Advanced Econometrics Lecture Notes

Burak Saltoğlu

March 2003

Contents

Chapter 1
MULTIPLE REGRESSION: MATRIX APPROACH

1.0.1 The k-variable regression model

$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + ... + \beta_k X_{k1} + u_1$

$Y_2 = \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + ... + \beta_k X_{k2} + u_2.$

.

.

.

$Y_n = \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + ... + \beta_k X_{kn} + u_n$

In vector notation

$$\begin{bmatrix} Y_1 \\ . \\ . \\ . \\ . \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & . & . & . & . & X_{k1} \\ 1 & X_{12} & . & & & & & X_{k2} \\ 1 & & & . & & & & \\ . & & & & . & & & \\ . & & & & & . & & \\ . & & & & & & . & \\ 1 & X_{1n} & . & . & . & . & . & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \ddots \\ . \\ . \\ . \\ \beta_n \end{bmatrix} + \begin{bmatrix} u_1 \\ . \\ . \\ . \\ . \\ u_n \end{bmatrix}$$

$y = X\beta + u$

1.0.2 Assumptions

1.) $E(u) = 0$

$$2.)\ E(uu^{'}) = \begin{bmatrix} E(u_1^2) & E(u_1u_2) & . & . & . & . & E(u_1u_n) \\ E(u_1u_2) & E(u_2^2) & & & & & \\ & & E(u_3^2) & & & & \\ & & & . & & & \\ & & & & . & & \\ & & & & & . & \\ & . & . & . & . & . & E(u_n^2) \end{bmatrix} =$$

$$= \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & .0 & 0. & 0 \\ 0 & \sigma^2 & & & & & \\ 0 & 0 & \sigma^2 & & & & \\ 0 & 0 & 0 & . & & & \\ 0 & 0 & 0 & & . & & \\ 0 & 0 & 0 & & & . & \\ 0 & 0. & . & . & . & . & \sigma^2 \end{bmatrix}_1$$

$$=\sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 & .0 & 0. & 0 \\ 0 & 1 & & & & & \\ 0 & 0 & 1 & & & & \\ 0 & 0 & 0 & . & & & \\ 0 & 0 & 0 & & . & & \\ 0 & 0 & 0 & & & . & \\ 0 & 0. & . & . & . & . & 1 \end{bmatrix}$$

or

$$= \qquad \sigma I_N$$

$$\text{where } I_N = \begin{bmatrix} 1 & 0 & 0 & 0 & .0 & 0. & 0 \\ 0 & 1 & & & & & \\ 0 & 0 & 1 & & & & \\ 0 & 0 & 0 & . & & & \\ 0 & 0 & 0 & & . & & \\ 0 & 0 & 0 & & & . & \\ 0 & 0. & . & . & . & . & 1 \end{bmatrix}$$

3.) X consists of &xed numbers.
4.) The rank of X is$k$ where k is the number of columns in X and k$\leq N$
5.) $u\tilde{} N(0, \sigma^2)$

1.0.3 OLS Estimation

$Y = X\hat{\beta} + \hat{u}$

$\hat{u} = (Y - X\hat{\beta})$

$\hat{u}' = (Y - X\hat{\beta})$

One way to estimate the coefficients it has to be squared and minimise the function. There are other ways i.e minimising the absolute value of residuals (Least Absolute Deviations: LAD)

$$\hat{u}'\hat{u} = (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

$$SSR = \hat{\beta}X'y - \hat{\beta}X'X\hat{\beta} - yy' + y'X\hat{\beta}'$$

$$SSR = yy' - 2\beta' X'y + \hat{\beta}' X'X\hat{\beta}$$

To minimize SSR

$$\frac{\partial SSR}{\partial \beta} = -X'y + (X'X)\hat{\beta}' = 0$$

$$\Longrightarrow \hat{\beta} = (X'X)^{-1}X'Y$$

1.0.4 Properties of OLS Estimates

$$\hat{\beta} = (X'X)^{-1}X'Y$$

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

$$E(\hat{\beta}) = E\{(X'X)^{-1}X'(X\beta + u)\}$$

$$E(\hat{\beta}) = \beta$$

$$var(\hat{\beta}) = E\{[((X'X)^{-1}X'X\beta) - E(\hat{\beta})][(X'X)^{-1}XX'\hat{\beta} + (X'X)^{-1}X'u - \beta]'\}$$

after some manipulations:

$$var(\hat{\beta}) = E\,[[(X'X)^{-1}\,\mathrm{u\,u'}\,X'X(X'X)^{-1}]$$

$$var(\hat{\beta}) = \sigma^2(XX')^{-1}$$

GAUSS MARKOV THEOREM:OLS estimators are Best Linear Unbiased Estimators (BLUE)

Proof: let $\hat{\beta}^*$be another linear estimator of β

1.) Unbiasedness

$$\hat{\beta}^* = \hat{\beta} + cy$$

Now

$$\hat{\beta}^* = [(X'X)^{-1}X' + c][X\beta + u]$$

$$\hat{\beta}^* = \beta + (X'X)^{-1}X'u + cX\beta + cu$$

$$E(\hat{\beta}^*) = \beta + cX\beta + cu$$

The only way to make this estimator to be an unbiased estimator of β is to make c=0 which is OLS estimator.

2.) Minimum variance

$$E\{(\hat{\beta}^* - \beta)\,(\hat{\beta}^* - \beta)'\} = E\{[(X'X)^{-1}X'u + cu][(X'X)^{-1}X'u + cu]'\}$$
$$= \sigma^2(XX') + \sigma^2(cc') + \sigma^2 cX(XX')^{-1}$$

where $\sigma^2 cX(XX')^{-1} = 0$ since $cX = 0$ from above

$$E\{(\hat{\beta}^* - \beta)\,(\hat{\beta}^* - \beta)'\} = \sigma^2(XX') + \sigma^2(cc')$$

The only condition that this linear estimator is more efficient is when the $(cc') = 0$ which implies that c is the null matrix.
This condition implies that both estimates are the same i.e
$\hat{\beta} = \hat{\beta}^*$i.e the OLS estimator is unbiased and most efficient among other linear estimators.

Chapter 2
HYPOTHESIS TESTING (K-VARIABLE) MULTIPLE REGRESSION

A general way to test the existence of certian linear restriction is to use

$$R\beta = r$$

Suppose we want to test the restriction of

$$H_0 : \beta_2 - \beta_3 = 0$$

which can be represented as

$$R = \begin{bmatrix} 0 & 1 & -1 & . & 0. \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_{k-1} \\ \beta_k \end{bmatrix}$$

Distribution of test statistics

$$\begin{aligned} H_0 &: R\beta = r \\ H_1 &: R\beta \neq r \end{aligned}$$

Then

$$E(R\hat{\beta}) = R\beta$$

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

$$\sigma^2 R(X'X)^{-1}R'$$

Then it can be shown that

$$(R\hat{\beta} - r)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \sim \chi^2_q$$

. It can also be shown

since $(R\hat{\beta} - r) \sim N\ (0,'\ [\sigma^2 R(X'X)^{-1}R'])$ and the term above is the sum of square of standard normal variable which is distributed chi-square. We also know that

$$\frac{\hat{u}'\hat{u}}{\sigma^2} \sim \chi^2_{N-k-1}$$

Then,

$$F = \frac{\frac{(R\hat{\beta}-r)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta}-r)/q}{\sigma^2}}{\frac{\hat{u}'\hat{u}/(N-k-1)/(n-k-1)}{\sigma^2}}$$

which can be simpli&ed to yield

$$F = \frac{(R\hat{\beta} - r)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)/q}{\hat{u}'\hat{u}/(N - k - 1)} \sim F^{\alpha}_{q,N-k-1}$$

The reason why it has a F-distribution is becasues F is the ratio of two chi-squares with their respective degrees of freedom.

Decision rule: Reject
$H_0 : R\beta = r$
if F>$F^{\alpha}_{q,N-k-1}$.

Chapter 3

THEORY OF MAXIMUM LIKELIHOOD

New developments in the computer technology allow econometricians to use more ! exible methods.

3.1 Maximum Likelihood

Maximum likelihood method is based on the priniple that the parameter estimates can be obtained by maxising the likelihood of the selected sample to re! ect the population. In other words, we choose the parameters in a way that it maximises the likelihood of representing the population. Suppose we are given

$$y^{'} = (y_1, ...y_N) \tag{3.1}$$

be an n-vector dependent on k-vector unknown parameters θ.

Let the vector $\theta = (\theta_1, ..., \theta_K)$. If we write the joint density as

$$f(y; \theta) \tag{3.2}$$

Equation (2) can be interpreted in two ways:

i.) For a given θ,it indicates the probability of a set of sample outcomes.

ii.) It is a function of θ, conditional on a set of outcomes.

3.1.1 Likelihood function

$$L(\theta; y) = f(y; \theta) \tag{3.3}$$

Maximising the above likelihood function with respect to θ &nding a speci&c value of $\hat{\boldsymbol{\theta}}$ that maximises the probability of obtaining the sample values that have actually observed. Then $\hat{\boldsymbol{\theta}}$ is said to be Maximum Likelihood Estimator of the unknown parameter vector. In most applications it is more convenient to work with the log-likelihood function

$$l = \ln L \tag{3.4}$$

and then

$$\frac{\partial l}{\partial \theta} = \frac{1}{L}\frac{\partial L}{\partial \theta} \tag{3.5}$$

The MLE $\hat{\boldsymbol{\theta}}$ can be obtained by &nding $\hat{\boldsymbol{\theta}}$ such that
The above equation is known as score

3.1.2 Properties of MLE

1. Consistency:

$$p\lim(\hat{\boldsymbol{\theta}}) = \theta \tag{3.6}$$

2. Asymtotic Normality $\hat{\boldsymbol{\theta}} \overset{a}{\sim} N(\theta, I^{-1}(\theta))$
where

$$I(\theta) = E\left[(\frac{\partial l}{\partial \boldsymbol{\theta}})(\frac{\partial l}{\partial \boldsymbol{\theta}})'\right] = -E\left[(\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'})\right] \tag{3.7}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{bmatrix} \frac{\partial^2 l}{\partial \theta_1^2} & \frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} & & \frac{\partial^2 l}{\partial \theta_1 \partial \theta_k} \\ \frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 l}{\partial \theta_2^2} & & \\ . & & . & \\ . & & . & \\ . & & . & \\ \frac{\partial^2 l}{\partial \theta_k \partial \theta_1} & & & \frac{\partial^2 l}{\partial \theta_k^2} \end{bmatrix}$$

3. Asymptotic efficiency
If $\hat{\boldsymbol{\theta}}$ is the maximum liklelihood estimator of a single parameter

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2)$$

This states that If there exists another consistent and asymptotcially normal estimator $\tilde{\theta}$ of θ, then $\sqrt{n}\tilde{\theta}$ has a normal lmitting distribution whos variance is greater than or equal to σ^2.

4. Invariance
If $\hat{\theta}$ is the MLE of θ, and $g(\theta)$ is a continuous function of θ, then $g(\hat{\theta})$ is the MLE of $g(\theta)$.

5. The score has mean zero and variance $I(\theta)$

3.1.3 Estimation of the linear regression model

$$y = X\beta + u \tag{3.8}$$

$$f(u) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)(\mathrm{u\,u}')} \tag{3.9}$$

$$l = \ln f(y|X) = \ln f(u) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(uu^{'}) \tag{3.10}$$

$$l = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)^{'}(y - X\beta) \tag{3.11}$$

The vector of unknowns are
$\acute{\theta} = (\beta, \acute{\sigma}^2)$ which be solved by

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = -\frac{1}{\sigma^2}(-\acute{X}y + X^{'}X\beta) \tag{3.12}$$

$$\frac{\partial l}{\partial \boldsymbol{\sigma}^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X\acute{\beta}) \tag{3.13}$$

Then

$$\beta = (\acute{X}X)^{-1}(X^{'}y) \tag{3.14}$$

$$\sigma = \acute{u}u/n \tag{3.15}$$

3.1.4 SOME APPLICATIONS

Time series models AR(1) Model:

Suppose you are given an AR(1) model as

$$y_t = c + \phi y_{t-1} + \varepsilon_t \tag{3.16}$$

where $\varepsilon_t \backsim N(0, \sigma^2)$. Since all the errors are Gaussian, disregarding the &rst obervation in the sample the probability distribution of Y_2.

$$(Y_2) = c + \phi Y_1 + \varepsilon_2 \tag{3.17}$$

If we treat the &rst variable &xed then

$$(Y_2|Y_1 = y_1) \backsim N\left[(c + \phi Y_1 + \varepsilon_2), \sigma^2\right] \tag{3.18}$$

$$f_{Y_2|Y_1}(y_2|y_1;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(y_{2-}c - \phi y_1)^2}{2\sigma^2}\} \tag{3.19}$$

Then by using the principle of likelihood

$$f_{Y_3|Y_1Y_2}(y_3|y_{1,}y_2;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(y_{3-}c - \phi y_2)^2}{2\sigma^2}\} \tag{3.20}$$

$$f_{Y_3Y_2Y_1}(y_3, y_{2,}y_1;\theta) = f_{Y_3|Y_1Y_2}(y_3|y_{1,}y_2;\theta) f_{Y_2|Y_1}(y_2|y_1;\theta) \tag{3.21}$$

Then

$$f_{Y_TY_{T-1},...,Y_1}(y_T, y_{T-1,},...,y_1;\theta) = f_{Y1}(y_1;\theta) \prod_{i=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\theta) \tag{3.22}$$

$$L(\theta) = \log f_{Y_1}(y_1;\theta) + \sum_{i=1}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\theta) \tag{3.23}$$

$$L(\theta) = const + [(T-1)/2]\log(2\pi) - [(T-1)/2]\log(\sigma^2) - \sum_{i=2}^{T}\left[\frac{(y_t - c - \phi y_{t-1})^2}{\sigma^2}\right]() \tag{3.24}$$

3.1.5 Maximum Likelihood Estimation of Linear Model with Nonpherical Disturbances

$$y = X\beta + u \quad \text{with, } u \backsim N(0, \sigma^2\Omega) \tag{3.25}$$

$$var(u_i) = \sigma_i^2 X_{2i}^2 \tag{3.26}$$

$$var(u_i) = \sigma^2 \begin{bmatrix} X_{21}^2 & 0 & 0 & & 0 \\ 0 & X_{22}^2 & & & 0 \\ . & & . & & . \\ & & & . & . \\ & & & & . \quad . \\ 0 & . & . & . \quad . & X_{2n}^2 \end{bmatrix} \tag{3.27}$$

Multivariate normal density for u:

$$f(u) = (2\pi)^{-n/2}|\sigma^2\Omega| \exp\left[-1/2u^{'}(\sigma^2\Omega)^{-1}u\right] \tag{3.28}$$

for which the likelihood function can be written as

$$l = -n/2\ln(2\pi) - n/2\ln\sigma^2 - 1/2\ln|\Omega| - \frac{1}{2\sigma^2}(y - X\beta\acute{)}(y - X\beta) \tag{3.29}$$

3.1.6 Likelihood RatioTest

As before we may wan to test whether

$$H_0 : R\beta = r \tag{3.30}$$

The validity of the restrictions imposed on the model can be tested by likelihood ratio:

$$\lambda = \frac{L(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\sigma}}^2)}{L(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\sigma}})} \tag{3.31}$$

Rejec the null when lamda is small. A large sample test is

$$LR = 2\left[\ln L(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\sigma}}) - \ln L(\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\sigma}}^2)\right] \backsim \chi_q^2 \tag{3.32}$$

Chapter 4

SIMULTANEOUS EQUATIONS SYSTEMS

Since in economics we deal mostly with more than one equation we may need to consider multiple equation systems. Typical example is demand supply system and macroeconomic equilibrium. There are various forms of simultaneous equations systems one of them is Vector Autoregression (VAR).

4.1 VAR SYSTEMS

A simple autoregressive process is shown to be

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... \phi_k y_{t-k} + \varepsilon_t \tag{4.1}$$

$$\mathbf{y}_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + ... \Phi_k y_{t-k} + \boldsymbol{\varepsilon}_t \tag{4.2}$$

Φ is nxn
$\mathbf{y}_t$ nx1
etc.

4.1.1 *A simple VAR*

$$\mathbf{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \tag{4.3}$$

The above model is called a VAR(1) model with dimension 2.

4.1.2 *Forecasting, Impulse response functions and variance decomposition*

$$\hat{\mathbf{y}}_{t+1} = E(\mathbf{y}_{t+1} | y_t, ..., y_{t1}) \tag{4.4}$$

$$\hat{\mathbf{y}}_{t+1} = \Phi(\mathbf{y}_t) \tag{4.5}$$

$$\hat{\mathbf{y}}_{t+2} = \Phi^2(\mathbf{y}_t) \tag{4.6}$$

$$\hat{\mathbf{y}}_{t+n} = \Phi^N(\mathbf{y}_t) \tag{4.7}$$

Impulse response functions

If we want to see how does the system of equtions responses to impulses say

$\boldsymbol{\varepsilon}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$$y_1 = \begin{bmatrix} 4 \\ 0 \end{bmatrix} \tag{4.8}$$

$$\Phi = \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix} \tag{4.9}$$

$$\mathbf{y}_2 = \Phi \mathbf{y}_1 + \boldsymbol{\varepsilon}_2 \tag{4.10}$$

$$\mathbf{y}_2 = \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.6 \\ 0.8 \end{bmatrix} \tag{4.11}$$

$$\mathbf{y}_3 = \Phi \mathbf{y}_2 + \boldsymbol{\varepsilon}_3 \tag{4.12}$$

4.1.3 Structural Equations

$$\begin{aligned} y_{1t} + \beta_{12} y_{2t} + \gamma_{11} &= u_{1t} \\ \beta_{21} y_{1t} + y_{2t} + \gamma_{21} &= u_{2t} \end{aligned} \tag{4.13}$$

In matrix notation

$$\mathbf{B}\mathbf{y}_t + \mathbf{C}\mathbf{x}_t = \mathbf{u}_t \tag{4.14}$$

$$\mathbf{B} = \begin{bmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{bmatrix}, \mathbf{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} \tag{4.15}$$

$$\mathbf{C} = \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \end{bmatrix} \tag{4.16}$$

$$\mathbf{x}_t = [1] \tag{4.17}$$

Reduced form

$$\mathbf{y}_t = \Pi \mathbf{x}_t + \mathbf{v}_t \tag{4.18}$$

$$\Pi = -\mathbf{B}^{-1}\mathbf{C}, \tag{4.19}$$

$$\mathbf{v}_t = \mathbf{B}^{-1}\mathbf{u}_t \tag{4.20}$$

4.1.4 Identification problem

In the below model

$$\mathbf{B}\mathbf{y}_t + \mathbf{C}\mathbf{x}_t = \mathbf{u}_t \tag{4.21}$$

B is GxG matr,x of coefficients, **C** is GxK exogenous or predetermined variables y_t,x_t and u_t s are G, K and G vector elements.

Suppose

G: is the number of endogenous variables in the system

g: is the number of endogenous variables in the equation

k: is the number of exogenous variables in the equation

Order Condition

G-g$\geq k-1$ is the order condition.

G-g=k-1$\Longrightarrow$just ideneti&ed

G-g>k-1 overidenti&ed

G-g<m-1 underidenti&ed

4.1.5 Estimation

If we use OLS for a simultaneous equation system then the estimators will be biased and inconsistent estimates. Suppose we have the following

$$y_t = \beta_1 + \beta_2 z_{2t} + \beta_3 z_{3t} + \beta_4 z_{4t} + \beta_5 z_{5t} + u_t \tag{4.22}$$

where z_{4t} and z_{5t} are endogeneous and z_{2t} and z_{3t} are predetermined

$$\mathbf{x}_t = (1, z_{2t}, z_{3t}, \varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})^{\prime} \tag{4.23}$$

$$z_{it} = \delta_i \mathbf{x}_t + e_{it} \tag{4.24}$$

$$\beta_{2SLS} = \left[\sum_{t=1}^{T} \hat{z}_t x_t^{\prime}\right]^{-1} \left[\sum_{t=1}^{T} \hat{z}_t y_t\right] \tag{4.25}$$

2SLS will be a consistent estimate.

4.1.6 Full Information Maximum Likelihood

$$\mathbf{B}\mathbf{y}_t + \mathbf{C}\mathbf{x}_t = \mathbf{u}_t \tag{4.26}$$

$$\mathbf{y}_t = \Pi\mathbf{x}_t + \mathbf{v}_t \tag{4.27}$$

$$L = (2\pi)^{-n}|\Omega|^{-n/2}\exp\frac{1}{2}\left[\sum_{t=1}^{T}(\mathbf{y}_t - \Pi\mathbf{x}_t)'\Omega^{-1}(\mathbf{y}_t - \Pi\mathbf{x}_t)\right] \tag{4.28}$$

Chapter 5

GENERALIZED METHOD OF MOMENTS

Generalized Method of Moments (GMM) is another estimation method for a parameter θ. It has various advantages over other LS and MLE methods. Advantages: It is more ! exible than LS methods. It has very desirable asymptotic features. Of course these can turn into its disadvantages such as computational difficulties and undesirable small sample properties.

5.1 Method of moments

Suppose γ is the population moment which can be de&ned as the expectation of a continuous function

$$\gamma = E[g(x)] \tag{5.1}$$

Most known moments are &rst and the second moments

$$\gamma_{(1)} = \mathbf{E}[g(x)] \tag{5.2}$$

$$\gamma_{(2)} = \mathbf{E}[g(x)]^2 \tag{5.3}$$

Then if g(x) is a function of identity then the above &rst moment is mean second moment can be resembled to variance.

Sample Moments

$$\gamma_{(1)} = \frac{1}{n}\sum_{\mathbf{1}}^{\mathbf{n}} g(x) \tag{5.4}$$

$$\gamma_{(2)} = \frac{1}{n}\sum_{\mathbf{1}}^{\mathbf{n}} (g(x)^2) \tag{5.5}$$

5.1.1 OLS as a special case of Moment estimation

It can be shown that OLS is a special case of GMM. Remember standard OLS problem given as

$$y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{5.6}$$

where X nxK variable vector and $\boldsymbol{\epsilon} \backsim \mathbf{Q}(0, \boldsymbol{\sigma}^2)$.

If the model is correctly speci&ed then the restrictions can be written as

$$E(\mathbf{X}'\boldsymbol{\epsilon}) = \mathbf{0} \tag{5.7}$$

$$E[\mathbf{X}'(y - \mathbf{X}\boldsymbol{\beta})] = \mathbf{0} \tag{5.8}$$

This condition is known as moment or *orthogonality condition.* Then under normal conditions Then the estimation through GMM yields to equate the moment condition to zero

$$\frac{1}{n}\mathbf{X}'(y - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0} \tag{5.9}$$

$$\hat{\boldsymbol{\beta}}_{GMM} = (X'X)^{-1}X'y \tag{5.10}$$

This is nothing but the OLS estimator.

5.1.2 IV method as a Moment estimation

Consider the following model

$$y = \alpha + x_1\beta_1 + \boldsymbol{\epsilon} \tag{5.11}$$

However, using standard OLS is not valid here since $E(\mathbf{x}_1\boldsymbol{\epsilon}) \neq \mathbf{0}$ and hence the estimators will be inconsistent. One proposal is to use the *instrumental variable* estimation which is to &nd an instrument correlated with x_1 s and uncorrelated with $\boldsymbol{\epsilon}$. In other words, &nd a $\mathbf{Z}$ such that $E(\mathbf{Z}'\boldsymbol{\epsilon})$ =0 and $E(\mathbf{Z}'x_1)$. Then the IV estimator would be a solution to the below equation

$$\frac{1}{n}\mathbf{Z}'(y - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0} \tag{5.12}$$

Then the optimal GMM method developed by Hansen (1982) can be given as

$$\min_{\beta}\left([\frac{1}{n}\mathbf{Z}^{'}(y-\mathbf{X}\boldsymbol{\beta})]^{'}\hat{V}^{-1}[\frac{1}{n}\mathbf{Z}^{'}(y-\mathbf{X}\boldsymbol{\beta})]\right) \tag{5.13}$$

where $\hat{V}^{-1}$ is a consistent estimate of $[var(\frac{1}{n}\mathbf{Z}^{'}(\epsilon))]^{-1}$. In serially independent errors and homoscedastic the above will yield the 2SLS estimators.

5.1.3 GMM and Orthogonality condition

In general we can have all sorts of nonlinear orthogonality conditions arising from economics and &nance. Suppose this condition is given as $E[g(y,X,\theta)] = 0$, where g(.) is some (possibly nonlinear) continuous function of data and unkown parameters θ. The sample analog of these conditons will yield

$$m(\mathbf{y},\mathbf{X},\boldsymbol{\theta})^{'}\ \mathbf{W}\ m(\mathbf{y},\mathbf{X},\boldsymbol{\theta}) = 0 \tag{5.14}$$

It is also important to note that $\mathbf{W}$ is a consistent estimator of $var(m(.))^{-1}$. Then the GMM estimator can be found as

$$\min_{\theta} m(\mathbf{y},\mathbf{X},\boldsymbol{\theta})^{'}\ \mathbf{W}\ m(\mathbf{y},\mathbf{X},\boldsymbol{\theta}) = 0 \tag{5.15}$$

Furthermore,

$$m(\mathbf{y},\mathbf{X},\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} g(y_i, X_i, \theta) \tag{5.16}$$

5.1.4 Estimation of Dynamic Rational Expectation Models

Suppose an individual maximixes his/her lifetime utility. He/she also has portfolio consisted of a stock. This individual either consumes or saves and invests in the &nancial market instruments So his optimization problem becomes

$$\sum_{\tau=0}^{\infty}\delta^{\tau}E\{u(c_{t+}\mathcal{T})\} \tag{5.17}$$

subject to

$$A_t = \sum_{\tau=0}^{\infty}(1+r)^{-\tau}(c_{t+\tau} - w_{t+\tau}) \tag{5.18}$$

$$EZ_t'[u(c_{t+1}) - \zeta u'(c_t)] = 0 \tag{5.19}$$

$$u'(c_{t+1}) = \zeta u'(c_{t+1}) + \epsilon_{t+1} \tag{5.20}$$

If the utility function is logarithmic then consumption follows a random walk i.e

$$(c_{t+1}) = \beta_0 + \lambda(c_{t+1}) + \epsilon_{t+1} \tag{5.21}$$

What to include in Z?
Past consumption etc

If the utility is not logarithmic GMM

If the utility is not logarithmic but *constant relative risk aversion*

$$u(c_t) = \left\{ \begin{array}{cc} \frac{c_t^{1-\gamma}}{1-\gamma} & 1{>}\gamma > 0 \\ \log c_t & \gamma = 1 \end{array} \right\} \tag{5.22}$$

$$c_t^{-\gamma} = \zeta c_{t+1}^{-\gamma} \tag{5.23}$$

In other words the population moment condition $m(\mathbf{y}, \mathbf{X}, \boldsymbol{\theta})$

$$\frac{1}{n}\sum_{i=1}^{n} g(y_i, X_i, \theta) = c_t^{-\gamma} - \zeta E(c_{t+1}^{-\gamma}) = 0 \tag{5.24}$$

$$\min_{\theta} m(\mathbf{y}, \mathbf{X}, \boldsymbol{\theta})' \ \mathbf{W} \ m(\mathbf{y}, \mathbf{X}, \boldsymbol{\theta}) = 0 \tag{5.25}$$

Chapter 6
NONLINEAR REGRESSION MODELS

So far we have only dealt with linear models but the cost of convenince in modelling is the in! exibility. A general form of Non-linear form is as follows

$$y_i = h(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i \tag{6.1}$$

Special cases

$$y_i = \beta_1 + \beta_2 e^{\beta_3 x} + \varepsilon_i \tag{6.2}$$

$$y_i = \beta_0 x_1^{\beta_1} x_2^{\beta_2} e^{\varepsilon_i} \tag{6.3}$$

If the error terms above is normally distributed then the nonlinear regression model

$$SSR(\beta) = \frac{1}{2}\sum_{i=1}^{N} \varepsilon_i^2 = \frac{1}{2}\sum_{i=1}^{N} (y_i - h(\mathbf{x}_i, \boldsymbol{\beta}))^2 \tag{6.4}$$

However, unlike in the case of OLS, we can not take the partial derivatives of the above functional and solve anaylitically. Therefore, the major difference between OLS and NLS is that the &rst order conditions of the latter are non-linear functions of parameters β.

6.1 Two major methods in estimating the NLS

6.1.1 Linearizing the NLS or (GAUSS-Newton Regression)

In this method the NL equation is linearized by Taylor series expansion around the parameter vector $\boldsymbol{\beta}^0$:

$$h(\mathbf{x}, \boldsymbol{\beta}) \cong h(\mathbf{x}, \boldsymbol{\beta}^0) + \sum_{k=1}^{K} \frac{\partial h(\mathbf{x}, \boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}^0}(\beta_k - \beta_k^0) \tag{6.5}$$

Let x_k^0 equal the kth partial derivative

$$y^0 = y - h^0 + x^{0\prime}\boldsymbol{\beta}^0 \tag{6.6}$$

Example

Suppose we have the model 1

$$y_i = \beta_1 + \beta_2 e^{\beta_3 x} + \varepsilon_i \tag{6.7}$$

Then

$$x_1^0 = \frac{\partial h(\mathbf{x}, \boldsymbol{\beta}^0)}{\partial \beta_1^0} = 1 \tag{6.8}$$

$$x_2^0 = \frac{\partial h(\mathbf{x}, \boldsymbol{\beta}^0)}{\partial \beta_2^0} = e^{\beta_3^0 x} \tag{6.9}$$

$$x_3^0 = \frac{\partial h(\mathbf{x}, \boldsymbol{\beta}^0)}{\partial \beta_3^0} = \beta_2^0 x e^{\beta_3^0 x} \tag{6.10}$$

With a set of values of $\boldsymbol{\beta}^0$

$$y^0 = y - h(x, \beta_1^0, \beta_2^0, \beta_3^0) + \beta_1^0 x_1^0 + \beta_2^0 x_2^0 + \beta_3^0 x_3^0 \tag{6.11}$$

so it is regressed on the right hand side and &nd β_1, β_2 and β_3.

Gauss Newton Regression: GNR can be done by repetitive OLS for the above until $\boldsymbol{\beta}^0 \cong \boldsymbol{\beta}$.

6.1.2 Nonlinear Least Squares via Maximum Likelihood

Suppose in general we have model

$$g(y_i, \boldsymbol{\theta}) = h(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i \tag{6.12}$$

$$S(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{i=1}^{N} [g(y_i, \boldsymbol{\theta}) - h(\mathbf{x}_i, \boldsymbol{\beta})]^2 \tag{6.13}$$

For normal disturbances the density of y_i

$$y_i = |\frac{\partial \epsilon_i}{\partial y_i}|(2\pi\sigma^2)^{-1/2} \exp[-(g(y_i, \boldsymbol{\theta}) - h(\mathbf{x}_i, \boldsymbol{\beta}))/\mathbf{2}\sigma^2] \tag{6.14}$$

$$\ln L = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 + \sum_{i=1}^{n} \ln J(y_i, \theta) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} [g(y_i, \boldsymbol{\theta}) - h(\mathbf{x}_i, \boldsymbol{\beta})]^2 \tag{6.15}$$

where

$$J(y_i, \theta) = |\frac{\partial \epsilon_i}{\partial y_i}| \tag{6.16}$$

6.1.3 Properties of NLS estimator

Consistency and asymptotic normality is attained. However, it is hard judge whether it is the most efficient estimator except i has normally distributed errors.

Chapter 7
ASYMPTOTIC DISTRIBUTION THEORY

Asymptotic theory describes the properties of estimators when the sample size T goes to in&nity.

7.1 Some de&nitions

7.1.1 Convergence in probability

Consider a scalar random variables, $\{X_T\}_{T=1}^{\infty}$. The sequence is said to converge in probability to c if for every $\varepsilon > 0$, and every δ there exists a value N such that T$\geq N$

$$P\{|X_T - c|\} > \delta < \varepsilon \tag{7.1}$$

If we have sufficient observations te probability of X_T differs from c by more than δ can be made arbitrarily small for any δ. Plim can be represented as

$$P\lim X_T = c \tag{7.2}$$

$$X_T \xrightarrow{P} c \tag{7.3}$$

7.1.2 Convergence in Mean Square

A stronger condition than plim is convergence in Mean Square. A random sequence is said to coverge in MS to c is indicated as

$$X_T \xrightarrow{MS} c \tag{7.4}$$

if for every $\varepsilon > 0$, there exists a value N such that T$\geq N$

$$E(X_T - c)^2 < \varepsilon \tag{7.5}$$

7.1.3 Law of Large Numbers for iid Variables

Suppose the sample mean of $\bar{Y}_T = (\frac{1}{T}\sum_{t=1}^{T} Y_t)$ where $\{Y_t\}$ is i.i.d with mean μ and variance σ^2. Then $\bar{Y}_T$ has expectation of μ and variance

$$E(\bar{Y}_T - \mu)^2 = \frac{1}{T^2} var(\sum_{t=1}^{T} Y_t) = \frac{1}{T^2} var(\sum_{t=1}^{T} Y_t) = \frac{\sigma^2}{T} \tag{7.6}$$

Since $\frac{\sigma^2}{T} \to 0$ as T$\to \infty$, this means that $\bar{Y}_T \xrightarrow{MS} \mu$ implying also that $\bar{Y}_T \xrightarrow{p} \mu$.

```
\textbf{LAW\ of LARGE\ NUMBERS:
}
Sample mean is the consistent estimate of the population mean.
```

7.1.4 Convergence in Distribution

Let $\{X_T\}_{T=1}^{\infty}$ be a sequence of random variables, and let $F_{X_T}(x)$ denote the cumulative distribution function of X_T. Suppose there exists a cumulative distribution function $F_X(x)$ such that

$$\lim_{T\to\infty} F_{X_{T(x)}} = F_X(x) \tag{7.7}$$

at any value of x. Then X_T is said to *converge in distribution* to X denoted

$$X_T \xrightarrow{L} X \tag{7.8}$$

7.1.5 Central Limit Theorem

We have noticed that when T tends to in&nity, the sample mean Y_T has a degenrate probability density, collapsing toward a point mass at μ. However, we would like to describe the distribution of Y_T in more detail. For this purpose we transform a new random variable $\sqrt{T}(\bar{Y}_T - \mu)$ has a mean zero and variance given $(\sqrt{T})^2 var(\bar{Y}_T) = \sigma^2$. Then the *Central Limit Theorem* describes the result that the sequence $\sqrt{T}(\bar{Y}_T - \mu)$ converges in distribution to a Gaussian random variable. One form of CLT

$$\sqrt{T}(\bar{Y}_T - \mu) \xrightarrow{L} N(0, \sigma^2) \tag{7.9}$$