Introductory Econometrics

Lecture Notes*

Prof. Dr. Joachim Grammig

Winter Term 2006/07

*Prof. Dr. Joachim Grammig, University of Tübingen

1. Introduction

What is econometrics?

Econometrics = economic statistics $\cap$ economic theory $\cap$ mathematics[1]

Conceptional:

Data perceived as realizations of random variables

Parameters are real numbers, not random variables

Joint distributions of random variables depend on parameters

Motivating examples: Glosten-Harris model & Mincer equation

[1] according to Ragnar Frisch

Example 1: Derivation of key equation of the Glosten-Harris model[2]

Evolution of financial asset prices

Public and private information

Notation:

- Transaction price: P_t
- Indicator of transaction type: $Q_t = \begin{cases} 1 & \text{buyer initiated trade} \\ -1 & \text{seller initiated trade} \end{cases}$
- Trade volume: v_t
- Drift parameter: μ
- Earnings/costs of the market maker: c
- Unobserved component (public information): ε_t

[2] Journal of Financial Economics, 1988

Glosten-Harris model

Efficient/'fair' price: $m_t = \mu + m_{t-1} + \varepsilon_t + Q_t z_t, \quad z_t = z_0 + z_1 v_t$

Private information: $Q_t z_t$

Public information: ε_t

Market maker sets:

Sell price (ask): $P_t^a = \mu + m_{t-1} + \varepsilon_t + z_t + c$
Buy price (bid): $P_t^b = \mu + m_{t-1} + \varepsilon_t - z_t - c$

$\Rightarrow$ Transaction price change $\boxed{\Delta P_t = \mu + z_0 Q_t + z_1 v_t Q_t + c\Delta Q_t + \varepsilon_t}$

$\Rightarrow$ Estimation of unknown structural parameters $\boldsymbol{\beta}=(\mu, z_0, z_1, c)$'

Example 2: The influence of schooling on wages - Mincer equation

$$\boxed{ln(WAGE_i) = \beta_1 + \beta_2 S_i + \beta_3 TENURE_i + \beta_4 EXPR_i + \varepsilon_i}$$

Notation:

Logarithm of the wage rate: $ln(WAGE_i)$

Years of schooling: S_i

Experience in the current job: $TENURE_i$

Experience in the labor market: $EXPR_i$

$\Rightarrow$ Estimation of the parameters β_k, where β_2: return to schooling

General regression equations

Generalization: $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_K x_{iK} + \varepsilon_i$

Index for observations $i = 1, 2, ..., n$ and regressors $k = 1, 2, ..., K$

y_i	$=$	$\boldsymbol{\beta}'$	$\cdot$	$\mathbf{x}_i$	$+$	ε_i
(1x1)		(1xK)		(Kx1)		(1x1)

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix} \quad \text{and} \quad \mathbf{x_i} = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{pmatrix}$$

The key problem of econometrics: We deal with non-experimental data

Unobservable variables, interdependence, endogeneity, causality

Examples:

- Ability bias in Mincer equation
- Reverse causality problem if unemployment is regressed on liberalization index
- Causal effect on policy force and crime is not an independent outcome
- Simultaneity problem in demand price equation

2. The CLRM: Parameter Estimation by OLS

Hayashi p. 6/15-18

Classical linear regression model (CLRM)

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_K x_{iK} + \varepsilon_i = \underset{(1\text{x}K)}{\mathbf{x}_i'} \cdot \underset{(K\text{x}1)}{\boldsymbol{\beta}} + \varepsilon_i$$

y_i: Dependent variable, observed

$\mathbf{x}_i' = (x_{i1}, x_{i2}, ..., x_{iK})$: Explanatory variables, observed

$\boldsymbol{\beta}' = (\beta_1, \beta_2, ..., \beta_K)$: Unknown parameters

ε_i: 'Disturbance' component, unobserved

$\Rightarrow$ $\mathbf{b}' = (b_1, b_2, ..., b_K)$ estimator of $\boldsymbol{\beta}'$

$\Rightarrow$ $e_i = y_i - \mathbf{x}_i'\mathbf{b}$: Estimated residual

For convenience we introduce matrix notation

$$\underset{(nx1)}{\mathbf{y}} = \underset{(nxK)}{\mathbf{X}} \cdot \underset{(Kx1)}{\boldsymbol{\beta}} + \underset{(nx1)}{\boldsymbol{\varepsilon}}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{pmatrix} 1 & x_{12} & x_{13} & \dots & x_{1K} \\ 1 & x_{22} & & & \\ \vdots & \vdots & \ddots & & \vdots \\ 1 & x_{n2} & & \dots & x_{nK} \end{pmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Writing extensively: A system of linear equations

$$y_1 = \beta_1 + \beta_2 x_{12} + \ldots + \beta_K x_{1K} + \varepsilon_1$$

$$y_2 = \beta_1 + \beta_2 x_{22} + \ldots + \beta_K x_{2K} + \varepsilon_2$$

$$\vdots$$

$$y_n = \beta_1 + \beta_2 x_{n2} + \ldots + \beta_K x_{nK} + \varepsilon_n$$

We estimate the linear model and choose b such that SSR is minimized

Obtain an estimator $\mathbf{b}$ of $\boldsymbol{\beta}$ by minimizing the SSR (sum of squared residuals):

$$\underset{\{\mathbf{b}\}}{\text{argmin}} S(\mathbf{b}) = \text{argmin} \sum_{i=1}^{n} e_i^2 = \text{argmin} \sum_{i=1}^{n} (y_i - \mathbf{x}_i' \mathbf{b})^2$$

Differentiation with respect to $b_1, b_2, ..., b_K \quad \Rightarrow \quad$ FOC's:

$$\begin{aligned}
&(1)\ \frac{\partial S(\mathbf{b})}{\partial b_1} \stackrel{!}{=} 0 \quad \Rightarrow \sum e_i &= 0 \\
&(2)\ \frac{\partial S(\mathbf{b})}{\partial b_2} \stackrel{!}{=} 0 \quad \Rightarrow \sum e_i x_{i2} &= 0 \\
&\ \vdots \\
&(K)\ \frac{\partial S(\mathbf{b})}{\partial b_K} \stackrel{!}{=} 0 \quad \Rightarrow \sum e_i x_{iK} &= 0
\end{aligned}$$

$\Rightarrow$ FOC's can be conveniently written in matrix notation $\mathbf{X}'\mathbf{e} = 0$

The system of K equations is solved by matrix algebra

$$\mathbf{X}'\mathbf{e} = \mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\mathbf{b} = 0$$

Premultiplying by $(\mathbf{X}'\mathbf{X})^{-1}$:

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{b} = 0$$

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{I}\mathbf{b} = 0$$

OLS-estimator:

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

Alternatively:

$$\mathbf{b} = \left(\tfrac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\tfrac{1}{n}\mathbf{X}'\mathbf{y} = \left(\tfrac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\tfrac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i y_i$$

Zoom into the matrices $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{y}$

$$\mathbf{b} = \left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\frac{1}{n}\mathbf{X}'\mathbf{y} = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i y_i$$

$$\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' = \begin{pmatrix} \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \sum x_{i1}x_{i3} & \dots & \sum x_{i1}x_{iK} \\ \sum x_{i1}x_{i2} & \sum x_{i2}^2 & & & \sum x_{i2}x_{iK} \\ \vdots & \vdots & \ddots & & \vdots \\ \sum x_{i1}x_{iK} & \sum x_{i2}x_{iK} & & \dots & \sum x_{iK}^2 \end{pmatrix}$$

$$\sum_{i=1}^{n}\mathbf{x}_i y_i = \begin{pmatrix} \sum x_{i1}y_i \\ \sum x_{i2}y_i \\ \sum x_{i3}y_i \\ \vdots \\ \sum x_{iK}y_i \end{pmatrix}$$

3. Assumptions of the CLRM

Hayashi p. 3-13

The four core assumptions of CLRM

1.1 Linearity $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$

1.2 Strict exogeneity $E(\varepsilon_i|\mathbf{X}) = 0$

$\Rightarrow E(\varepsilon_i) = 0$ and $Cov(\varepsilon_i, x_{ik}) = E(\varepsilon_i x_{ik}) = 0$

1.3 No exact multicollinearity, $P(\text{rank}(X) = k) = 1$

$\Rightarrow$ No linear dependencies in the data matrix

1.4 Spherical disturbances: $Var(\varepsilon_i|\mathbf{X}) = E(\varepsilon_i^2|\mathbf{X}) = \sigma^2$

$Cov(\varepsilon_i, \varepsilon_j|\mathbf{X}) = 0; \quad E(\varepsilon_i \varepsilon_j|\mathbf{X}) = 0$

$\Rightarrow E(\varepsilon_i) = \sigma_i^2$ and $Cov(\varepsilon_i, \varepsilon_j) = 0$ by LTE (see Hayashi p. 18)

Interpreting the parameters β of different types of linear equations

Linear model $y_i = \beta_1 + \beta_2 x_{i2} + ... + \beta_K x_{iK} + \varepsilon_i$: A one unit increase in the independent variable x_{ik} increases the dependent variable by β_k units

Semi-log form $log(y_i) = \beta_1 + \beta_2 x_{i2} + ... + \beta_K x_{iK} + \varepsilon_i$: A one unit increase in the independent variable increases the dependent variable approximately by $100 \cdot \beta_k$ percent

Log linear model $log(y_i) = \beta_1 log(x_{i1}) + \beta_2 log(x_{i2}) + ... + \beta_K log(x_{iK}) + \varepsilon_i$: A one percent increase in x_{ik} increases the dependent variable y_i approximately by β_k percent

Some important laws

Law of Total Expectation (LTE):

$$E_X[E_{Y|X}(Y|X)] = E_Y(Y)$$

Double Expectation Theorem (DET):

$$E_X[E_{Y|X}(g(Y)|X)] = E_Y(g(Y))$$

Law of Iterated Expectations (LIE):

$$E_{Z|X}[E_{Y|X,Z}(Y|X,Z)|X] = E_{Y|X}(Y|X)$$

Some important laws (continued)

Generalized DET:

$$E_X[E_{Y|X}(g(X,Y))|X] = E_{X,Y}(g(X,Y))$$

Linearity of Conditional Expectations:

$$E_{Y|X}[g(X)Y|X] = g(X)E_{Y|X}[Y|X]$$

4. Finite sample properties of the OLS estimator

Hayashi p. 27-31

Finite sample properties of $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

1. $E(\mathbf{b}) = \boldsymbol{\beta}$: Unbiasedness of the estimator

   Holds for any sample size
   Holds under assumptions 1.1 - 1.3

2. $Var(\mathbf{b}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$: Conditional variance of $\mathbf{b}$

   Conditional variance depends on the data
   Holds under assumptions 1.1 - 1.4

3. $Var(\hat{\boldsymbol{\beta}}|\mathbf{X}) \geq Var(\mathbf{b}|\mathbf{X})$

   $\hat{\boldsymbol{\beta}}$ is any other linear unbiased estimator of $\boldsymbol{\beta}$
   Holds under assumptions 1.1 - 1.4

Some key results from mathematical statistics

$$\underset{(n\text{x}1)}{\mathbf{z}} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \qquad \underset{(m\text{x}n)}{\mathbf{A}} = \begin{pmatrix} a_{11} & a_{12} & & \dots & a_{1n} \\ a_{21} & a_{22} & & & \\ \vdots & \vdots & \ddots & & \vdots \\ a_{m1} & a_{m2} & & \dots & a_{mn} \end{pmatrix}$$

A new random variable: $\underset{(m\text{x}1)}{\mathbf{v}} = \underset{(m\text{x}n)}{\mathbf{A}} \cdot \underset{(n\text{x}1)}{\mathbf{z}}$

$$\underset{(m\text{x}1)}{E(\mathbf{v})} = \begin{pmatrix} E(v_1) \\ E(v_2) \\ \vdots \\ E(v_m) \end{pmatrix} = \mathbf{A}E(\mathbf{z})$$

$$\underset{(m\text{x}m)}{Var(\mathbf{v})} = \mathbf{A}Var(\mathbf{z})\mathbf{A}'$$

The OLS estimator's unbiasedness

$E(\mathbf{b}) = \boldsymbol{\beta} \;\; \Rightarrow \;\; E(\mathbf{b} - \boldsymbol{\beta}) = 0$

sampling error

$$\begin{aligned}
\mathbf{b} - \boldsymbol{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \boldsymbol{\beta} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) - \boldsymbol{\beta} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} - \boldsymbol{\beta} \\
&= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} - \boldsymbol{\beta} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}
\end{aligned}$$

$\Rightarrow \;\; E(\mathbf{b} - \boldsymbol{\beta}|X) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}|\mathbf{X}) = 0$ under assumption 1.2

$\Rightarrow \;\; E_{\mathbf{X}}(E(\mathbf{b}|\mathbf{X})) = E_{\mathbf{X}}(\boldsymbol{\beta}) = E(\mathbf{b})$ by the LTE

We show that $Var(\mathbf{b}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

$$\begin{aligned}
Var(\mathbf{b}|\mathbf{X}) &= Var(\mathbf{b} - \boldsymbol{\beta}|\mathbf{X}) \\
&= Var((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}|\mathbf{X}) = Var(\mathbf{A}\boldsymbol{\varepsilon}|\mathbf{X}) \\
&= \mathbf{A}Var(\boldsymbol{\varepsilon}|\mathbf{X})\mathbf{A}' = \mathbf{A}\sigma^2\mathbf{I}_n\mathbf{A}' \\
&= \sigma^2\mathbf{A}\mathbf{I}_n\mathbf{A}' = \sigma^2\mathbf{A}\mathbf{A}' \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

Note:

$\boldsymbol{\beta}$ non-random

$\mathbf{b} - \boldsymbol{\beta}$ sampling error

$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

$Var(\boldsymbol{\varepsilon}|\mathbf{X}) = \sigma^2\mathbf{I}_n$

Sketch of the proof of the Gauss Markov theorem

$Var(\hat{\beta}|\mathbf{X}) \geq Var(\mathbf{b}|\mathbf{X})$

$$
\begin{aligned}
Var(\hat{\beta}|\mathbf{X}) &= Var(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\mathbf{X}) = Var[(\mathbf{D}+\mathbf{A})\boldsymbol{\varepsilon}|\mathbf{X}] \\
&= (\mathbf{D}+\mathbf{A})Var(\boldsymbol{\varepsilon}|\mathbf{X})(\mathbf{D}'+\mathbf{A}') = \sigma^2(\mathbf{D}+\mathbf{A})(\mathbf{D}'+\mathbf{A}') \\
&= \sigma^2(\mathbf{DD}'+\mathbf{AD}'+\mathbf{DA}'+\mathbf{AA}') = \sigma^2[\mathbf{DD}'+(\mathbf{X}'\mathbf{X})^{-1}] \\
&\geq \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = Var(\mathbf{b}|\mathbf{X})
\end{aligned}
$$

where $\mathbf{C}$ is a function of $\mathbf{X}$

$\hat{\boldsymbol{\beta}} = \mathbf{Cy}$

$\mathbf{D} = \mathbf{C} - \mathbf{A}$

$\mathbf{A} \equiv (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

Details of proof: Hayashi pages 29 - 30

The OLS estimator is BLUE

- OLS is **l**inear

  Holds under assumption 1.1

- OLS is **u**nbiased

  Holds under assumption 1.1 - 1.3

- OLS is the **b**est **e**stimator

  Holds under the Gauss Markov theorem $Var(\hat{\beta}|\mathbf{X}) \geq Var(\mathbf{b}|\mathbf{X})$

5. Hypothesis Testing under Normality

Hayashi p. 33-45

Hypothesis testing

Economic theory provides hypotheses about parameters

$\Rightarrow$ If theory is right $\Rightarrow$ testable implications

But: Hypotheses can't be tested without distributional assumptions about $\boldsymbol{\varepsilon}$

Distributional assumption: Normality assumption about the conditional distribution of $\boldsymbol{\varepsilon}|\mathbf{X} \sim MVN(\mathbf{0}, \sigma^2\mathbf{I}_n)$ [Assumption 1.5]

Some facts from multivariate statistics

Vector of random variables: $\mathbf{x} = (x_1, x_2, ..., x_n)'$
Expectation vector:

$$E(\mathbf{x}) = \boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_n)' = (E(x_1), E(x_2), ..., E(x_n))'$$

Variance-covariance matrix:

$$Var(\mathbf{x}) = \Sigma = \begin{pmatrix} Var(x_1) & Cov(x_1, x_2) & \dots & Cov(x_1, x_n) \\ Cov(x_1, x_2) & Var(x_2) & & \\ \vdots & & \ddots & \vdots \\ Cov(x_1, x_n) & & \dots & Var(x_n) \end{pmatrix}$$

$\mathbf{y} = \mathbf{c} + \mathbf{A}\mathbf{x}$; $\mathbf{c}, \mathbf{A}$ non-random vector/matrix

$\Rightarrow \quad E(\mathbf{y}) = (E(y_1), E(y_2), ..., E(y_n))' = \mathbf{c} + \mathbf{A}\boldsymbol{\mu}$

$\Rightarrow \quad Var(\mathbf{y}) = \mathbf{A}\Sigma\mathbf{A}'$

$\Rightarrow \quad \mathbf{x} \sim MVN(\boldsymbol{\mu}, \Sigma) \Rightarrow \; \mathbf{y} = \mathbf{c} + \mathbf{A}\mathbf{x} \sim MVN(\mathbf{c} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\Sigma\mathbf{A}')$

Application of the facts from multivariate statistics and the assumptions 1.1 - 1.5

$$\underbrace{\mathbf{b}-\boldsymbol{\beta}}_{\text{sampling error}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

Assuming $\boldsymbol{\varepsilon}|\mathbf{X} \sim MVN(\mathbf{0}, \sigma^2\mathbf{I}_n)$

$$\Rightarrow \; \mathbf{b}-\boldsymbol{\beta}|\mathbf{X} \sim MVN\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}|\mathbf{X}), (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}_n\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right)$$

$$\Rightarrow \; \mathbf{b}-\boldsymbol{\beta}|\mathbf{X} \sim MVN\left(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$$

Note that $Var(\mathbf{b}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

OLS-estimator conditionally normal distributed if $\boldsymbol{\varepsilon}|\mathbf{X}$ is multivariate normal

Testing hypothesis about individual parameters (t-Test)

Null hypothesis: $H_0 : \beta_k = \bar{\beta}_k$, $\bar{\beta}_k$ a hypothesized value, a real number

Under assumption 1.5 and $\varepsilon|\mathbf{X} \sim MVN(\mathbf{0}, \sigma^2 \mathbf{I}_n) \Rightarrow$ alternative hypothesis:
$H_A : \beta_k \neq \bar{\beta}_k$

If H_0 is true $E(b_k) = \bar{\beta}_k$

$$\text{Test statistic: } t_k = \frac{b_k - \bar{\beta}_k}{\sqrt{\sigma^2[(\mathbf{X}'\mathbf{X})^{-1}]_{kk}}} \sim N(0, 1)$$

Note: $[(\mathbf{X}'\mathbf{X})^{-1}]_{kk}$ is the k-th row k-th column element of $(\mathbf{X}'\mathbf{X})^{-1}$

Nuisance parameter σ^2 can be estimated

$$\sigma^2 = E(\varepsilon_i^2|\mathbf{X}) = Var(\varepsilon_i|\mathbf{X}) = E(\varepsilon_i^2) = Var(\varepsilon_i)$$

We don't know ε_i but we use the estimator $e_i = y_i - \mathbf{x}_i'\mathbf{b}$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (e_i - \frac{1}{n}\sum_{i=1}^n e_i)^2 = \frac{1}{n}\sum_{i=1}^n e_i^2 = \frac{1}{n}\mathbf{e}'\mathbf{e}$$

$\hat{\sigma}^2$ is a biased estimator:

$$E(\hat{\sigma}^2|\mathbf{X}) = \frac{n-K}{n}\sigma^2$$

An unbiased estimator of σ^2

For $s^2 = \frac{1}{n-K}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-K}\mathbf{e}'\mathbf{e}$ we get an unbiased estimator

$$\Rightarrow E(s^2|\mathbf{X}) = \frac{1}{n-K}E(\mathbf{e}'\mathbf{e}|\mathbf{X}) = \sigma^2$$

$$E\left(E(s^2|\mathbf{X})\right) = E(s^2) = \sigma^2$$

Using this provides an unbiased estimator of $Var(\mathbf{b}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$:

$$\widehat{Var(\mathbf{b}|\mathbf{X})} = s^2(\mathbf{X}'\mathbf{X})^{-1}$$

$\Rightarrow$ t-statistic under H_0:

$$t_k = \frac{b_k - \bar{\beta}_k}{\sqrt{\left[\widehat{Var(\mathbf{b}|\mathbf{X})}\right]_{kk}}} = \frac{b_k - \bar{\beta}_k}{SE(b_k)} = \frac{b_k - \bar{\beta}_k}{\sqrt{\left[\widehat{Var(b_k|\mathbf{X})}\right]}} \sim t(n-K)$$

Decision rule for the t-test

1. $H_0 : \beta_k = \bar{\beta}_k$, is often $\bar{\beta}_k = 0$

   $H_A : \beta_k \neq \bar{\beta}_k$

2. Given $\bar{\beta}_k$, OLS-estimate b_k and s^2, we compute $t_k = \frac{b_k - \bar{\beta}_k}{SE(b_k)}$
3. Fix significance level α of two-sided test
4. Fix non-rejection and rejection regions $\Rightarrow$ decision

Remark:

$\sqrt{\sigma^2[(\mathbf{X}'\mathbf{X})^{-1}]_{kk}}$: standard deviation $b_k|\mathbf{X}$

$\sqrt{s^2[(\mathbf{X}'\mathbf{X})^{-1}]_{kk}}$: standard error $b_k|\mathbf{X}$

Testing joint hypotheses (F-test/Wald test)

Write hypothesis as:

$$H_0: \underset{(\#r \times K)}{\mathbf{R}} \quad \underset{(K \times 1)}{\boldsymbol{\beta}} = \underset{(\#r \times 1)}{\mathbf{r}}$$

$\mathbf{R}$: matrix of real numbers

$\mathbf{r}$: number of restrictions

Replacing the $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_k)$ by estimator $\mathbf{b} = (b_1, b_2, ..., b_K)'$:

$$\mathbf{R}\,\mathbf{b} = \mathbf{r}$$

Definition of the F-test statistic

Properties of $\mathbf{R}\,\mathbf{b}$:

$$\mathbf{R}\,E(\mathbf{b}|\mathbf{X}) = \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

$$\mathbf{R}\,Var(\mathbf{b}|\mathbf{X})\mathbf{R}' = \mathbf{R}\,\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$$

$$\boxed{\mathbf{R}\mathbf{b} = \mathbf{r} \sim MVN(\mathbf{R}\boldsymbol{\beta}, \mathbf{R}\,\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')}$$

Using some additional important facts from multivariate statistics

$$\mathbf{z} = (z_1, z_2, ..., z_m) \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Omega})$$
$$\Rightarrow \quad (\mathbf{z} - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\mathbf{z} - \boldsymbol{\mu}) \sim \chi^2(m)$$

Result applied: Wald statistic

$$\boxed{(\mathbf{R}\mathbf{b} - \mathbf{r})'[\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{r}) \sim \chi^2(\#\mathbf{r})}$$

Properties of the F-test statistic

Replace σ^2 by its unbiased estimate $s^2 = \frac{1}{n-K}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-K}\mathbf{e}'\mathbf{e}$ and dividing by $\#r$:

$\Rightarrow$ F-ratio:

$$
\begin{aligned}
F &= \frac{(\mathbf{Rb}-\mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb}-\mathbf{r})/\#\mathbf{r}}{(\mathbf{e}'\mathbf{e})/(n-K)} \\
&= (\mathbf{Rb}-\mathbf{r})'[\mathbf{R}\,\widehat{Var(\mathbf{b}|\mathbf{X})}\mathbf{R}']^{-1}(\mathbf{Rb}-\mathbf{r})/\#\mathbf{r} \sim F(\#\mathbf{r}, n-K)
\end{aligned}
$$

Note: F-test is one-sided

Proof: see Hayashi p. 41

Decision rule of the F-test

1. Specify H_0 in the form $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ and $H_A : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$.
2. Calculate F-statistic.
3. Look up entry in the table of the F-distribution for $\#\mathbf{r}$ and $n - K$ at given significance level.
4. Null is not rejected on the significance level α for F less than $F_\alpha(\#\mathbf{r}, n - K)$

Alternative representation of the F-statistic

Minimization of the unrestricted sum of squared residuals:

$$min \sum_{i=1}^{n}(y_i - \mathbf{x}_i'\mathbf{b})^2 \quad \Rightarrow \quad SSR_U$$

Minimization of the restricted sum of squared residuals:

$$min \sum_{i=1}^{n}(y_i - \mathbf{x}_i'\tilde{\mathbf{b}})^2 \quad \Rightarrow \quad SSR_R$$

F-ratio:

$$F = \frac{(SSR_R - SSR_U)/\#r}{SSR_U/(n-K)}$$

6. Confidence intervals and goodness of fit measures

Hayashi p. 38/20

Duality of t-test and confidence interval

Under $H_0 : \beta_k = \overline{\beta}_k$

$$t_k = \frac{b_k - \overline{\beta}_k}{SE(b_k)} \sim t(n-K)$$

Probability for non-rejection:

$$P\left(-t_{\frac{\alpha}{2}}(n-K) \leq t_k \leq t_{\frac{\alpha}{2}}(n-K)\right) = 1-\alpha$$

$-t_{\frac{\alpha}{2}}(n-K)$	lower critical value
$t_{\frac{\alpha}{2}}(n-K)$	upper critical value
t_k	random variable (value of test statistic)
$1-\alpha$	fixed number

$$\Rightarrow \; P\left(b_k - SE(b_k)t_{\frac{\alpha}{2}}(n-K) \leq \overline{\beta}_k \leq b_k + SE(b_k)t_{\frac{\alpha}{2}}(n-K)\right) = 1-\alpha$$

The confidence interval

Confidence interval for β_k:

$$P\left(b_k - SE(b_k)t_{\frac{\alpha}{2}}(n-K) \leq \beta_k \leq b_k + SE(b_k)t_{\frac{\alpha}{2}}(n-K)\right) = 1-\alpha$$

The confidence bounds are random variables!

$b_k - SE(b_k)t_{\frac{\alpha}{2}}(n-K)$: lower bound

$b_k + SE(b_k)t_{\frac{\alpha}{2}}(n-K)$: upper bound

Wrong Interpretation: True parameter β_k lies with probability $1-\alpha$ within the bounds of the confidence interval

Problem: Confidence bounds are not fixed; they are random!

H_0 is rejected at significance level α if the hypothesized value does not lie within the confidence bounds of the $1-\alpha$ interval.

Coefficient of determination: uncentered R^2

Measure of the variability of the dependent variable: $\sum y_i^2 = \mathbf{y}'\mathbf{y}$

Decomposition of $\mathbf{y}'\mathbf{y}$:

$$\begin{aligned}
\mathbf{y}'\mathbf{y} &= (\hat{\mathbf{y}} + \mathbf{e})'(\hat{\mathbf{y}} + \mathbf{e}) \\
&= \hat{\mathbf{y}}'\hat{\mathbf{y}} + 2\hat{\mathbf{y}}\mathbf{e} + \mathbf{e}'\mathbf{e} \\
&= \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e}
\end{aligned}$$

$$\Rightarrow \quad R^2_{uc} \equiv 1 - \tfrac{\mathbf{e}'\mathbf{e}}{\mathbf{y}'\mathbf{y}}$$

A good model explains much and therefore the residual variation is very small compared to the explained variation.

Coefficient of determination: centered R^2 and R^2_{adj}

Use centered R^2 if there is a constant in the model ($x_{i1} = 1$)

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{n} e_i^2$$

$$\Rightarrow \quad R_c^2 \equiv 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \equiv 1 - \frac{SSR}{SST}$$

Note, that R^2_{uc} and R^2_c lie both in the interval $[0, 1]$ but describe different models. They are not comparable!

R^2_{adj} is constructed with a penalty for heavy parametrization:

$$R^2_{adj} = 1 - \frac{SSR/(n-K)}{SST/(n-1)} = 1 - \frac{n-1}{n-K}\frac{SSR}{SST}$$

The R^2_{adj} is an accepted model selection criterion

Alternative goodness of fit measures

Akaike criterion (AIC): $\log\left(\frac{SSR}{n}\right) + \frac{2K}{n}$

Schwarz criterion (SBC): $\log\left(\frac{SSR}{n}\right) + \frac{\log(n)K}{n}$

Note:

Both criteria include a penalty term for heavy parametrization

Select model with smallest AIC/SBC

7. Introduction to Large Sample Theory

Hayashi p. 88-97/109-133

Basic concepts of large sample theory

Using large sample theory we can dispense with basic assumptions from finite sample theory

1.2 $E(\varepsilon_i|\mathbf{X}) = 0$: strict exogeneity

1.4 $Var(\boldsymbol{\varepsilon}|\mathbf{X}) = \sigma^2\mathbf{I}$: homoscedasticity

1.5 $\varepsilon|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$: normality of the error term

Approximate/assymptotic distribution of $\mathbf{b}$, and t- and the F-statistic can be obtained

Modes of convergence - Convergence in probability

$\{z_n\}$: sequence of random variables

$\{\mathbf{z}_n\}$: sequence of random vectors

<u>Convergence in probability:</u>

A sequence $\{z_n\}$ converges in probability to a constant α if for any $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(|z_n - \alpha| > \varepsilon\right) = 0$$

Short-hand we write: $\operatorname*{plim}_{n\to\infty} z_n = \alpha$ or $z_n \underset{p}{\to} \alpha$ or $z_n - \alpha \underset{p}{\to} 0$

<u>Extends to random vectors:</u>

If $\lim_{n\to\infty} P\left(|z_{kn} - \alpha_k| > \varepsilon\right) = 0 \;\forall\; k = 1, 2, ..., K$, then $\mathbf{z}_n \underset{p}{\to} \boldsymbol{\alpha}$

(element-wise convergence)

Modes of convergence - Convergence in mean square and distribution

Convergence in mean square:

$\lim_{n \to \infty} E\left[(z_n - \alpha)^2\right] = 0$ or $z_n \underset{m.s.}{\longrightarrow} \alpha$

Convergence in mean square implies convergence in probability.

Convergence in distribution:

$z_n \underset{d}{\longrightarrow} z$
if c.d.f. of z_n converges to the c.d.f. of z at each point of continuity.

Convergence in mean square and convergence in distribution extend to random vectors

Weak Law of Large Numbers (WLLN) according to Kinchin

$\{z_i\}$ i.i.d. with $E(z_i) = \mu$, then $\overline{z}_n = \frac{1}{n}\sum_{i=1}^{n} z_i$

we have:

$$\overline{z}_n \underset{p}{\longrightarrow} \mu \text{ or}$$

$$\lim_{n\to\infty} P\left(|\overline{z}_{kn} - \mu| > \varepsilon\right) = 0 \text{ or}$$

$$\text{plim } \overline{z}_n = \mu$$

Extensions of the Weak Law of Large Numbers (WLLN)

The WLLN holds for:

Extension (1): Multivariate Extension (sequence of random vectors $\{\mathbf{z}_i\}$)

Extension (2): Relaxation of independence

Extension (3): Functions of random variables $h(z_i)$

Extension (4): Vector valued functions $f(\mathbf{z}_i)$

Central Limit Theorems (Lindeberg-Levy)

$\{z_i\}$ i.i.d. with $E(z_i) = \mu$ and $Var(z_i) = \sigma^2$. Then for $\overline{z}_n = \frac{1}{n}\sum_{i=1}^{n}$:

$$\sqrt{n}(\overline{z}_n - \mu) \underset{d}{\rightarrow} N\left(0, \sigma^2\right) \quad \text{or}$$

$$\overline{z}_n - \mu \overset{a}{\sim} N\left(0, \frac{\sigma^2}{n}\right) \text{ or } \overline{z}_n \overset{a}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Remark: Read $\overset{a}{\sim}$ 'approximately distributed as'

CLT also holds for multivariate extension: sequence of random vectors $\{\mathbf{z}_i\}$

Useful lemmas of large sample theory

Lemma 1:

$\mathbf{z}_n \underset{p}{\to} \boldsymbol{\alpha}$ with a as a continuous function which does not depend on n then:

$$a(\mathbf{z}_n) \underset{p}{\to} a(\boldsymbol{\alpha}) \quad \text{or} \quad \underset{n\to\infty}{\text{plim}}\, a(\mathbf{z}_n) = a\left(\underset{n\to\infty}{\text{plim}}\,(\mathbf{z}_n)\right)$$

Examples:

$$x_n \underset{p}{\to} \alpha \quad \Rightarrow \quad \ln(x_n) \underset{p}{\to} \ln(\alpha)$$

$$x_n \underset{p}{\to} \beta \quad \text{and} \quad y_n \underset{p}{\to} \gamma \quad \Rightarrow \quad x_n + y_n \underset{p}{\to} \beta + \gamma$$

$$\mathbf{Y}_n \underset{p}{\to} \boldsymbol{\Gamma} \quad \Rightarrow \quad \mathbf{Y}_n^{-1} \underset{p}{\to} \boldsymbol{\Gamma}^{-1}$$

Useful lemmas of large sample theory (continued)

Lemma 2:

$\mathbf{z}_n \underset{d}{\rightarrow} \mathbf{z}$ then:

$a(\mathbf{z}_n) \underset{d}{\rightarrow} a(\mathbf{z})$

Examples:

$$z_n \underset{d}{\rightarrow} z, \quad z \sim N(0,1) \qquad \Rightarrow \qquad z^2 \sim \chi^2(1)$$

$$z_n \underset{d}{\rightarrow} N(0,1)$$

$$z^2 \underset{d}{\rightarrow} \chi^2(1)$$

Useful lemmas of large sample theory (continued)

Lemma 3:

$\mathbf{x}_n \underset{d}{\rightarrow} \mathbf{x}$ and $\mathbf{y}_n \underset{p}{\rightarrow} \boldsymbol{\alpha}$ then:

$$\mathbf{x}_n + \mathbf{y}_n \underset{d}{\rightarrow} \mathbf{x} + \boldsymbol{\alpha}$$

Examples:

$$x_n \underset{d}{\rightarrow} N(0,1), \quad y_n \underset{p}{\rightarrow} \alpha \quad \Rightarrow \quad x_n + y_n \underset{d}{\rightarrow} N(\alpha, 1)$$

$$\mathbf{x}_n \underset{d}{\rightarrow} \mathbf{x}, \quad \mathbf{y}_n \underset{p}{\rightarrow} \mathbf{0} \quad \Rightarrow \quad \mathbf{x}_n + \mathbf{y}_n \underset{d}{\rightarrow} \mathbf{x}$$

Lemma 4:

$\mathbf{x}_n \underset{d}{\rightarrow} \mathbf{x}$ and $\mathbf{y}_n \underset{p}{\rightarrow} \mathbf{0}$ then:

$$\mathbf{x}_n \cdot \mathbf{y}_n \underset{p}{\rightarrow} \mathbf{0}$$

Useful lemmas of large sample theory (continued)

Lemma 5:

$\mathbf{x}_n \underset{d}{\rightarrow} \mathbf{x}$ and $\mathbf{A}_n \underset{p}{\rightarrow} \mathbf{A}$ then:

$$\mathbf{A}_n \cdot \mathbf{x}_n \underset{p}{\rightarrow} \mathbf{A} \cdot \mathbf{x}$$

Example:

$$\mathbf{x}_n \underset{d}{\rightarrow} MVN(\mathbf{0}, \mathbf{\Sigma})$$

$$\mathbf{A}_n \cdot \mathbf{x}_n \underset{d}{\rightarrow} MVN(\mathbf{0}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}')$$

Lemma 6:

$\mathbf{x}_n \underset{d}{\rightarrow} \mathbf{x}$ and $\mathbf{A}_n \underset{p}{\rightarrow} \mathbf{A}$ then:

$$\mathbf{x}_n' \mathbf{A}_n^{-1} \mathbf{x}_n \underset{d}{\rightarrow} \mathbf{x}' \mathbf{A}^{-1} \mathbf{x}$$

Large sample assumptions for the OLS estimator

(2.1) Linearity: $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \quad \forall\, i = 1, 2, ..., n$

(2.2)/(2.5) Assumptions regarding dependence of $\{y_i, x_i\}$

(2.3) Orthogonality/predetermined regressors: $E(x_{ik} \cdot \varepsilon_i) = 0$
If $x_{ik} = 1 \;\; \Rightarrow \;\; E(\boldsymbol{\varepsilon}_i) = 0 \;\; \Rightarrow \;\; Cov(x_{ik}, \varepsilon_i) = 0$

(2.4) Rank condition: $E(\underset{K\times K}{\mathbf{x}_i\mathbf{x}_i'}) \equiv \boldsymbol{\Sigma}_{\mathbf{XX}}$ is non-singular

See Hayashi pp. 109-113

Large sample distribution of the OLS estimator

We get for $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$:

$$\underbrace{\mathbf{b}_n}_{} = \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{y}_i'$$

n indicates the dependence on the sample size

Under WLLN and lemma 1:

$$\mathbf{b}_n \underset{p}{\longrightarrow} \boldsymbol{\beta}$$

$$\sqrt{n}(\mathbf{b}_n - \boldsymbol{\beta}) \underset{d}{\longrightarrow} MVN\left(\mathbf{0}, Avar(\mathbf{b})\right) \quad \text{or} \quad \mathbf{b} \overset{a}{\sim} MVN\left(\boldsymbol{\beta}, \frac{Avar(\mathbf{b})}{n}\right)$$

$\Rightarrow$ $\mathbf{b}_n$ is **c**onsistent, **a**symptotically **n**ormal (**CAN**)

How to estimate $Avar(\mathbf{b})$

$Avar(\mathbf{b}) = \mathbf{\Sigma}_{\mathbf{xx}}^{-1} E(\mathbf{g}_i \mathbf{g}_i') \mathbf{\Sigma}_{\mathbf{xx}}^{-1}$ with $\mathbf{g}_i = \mathbf{X}_i \varepsilon_i$

$\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \underset{p}{\rightarrow} E(\mathbf{x}_i \mathbf{x}_i')$

Estimation of $E(\mathbf{g}_i \mathbf{g}_i')$: $\hat{\mathbf{S}} = \frac{1}{n} \sum e_i^2 \mathbf{x}_i \mathbf{x}_i' \underset{p}{\rightarrow} E(\mathbf{g}_i \mathbf{g}_i')$

$$\begin{aligned} \Rightarrow \quad \widehat{Avar(\mathbf{b})} &= \left[\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \hat{\mathbf{S}} \left[\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \underset{p}{\rightarrow} \\ Avar(\mathbf{b}) &= E(\mathbf{x}_i \mathbf{x}_i')^{-1} E(\mathbf{g}_i \mathbf{g}_i') E(\mathbf{x}_i \mathbf{x}_i')^{-1} \end{aligned}$$

Developing a test statistic under the assumption of conditional homoskedasticity

Assumption: $E(\varepsilon_i^2|\mathbf{x}_i) = \sigma^2$

$$\begin{aligned}\widehat{Avar(\mathbf{b})} &= \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right]^{-1} \hat{\sigma}^2 \ \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right]^{-1} \\ &= \hat{\sigma}^2 \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\end{aligned}$$

with $\hat{\mathbf{S}} = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'$

Note: $\frac{1}{n}\sum_{i=1}^{n} e_i^2$ is a biased estimate for σ^2

White standard errors

Adjusting the test statistics to make them robust against violations of conditional homoskedasticity

t-ratio

$$t_k = \frac{b_k - \bar{\beta}_k}{\sqrt{\left[\frac{\left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} e_i^2 \mathbf{x}_i \mathbf{x}_i' \left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'\right]^{-1}}{n} \right]_{kk}}} \overset{a}{\sim} N(0,1)$$

Holds under $H_0 : \beta_k = \overline{\beta}_k$

F-ratio

$$W = (\mathbf{Rb} - \mathbf{r})' \left[\mathbf{R} \frac{\widehat{Avar(\mathbf{b})}}{n} \mathbf{R}' \right]^{-1} (\mathbf{Rb} - \mathbf{r})' \overset{a}{\sim} \chi^2(\#r)$$

Holds under $H_0 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r} = 0$; allows for nonlinear restrictions on $\boldsymbol{\beta}$

We show that $\mathrm{b}_n = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is consistent

$$\mathbf{b}_n = [\tfrac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i']^{-1}\tfrac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{y}_i'$$

$$\Rightarrow \quad \underbrace{\mathbf{b}_n - \boldsymbol{\beta}}_{\text{sampling error}} = \left[\tfrac{1}{n}\sum \mathbf{x}_i\mathbf{x}_i'\right]^{-1}\tfrac{1}{n}\sum \mathbf{x}_i\boldsymbol{\varepsilon}_i$$

We show: $\mathbf{b}_n \underset{p}{\rightarrow} \boldsymbol{\beta}$

When sequence $\{\mathbf{y}_i, \mathbf{x}_i\}$ allows application of WLLN

$$\Rightarrow \quad \tfrac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' \underset{p}{\rightarrow} E(\mathbf{x}_i\mathbf{x}_i')$$

$$\tfrac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\boldsymbol{\varepsilon} \underset{p}{\rightarrow} E(\mathbf{x}_i\boldsymbol{\varepsilon}_i) \underset{p}{\rightarrow} 0$$

We show that $\mathbf{b}_n = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is consistent (continued)

Lemma 1 implies:

$$\begin{aligned} \mathbf{b}_n - \boldsymbol{\beta} &= \left[\frac{1}{n}\sum \mathbf{x}_i\mathbf{x}_i'\right]^{-1} \frac{1}{n}\sum \mathbf{x}_i\boldsymbol{\varepsilon}_i \\ &\underset{p}{\rightarrow} E(\mathbf{x}_i\mathbf{x}_i')^{-1}E(\mathbf{x}_i\boldsymbol{\varepsilon}_i) \\ &\underset{p}{\rightarrow} E(\mathbf{x}_i\mathbf{x}_i')^{-1}\cdot 0 = 0 \end{aligned}$$

$\mathbf{b}_n = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is consistent

We show that $\mathrm{b}_n = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is asymptotically normal

Sequence $\{\mathbf{g}_i\} = \{\mathbf{x}_i\varepsilon_i\}$ allows applying CLT for $\frac{1}{n}\sum \mathbf{x}_i\varepsilon_i = \overline{\mathbf{g}}$

$$\sqrt{n}(\overline{\mathbf{g}} - E(\mathbf{g}_i)) \underset{d}{\rightarrow} MVN(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{xx}}^{-1} E(\mathbf{g}_i\mathbf{g}_i')\mathbf{\Sigma}_{\mathbf{xx}}^{-1})$$

$$\sqrt{n}(\mathbf{b}_n - \boldsymbol{\beta}) = \left[\tfrac{1}{n}\sum \mathbf{x}_i\mathbf{x}_i'\right]^{-1} \sqrt{n}\,\overline{\mathbf{g}}$$

Applying lemma 5:

$$\mathbf{A}_n = \left[\tfrac{1}{n}\sum \mathbf{x}_i\mathbf{x}_i'\right]^{-1} \underset{p}{\rightarrow} \mathbf{A} = \mathbf{\Sigma}_{\mathbf{xx}}^{-1}$$

$$\mathbf{x}_n = \sqrt{n}\,\overline{\mathbf{g}} \underset{d}{\rightarrow} \mathbf{x} \underset{d}{\rightarrow} MVN(\mathbf{0}, E(\mathbf{g}_i\mathbf{g}_i'))$$

$\Rightarrow \quad \sqrt{n}(\mathbf{b}_n - \boldsymbol{\beta}) \underset{d}{\rightarrow} MVN(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{xx}}^{-1} E(\mathbf{g}_i\mathbf{g}_i')\mathbf{\Sigma}_{\mathbf{xx}}^{-1})$

$\Rightarrow \quad \mathbf{b}_n$ is CAN

8. Time Series Basics
(Stationarity and Ergodicity)

Hayashi p. 97-107

Dependence in the data

Certain degree of dependence in the data in time series analysis; only one realization of the data generating process is given

CLT and WLLN rely on i.i.d. data, but dependence in real world data

Examples:

- Inflation rate
- Stock market returns

Stochastic process: sequence of r.v.s. indexed by time $\{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, ...\}$ or $\{\mathbf{z}_i\}$ with $i = 1, 2, ...$

A realization/sample path: One possible outcome of the process

Dependence in the data - theoretical consideration

If we were able to 'run the world several times', we had different realizations of the process at one point in time

$\Rightarrow$ We could compute ensemble means and apply the WLLN

As the described repetition is not possible, we take the mean over the one realization of the process

Key question: Does $\frac{1}{T}\sum_{t=1}^{T} x_t \underset{p}{\rightarrow} E(x)$ hold?

Condition: Stationarity of the process

Definition of stationarity

Strict stationarity:

The joint distribution of $\mathbf{z}_i, \mathbf{z}_{i_1}, \mathbf{z}_{i_2}, ..., \mathbf{z}_{i_r}$ depends only on the relative position $i_1 - i, i_2 - i, ..., i_r - i$ but not on i itself

In other words: The joint distribution of $(\mathbf{z}_i, \mathbf{z}_{i_r})$ is the same as the joint distribution of $(\mathbf{z}_j, \mathbf{z}_{j_r})$ if $i - i_r = j - j_r$

Weak stationarity:

- $E(\mathbf{z}_i)$ does not depend on i
- $Cov(\mathbf{z}_i, \mathbf{z}_{i-j})$ depends on j (distance), but not on i (absolute position)

Ergodicity

A stationary process is also called ergodic if

$$\lim_{n\to\infty} E \quad [f(z_i, z_{i+1}, ..., z_{i+k}) \cdot g(z_{i+n}, z_{i+n+1}, ..., z_{i+n+l})] =$$
$$E \quad [f(z_i, z_{i+1}, ..., z_{i+k})] \cdot E\,[g(z_{i+n}, z_{i+n+1}, ..., z_{i+n+l})]$$

Ergodic Theorem:

Sequence $\{\mathbf{z}_i\}$ is stationary and ergodic with $E(\mathbf{z}_i) = \boldsymbol{\mu}$, then

$$\overline{\mathbf{z}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i \underset{p}{\rightarrow} \boldsymbol{\mu}$$

Martingale difference sequence

Stationarity and Ergodicity are not enough for applying the CLT. To derive the CAN property of the OLS-estimator we assume:

$$\{\mathbf{g}_i\} = \{\mathbf{x}_i\boldsymbol{\varepsilon}_i\}$$

$\{\mathbf{g}_i\}$ is a stationary and ergodic martingale difference sequence (m.d.s.):

$$E(\mathbf{g}_i|\mathbf{g}_{i-1}, \mathbf{g}_{i-2}, ..., \mathbf{g}_{i-j}) = 0$$

$$\Rightarrow E(\mathbf{g}_i) = \mathbf{0}$$

Implications of m.d.s. when $1 \in \mathbf{x}_i$:

ε_i and ε_{i-j} are uncorrelated, i.e. $Cov(\varepsilon_i, \varepsilon_{i-j}) = 0$

9. Generalized Least Squares

Hayashi p. 54-58

Assumptions of GLS

Linearity $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$

Full rank: rank$(\mathbf{X}) = K$

Strict exogeneity $E(\varepsilon_i|\mathbf{X}) = 0$

$\Rightarrow E(\varepsilon_i) = 0$ and $Cov(\varepsilon_i, x_{ik}) = E(\varepsilon_i x_{ik}) = 0$

NOT assumed: $Var(\boldsymbol{\varepsilon}|\mathbf{X}) = \sigma^2\mathbf{I}_n$

Instead:

$$Var(\boldsymbol{\varepsilon}|\mathbf{X}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}) = \begin{pmatrix} Var(\varepsilon_1|\mathbf{X}) & Cov(\varepsilon_1,\varepsilon_2|\mathbf{X}) & \dots & Cov(\varepsilon_1,\varepsilon_n|\mathbf{X}) \\ Cov(\varepsilon_1,\varepsilon_2|\mathbf{X}) & Var(\varepsilon_2|\mathbf{X}) & & \vdots \\ Cov(\varepsilon_1,\varepsilon_3|\mathbf{X}) & Cov(\varepsilon_2,\varepsilon_3|\mathbf{X}) & Var(\varepsilon_3|\mathbf{X}) & \\ \vdots & & \ddots & \vdots \\ Cov(\varepsilon_1,\varepsilon_n|\mathbf{X}) & & \dots & Var(\varepsilon_n|\mathbf{X}) \end{pmatrix}$$

$$\Rightarrow \; Var(\boldsymbol{\varepsilon}|\mathbf{X}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}) = \sigma^2\mathbf{V}(\mathbf{X})$$

Deriving the GLS estimator

Derived under the assumption that $\mathbf{V}(\mathbf{X})$ is known, symmetric and positive definite

$\Rightarrow \quad \mathbf{V}(\mathbf{X})^{-1} = \mathbf{C}'\mathbf{C}$

Transformation:

$$\tilde{\mathbf{y}} = \mathbf{C}\mathbf{y}$$

$$\tilde{\mathbf{X}} = \mathbf{C}\mathbf{X}$$

$$\Rightarrow \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\mathbf{C}\mathbf{y} = \mathbf{C}\mathbf{X}\boldsymbol{\beta} + \mathbf{C}\boldsymbol{\varepsilon}$$

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}$$

Least squares estimation of $\tilde{\beta}$ using transformed data

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{GLS} &= (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}} \\
&= (\mathbf{X}'\mathbf{C}'\mathbf{C}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}'\mathbf{C}\mathbf{y} \\
&= (\mathbf{X}'\frac{1}{\sigma^2}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\frac{1}{\sigma^2}\mathbf{V}^{-1}\mathbf{y} \\
&= \left[\mathbf{X}'[Var(\boldsymbol{\varepsilon}|\mathbf{X})]^{-1}\right]^{-1}\mathbf{X}'\left[Var(\boldsymbol{\varepsilon}|\mathbf{X})\right]^{-1}\mathbf{y}
\end{aligned}
$$

GLS estimator is the **b**est **l**inear **u**nbiased **e**stimator (**BLUE**)

Problems:

- Difficult to work out the asymptotic properties of $\hat{\boldsymbol{\beta}}_{GLS}$
- In real world applications $Var(\boldsymbol{\varepsilon}|\mathbf{X})$ not known
- If $Var(\boldsymbol{\varepsilon}|\mathbf{X})$ is estimated the BLUE-property of $\hat{\boldsymbol{\beta}}_{GLS}$ is lost

Special case of GLS - weighted least squares

$$E(\varepsilon\varepsilon'|\mathbf{X}) = Var(\varepsilon|\mathbf{X}) = \sigma^2 \begin{pmatrix} V_1(\mathbf{X}) & 0 & \dots & 0 \\ 0 & V_2(\mathbf{X}) & & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & V_N(\mathbf{X}) \end{pmatrix} = \sigma^2 V$$

As $\mathbf{V}(\mathbf{X})^{-1} = \mathbf{C}'\mathbf{C}$

$$\Rightarrow \quad \mathbf{C} = \begin{pmatrix} \frac{1}{\sqrt{V_1(\mathbf{X})}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{V_2(\mathbf{X})}} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{\sqrt{V_n(\mathbf{X})}} \end{pmatrix} = \begin{pmatrix} \frac{1}{s_1} & 0 & \dots & 0 \\ 0 & \frac{1}{s_2} & & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{s_n} \end{pmatrix}$$

$$\Rightarrow \quad \text{argmin} \sum_{i=1}^{n} \left(\frac{y_1}{s_i} - \hat{\beta}_1 s_i^{-1} - \hat{\beta}_2 \frac{\mathbf{x}_{i2}}{s_i} ... - \hat{\beta}_K \frac{\mathbf{x}_{iK}}{s_i} \right)^2$$

Observations are weighted by standard deviation

10. Multicollinearity

Exact multicollinearity

Expressing a regressor as linear combination of (an)other regressor(s)

rank($\mathbf{X}$) $\neq K$: No full rank

$\Rightarrow$ Assumption 1.3 or 2.4 is violated

$(\mathbf{X}'\mathbf{X})^{-1}$ does not exist

Often economic variables are correlated to some degree

BLUE result is not affected

Large sample results are not affected

$\lightning$ relative results

$\lightning$ $Var(\mathbf{b}|\mathbf{X})$ is affected in absolute terms

Effects of Multicollinearity and solutions to the problem

Effects:

- Coefficients may have high standard errors and low significance levels
- Estimates may have the wrong sign
- Small changes in the data produces wide swings in the parameter estimates

Solutions:

- Increasing precision by implementing more data. (Costly!)
- Building a better fitting model that leaves less unexplained.
- Excluding some regressors. (Dangerous! Omitted variable bias!)

11. Endogeneity

Hayashi p. 186-196

Omitted variable bias

Correctly specified model: $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$

Regression of $\mathbf{y}$ on $\mathbf{X}_1 \quad \Rightarrow \quad \mathbf{X}_2$ gets into the error term

$\Rightarrow$ Omitted variable bias

$$\begin{aligned}
\mathbf{b}_1 &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} \\
&= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta}_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\boldsymbol{\beta}_2 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\varepsilon}
\end{aligned}$$

OLS estimator is biased:

If $\boldsymbol{\beta}_2 \neq \mathbf{0} \quad \Rightarrow \quad (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\boldsymbol{\beta}_2 \neq \mathbf{0}$

If $(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2 \neq \mathbf{0} \quad \Rightarrow \quad (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\boldsymbol{\beta}_2 \neq \mathbf{0}$

Endogeneity bias: Working example

Simultaneous equations model of market equilibrium (structural form):

$$q_i^d = \alpha_0 + \alpha_1 p_i + u_i$$

$$q_i^s = \beta_0 + \beta_1 p_i + v_i$$

Clear markets: $q_i^d = q_i^s$

It is not possible to estimate $\alpha_0, \alpha_1, \beta_0, \beta_1$ as we do not know whether changes in the market equilibrium are due to supply or demand shocks

We observe many possible equilibria, however we can not explain the slope of the demand and the supply curve from the data

Endogeneity: Correlation between errors and regressors, regressors are not predetermined

Here: Simultaneous equation bias

From structural form to reduced form

Solving q_i and p_i yields reduced form:

$$p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{v_i + u_i}{\alpha_1 - \beta_1}$$

$$q_i = \frac{\alpha_1 \beta_0 + \alpha_0 \beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 v_i + \beta_1 u_i}{\alpha_1 - \beta_1}$$

Price is a function of the two error terms

v_i: supply shifter

u_i: demand shifter

Calculating the covariance of p_i and the demand shifter u_i:

$$\Rightarrow \quad Cov(p_i, u_i) = -\frac{Var(u_i)}{\alpha_1 - \beta_1}$$

If endogeneity is present the OLS-estimator is not consistent

FOCs in simple regression context yield:

$$\hat{\alpha}_1 = \frac{\frac{1}{n}\sum(q_i-\overline{q})(p_i-\overline{p})}{\frac{1}{n}\sum(p_i-\overline{p})^2} \underset{p}{\longrightarrow} \frac{Cov(p_i,q_i)}{Var(p_i)}$$

But here: $Cov(p_i, q_i) = \alpha_1 Var(p_i) + Cov(p_i, u_i)$

$\Rightarrow \quad \frac{Cov(p_i,q_i)}{Var(p_i)} = \alpha_1 + \frac{Cov(p_i,u_i)}{Var(p_i)} \neq \alpha_1$

$\Rightarrow$ OLS is not consistent

Same holds for β_1

Instruments for the market model

Properties of the instruments:

Uncorrelated with the errors, instruments are predetermined

Correlated with the endogenous regressors

$$Cov(x_i, p_i) = \frac{\beta_2}{\alpha_1 - \beta_1} Var(x_i)$$

$$Cov(x_i, u_i) = 0$$

$\Rightarrow$ x_i an instrument for p_i $\Rightarrow$ yields new reduced form

$$\Rightarrow \quad p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1} x_i + \frac{\zeta_i - u_i}{\alpha_1 - \beta_1}$$

$$\Rightarrow \quad q_i = \frac{\alpha_1 \beta_0 + \alpha_0 \beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 \beta_1}{\alpha_1 - \beta_1} x_i + \frac{\alpha_1 \zeta_i + \beta_1 u_i}{\alpha_1 - \beta_1}$$

$$Cov(x_i, q_i) = \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1} Var(x_i) = \alpha_1 Cov(x_i, p_i)$$

$\Rightarrow \quad \alpha_1 = \frac{Cov(x_i, q_i)}{Cov(x_i, p_i)}$ by WLLN $\hat{\alpha}_1 \underset{p}{\rightarrow} \alpha_1$

A simple macroeconometric model: Haavelmo (1943)

Aggregated consumption function: $C_i = \alpha_0 + \alpha_1 Y_i + u_i$

GDP identity: $Y_i = C_i + I_i$

Y_i affects C_i, but at the same time C_i influences Y_i

Reduced form: $Y_i = \frac{\alpha_0}{1-\alpha_1} + \frac{1}{1-\alpha_1} I_i + \frac{u_i}{1-\alpha_1}$

$\Rightarrow$ C_i can not be regressed on Y_i as the regressor is correlated with the residual:

$$Cov(Y_i, u_i) = \frac{Var(u_i)}{1-\alpha_1} > 0$$

$\Rightarrow$ OLS-estimator is inconsistent: upwards biased

$$\frac{Cov(C_i, Y_i)}{Var(Y_i)} = \alpha_1 + \frac{Cov(Y_i, u_i)}{Var(Y_i)} \neq \alpha_1$$

Valid instrument for income Y_i: investment I_i

Errors in variables

Explanatory variable is measured with error (e.g. reporting errors)

Classical example: Friedman's permanent income hypothesis

Permanent consumption is proportinal to permanent income $C_i^* = kY_i^*$

Observed variables:

$$Y_i = Y_i^* + y_i$$
$$C_i = C_i^* + c_i$$
$$c_i = ky_i + u_i$$

Endogeneity due to measurement errors

$\Rightarrow$ Solution: IV-estimators; here: $x_i = 1$

12. IV estimation

Hayashi p. 186-1196

General solution to endogeneity problem: IV

Linear regression: $\mathbf{y}_i = \mathbf{z}_i'\boldsymbol{\delta} + \boldsymbol{\varepsilon}_i$

But the assumption of predetermined regressors does not hold: $E(\mathbf{z}_i\boldsymbol{\varepsilon}_i) \neq 0$

$\Rightarrow$ To get consistent estimators instrumental variables $\mathbf{x}_i$ are needed:

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{bmatrix}$$

$\mathbf{x}_i$ is correlated with endogenous regressor but uncorrelated with error term

Assumptions for IV-estimators

3.1 Linearity $\mathbf{y}_i = \mathbf{z}_i'\boldsymbol{\delta} + \boldsymbol{\varepsilon}_i$

3.2 Ergodic stationarity

- K instruments $\mathbf{x}_i$
- L regressors $\mathbf{z}_i$
- Data sequence $\mathbf{w}_i \equiv \{\mathbf{y}_i, \mathbf{z}_i, \mathbf{x}_i\}$ is stationary and ergodic

3.3 Orthogonality conditions: $E(\mathbf{x_i}\boldsymbol{\varepsilon}_i) = 0$

$$
\begin{gathered}
E(x_{i1}(\mathbf{y}_i - \mathbf{z}_i'\boldsymbol{\beta})) = 0 \\
E(x_{i2}(\mathbf{y}_i - \mathbf{z}_i'\boldsymbol{\beta})) = 0 \\
\vdots \\
E(x_{iK}(\mathbf{y}_i - \mathbf{z}_i'\boldsymbol{\beta})) = 0
\end{gathered}
$$

$$\Rightarrow \quad E(\mathbf{x}_i(\mathbf{y}_i - \mathbf{z}_i'\boldsymbol{\beta})) = \mathbf{0}$$

Assumptions for IV-estimators (continued)

3.4 Rank condition for identification: rank($\mathbf{\Sigma}_{\mathbf{XZ}}$) $= L$ with

$$E(\mathbf{x}_i \mathbf{z}_i') = \begin{pmatrix} E(x_{i1} z_{i1}) & \dots & E(x_{i1} z_{iL}) \\ \vdots & \ddots & \vdots \\ E(x_{iK} z_{i1}) & \dots & E(x_{iK} z_{iL}) \end{pmatrix} = \mathbf{\Sigma}_{\mathbf{XZ}}$$

$\mathbf{\Sigma}_{\mathbf{XZ}}^{-1}$ exists

Core question: Do moment conditions provide enough information to determine $\boldsymbol{\delta}$ uniquely?

Deriving the IV-estimator

$$E(\mathbf{x}_i \boldsymbol{\varepsilon}_i) = E\left(\mathbf{x}_i(\mathbf{y}_i - \mathbf{z}_i \boldsymbol{\delta}\right) = \mathbf{0}$$

$$E(\mathbf{x}_i \mathbf{y}_i) - E(\mathbf{x}_i \mathbf{z}_i' \boldsymbol{\delta}) = \mathbf{0}$$

$$\boldsymbol{\delta} = \left[E(\mathbf{x}_i \mathbf{z}_i')\right]^{-1} E(\mathbf{x}_i \mathbf{y}_i)$$

$$\hat{\boldsymbol{\delta}}_{IV} = \left[\tfrac{1}{n}\sum \mathbf{x}_i \mathbf{z}_i'\right]^{-1} \left[\tfrac{1}{n}\sum \mathbf{x}_i \mathbf{y}_i\right] \underset{p}{\rightarrow} \boldsymbol{\delta}$$

If $K = L$ the rank condition implies that $\boldsymbol{\Sigma}_{\mathbf{XZ}}^{-1}$ exists and the system is exactly identified

Applying WLLN, CLT and the lemmas it can be shown that IV-estimator $\hat{\boldsymbol{\delta}}_{IV}$ is CAN