear constraints are true: $\beta_2 = 0$, $\beta_3 = 1$, and $\beta_4 + \beta_5 = 0$.
   b. Specify the matrices *R* and *q*
   c. For testing, use (i) the substitution method and (ii) the Wald test.