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 ∩ economic theory ∩ mathematics

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

\[\text{1} \text{according to Ragnar Frisch}\]
Example 1: Derivation of key equation of the Glosten-Harris model\textsuperscript{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: $\mu$
- Earnings/costs of the market maker: $c$
- Unobserved component (public information): $\varepsilon_t$

\textsuperscript{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: \( \varepsilon_t \)

Market maker sets:

Sell price (ask): \( P^a_t = \mu + m_{t-1} + \varepsilon_t + z_t + c \)
Buy price (bid): \( P^b_t = \mu + m_{t-1} + \varepsilon_t - z_t - c \)

\( \Rightarrow \) Transaction price change \( \Delta P_t = \mu + z_0 Q_t + z_1 v_t Q_t + c \Delta Q_t + \varepsilon_t \)

\( \Rightarrow \) Estimation of unknown structural parameters \( \beta = (\mu, z_0, z_1, c)' \)
Example 2: The influence of schooling on wages - Mincer equation

\[
\ln(WAGE_i) = \beta_1 + \beta_2 S_i + \beta_3 TENURE_i + \beta_4 EXP R_i + \varepsilon_i
\]

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: \( EXP R_i \)

\( \Rightarrow \) Estimation of the parameters \( \beta_k \), where \( \beta_2 \): return to schooling
General regression equations

Generalization:  \[ y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} + \varepsilon_i \]

Index for observations \( i = 1, 2, \ldots, n \) and regressors \( k = 1, 2, \ldots, K \)

\[
\begin{align*}
y_i & = \beta' \cdot x_i + \varepsilon_i \\
(1x1) & (1xK) & (Kx1) & (1x1)
\end{align*}
\]

\[
\beta = \begin{pmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_K 
\end{pmatrix}
\quad \text{and} \quad
x_i = \begin{pmatrix}
x_{i1} \\
x_{i2} \\
\vdots \\
x_{iK}
\end{pmatrix}
\]
Introduction Econometrics

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)

\[ y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} + \varepsilon_i = x_i' \cdot \beta + \varepsilon_i \]

\( y_i \): Dependent variable, observed
\( x_i' = (x_{i1}, x_{i2}, \ldots, x_{iK}) \): Explanatory variables, observed
\( \beta' = (\beta_1, \beta_2, \ldots, \beta_K) \): Unknown parameters
\( \varepsilon_i \): 'Disturbance' component, unobserved

\[ \Rightarrow b' = (b_1, b_2, \ldots, b_K) \text{ estimator of } \beta' \]
\[ \Rightarrow e_i = y_i - x_i' b : \text{Estimated residual} \]
For convenience we introduce matrix notation

\[
\begin{align*}
\mathbf{y} &= \mathbf{X} \cdot \mathbf{\beta} + \mathbf{\varepsilon} \\
\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} &= \begin{pmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1K} \\ 1 & x_{22} & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & x_{n2} & \cdots & x_{nK} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}
\end{align*}
\]
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 $b$ of $\beta$ by minimizing the SSR (sum of squared residuals):

$$
\arg\min_{\{b\}} S(b) = \arg\min \sum_{i=1}^{n} e_i^2 = \arg\min \sum_{i=1}^{n} (y_i - x'_i b)^2
$$

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

1. $\frac{\partial S(b)}{\partial b_1} = 0 \Rightarrow \sum e_i = 0$
2. $\frac{\partial S(b)}{\partial b_2} = 0 \Rightarrow \sum e_i x_{i2} = 0$
   
   $\vdots$
3. $\frac{\partial S(b)}{\partial b_K} = 0 \Rightarrow \sum e_i x_{iK} = 0$

$\Rightarrow$ FOC’s can be conveniently written in matrix notation $X' e = 0$
The system of $K$ equations is solved by matrix algebra

$$X' e = X'(y - Xb) = X'y - X'Xb = 0$$

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

$$(X'X)^{-1}X'y - (X'X)^{-1}X'Xb = 0$$

$$(X'X)^{-1}X'y - Ib = 0$$

OLS-estimator:

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

Alternatively:

$$b = \left(\frac{1}{n}X'X\right)^{-1}\frac{1}{n}X'y = \left(\frac{1}{n}\sum_{i=1}^{n}x_ix'_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}x_iy_i$$
**Zoom into the matrices** $X'X$ and $X'y$

\[
b = \left(\frac{1}{n}X'X\right)^{-1} \frac{1}{n}X'y = \left(\frac{1}{n} \sum_{i=1}^{n} x_ix_i'\right)^{-1} \frac{1}{n} \sum_{i=1}^{n} x_iy_i
\]

\[
\sum_{i=1}^{n} x_ix_i' = \begin{pmatrix}
\sum x_{i1}^2 & \sum x_{i1}x_{i2} & \sum x_{i1}x_{i3} & \cdots & \sum x_{i1}x_{iK} \\
\sum x_{i1}x_{i2} & \sum x_{i2}^2 & \sum x_{i1}x_{i3} & \cdots & \sum x_{i2}x_{iK} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\sum x_{i1}x_{iK} & \sum x_{i2}x_{iK} & \cdots & \cdots & \sum x_{iK}^2
\end{pmatrix}
\]

\[
\sum_{i=1}^{n} x_iy_i = \begin{pmatrix}
\sum x_{i1}y_i \\
\sum x_{i2}y_i \\
\sum x_{i3}y_i \\
\vdots \\
\sum x_{iK}y_i
\end{pmatrix}
\]
3. Assumptions of the CLRM

Hayashi p. 3-13
The four core assumptions of CLRM

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

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

$\Rightarrow E(\varepsilon_i) = 0$ 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: $Var(\varepsilon_i|X) = E(\varepsilon_i^2|X) = \sigma^2$

$Cov(\varepsilon_i, \varepsilon_j|X) = 0; \ E(\varepsilon_i \varepsilon_j|X) = 0$

$\Rightarrow E(\varepsilon_i) = \sigma_i^2$ and $Cov(\varepsilon_i, \varepsilon_j) = 0$ by LTE (see Hayashi p. 18)
Interpreting the parameters $\beta$ 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 $\beta_k$ 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 $\beta_k$ 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 $b = (X'X)^{-1}X'y$

1. $E(b) = \beta$: Unbiasedness of the estimator
   - Holds for any sample size
   - Holds under assumptions 1.1 - 1.3

2. $Var(b|X) = \sigma^2(X'X)^{-1}$: Conditional variance of $b$
   - Conditional variance depends on the data
   - Holds under assumptions 1.1 - 1.4

3. $Var(\hat{\beta}|X) \geq Var(b|X)$
   - $\hat{\beta}$ is any other linear unbiased estimator of $\beta$
   - Holds under assumptions 1.1 - 1.4
Some key results from mathematical statistics

\[
\mathbf{z} = \begin{pmatrix}
z_1 \\
z_2 \\
\vdots \\
z_n
\end{pmatrix}
\]

\[
\mathbf{A} = \begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{pmatrix}
\]

A new random variable:

\[
\mathbf{v} = \mathbf{A} \cdot \mathbf{z}
\]

\[
E(\mathbf{v}) = \begin{pmatrix}
E(v_1) \\
E(v_2) \\
\vdots \\
E(v_m)
\end{pmatrix}
= \mathbf{A} E(\mathbf{z})
\]

\[
Var(\mathbf{v}) = \mathbf{A} Var(\mathbf{z}) \mathbf{A}'
\]
The OLS estimator’s unbiasedness

\[ E(b) = \beta \implies E(b - \beta) = 0 \]

sampling error

\[
\begin{align*}
b - \beta &= (X'X)^{-1}X'y - \beta \\
&= (X'X)^{-1}X'(X\beta + \varepsilon) - \beta \\
&= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon - \beta \\
&= \beta + (X'X)^{-1}X'\varepsilon - \beta \\
&= (X'X)^{-1}X'\varepsilon
\end{align*}
\]

\[ \implies E(b - \beta | X) = (X'X)^{-1}X'E(\varepsilon | X) = 0 \quad \text{under assumption 1.2} \]

\[ \implies E_X(E(b | X)) = E_X(\beta) = E(b) \quad \text{by the LTE} \]
**Introductory Econometrics**

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

\[
\begin{align*}
\text{Var}(b|X) &= \text{Var}(b - \beta|X) \\
&= \text{Var}((X'X)^{-1}X'\varepsilon|X) = \text{Var}(A\varepsilon|X) \\
&= AV\text{ar}(\varepsilon|X)A' = A\sigma^2I_nA' \\
&= \sigma^2AI_nA' = \sigma^2AA' \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}
\end{align*}
\]

**Note:**

- \( \beta \) non-random
- \( b - \beta \) sampling error
- \( A = (X'X)^{-1}X' \)
- \( \text{Var}(\varepsilon|X) = \sigma^2I_n \)
Sketch of the proof of the Gauss Markov theorem

\[ \text{Var}(\hat{\beta}|X) \geq \text{Var}(b|X) \]

\[
\text{Var}(\hat{\beta}|X) = \text{Var}(\hat{\beta} - \beta|X) = \text{Var}[(D + A)\varepsilon|X] \\
= (D + A)\text{Var}(\varepsilon|X)(D' + A') = \sigma^2(D + A)(D' + A') \\
= \sigma^2(DD' + AD' + DA' + AA') = \sigma^2[DD' + (X'X)^{-1}] \\
\geq \sigma^2(X'X)^{-1} = \text{Var}(b|X)
\]

where \( C \) is a function of \( X \)

\[
\hat{\beta} = Cy \\
D = C - A \\
A \equiv (X'X)^{-1}X'
\]

Details of proof: Hayashi pages 29 - 30
The OLS estimator is BLUE

- OLS is linear
  - Holds under assumption 1.1
- OLS is unbiased
  - Holds under assumption 1.1 - 1.3
- OLS is the best estimator
  - Holds under the Gauss Markov theorem $Var(\hat{\beta}|X) \geq Var(b|X)$
5. Hypothesis Testing under Normality

Hayashi p. 33-45
Hypothesis testing

Economic theory provides hypotheses about parameters

⇒ If theory is right ⇒ testable implications

But: Hypotheses can’t be tested without distributional assumptions about $\varepsilon$

Distributional assumption: Normality assumption about the conditional distribution of $\varepsilon|X \sim \text{MVN}(0, \sigma^2 I_n)$ [Assumption 1.5]
Introductory Econometrics

Some facts from multivariate statistics

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

Expectation vector:

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

Variance-covariance matrix:

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

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

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

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

$$\Rightarrow \quad \mathbf{x} \sim \text{MVN}(\mathbf{\mu}, \Sigma) \Rightarrow \quad \mathbf{y} = \mathbf{c} + \mathbf{A} \mathbf{x} \sim \text{MVN}(\mathbf{c} + \mathbf{A} \mathbf{\mu}, \mathbf{A} \Sigma \mathbf{A}')$$
Application of the facts from multivariate statistics and the assumptions 1.1 - 1.5

\[ b - \beta = (X'X)^{-1}X'\varepsilon \]

Sampling error

Assuming \( \varepsilon|X \sim MVN(0, \sigma^2 I_n) \)

\[ \Rightarrow b - \beta|X \sim MVN \left( (X'X)^{-1}X'E(\varepsilon|X), (X'X)^{-1}X'\sigma^2 I_nX(X'X)^{-1} \right) \]

\[ \Rightarrow b - \beta|X \sim MVN \left( 0, \sigma^2(X'X)^{-1} \right) \]

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

OLS-estimator conditionally normal distributed if \( \varepsilon|X \) 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 $\varepsilon|X \sim MVN(0, \sigma^2 I_n) \Rightarrow$ alternative hypothesis:

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

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

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

Note: $[(X'X)^{-1}]_{kk}$ is the $k$-th row $k$-th column element of $(X'X)^{-1}$
Nuisance parameter $\sigma^2$ can be estimated

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

We don’t know $\varepsilon_i$ but we use the estimator $e_i = y_i - x_i'b$

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

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

$$E(\hat{\sigma}^2 | X) = \frac{n-K}{n} \sigma^2$$
An unbiased estimator of $\sigma^2$

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

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

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

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

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

$\Rightarrow$ t-statistic under $H_0$:

$t_k = \frac{b_k - \beta_k}{\sqrt{\hat{Var}(b|X)_{kk}}} = \frac{b_k - \beta_k}{\hat{SE}(b_k)} = \frac{b_k - \beta_k}{\sqrt{\hat{Var}(b_k|X)}} \sim t(n - K)$
**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 \( \alpha \) of two-sided test

4. Fix non-rejection and rejection regions \( \Rightarrow \) decision

Remark:

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

\[ \sqrt{s^2[(X'X)^{-1}]_{kk}} : \text{standard error } b_k | X \]
Testing joint hypotheses (F-test/Wald test)

Write hypothesis as:

\[ H_0 : \begin{bmatrix} R \\ \beta \end{bmatrix} = \begin{bmatrix} r \end{bmatrix} \]

\( R \): matrix of real numbers

\( r \): number of restrictions

Replacing the \( \beta = (\beta_1, \beta_2, \ldots, \beta_K) \) by estimator \( b = (b_1, b_2, \ldots, b_K)' \):

\[ R b = r \]
Definition of the F-test statistic

Properties of $Rb$:

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

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

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

Using some additional important facts from multivariate statistics

$z = (z_1, z_2, ..., z_m) \sim MVN(\mu, \Omega)$

$\Rightarrow (z - \mu)'\Omega^{-1}(z - \mu) \sim \chi^2(m)$

Result applied: Wald statistic

$(\mathbf{R}b - r)'[\sigma^2\mathbf{R}(X'X)^{-1}\mathbf{R}']^{-1}(\mathbf{R}b - r) \sim \chi^2(\#r)$
Properties of the F-test statistic

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

$\Rightarrow F$-ratio:

$$ F = \frac{(Rb - r)'[R(X'X)^{-1}R']^{-1}(Rb - r)/\#r}{(e'e)/(n-K)} $$

$$ = (Rb - r)'[\hat{R} Var(b|X)R']^{-1}(Rb - r)/\#r \sim F(\#r, n-K) $$

Note: F-test is one-sided

Proof: see Hayashi p. 41
Decision rule of the F-test

1. Specify $H_0$ in the form $R\beta = r$ and $H_A : R\beta \neq r$.

2. Calculate F-statistic.

3. Look up entry in the table of the F-distribution for $\#r$ and $n - K$ at given significance level.

4. Null is not rejected on the significance level $\alpha$ for $F$ less than $F_{\alpha}(\#r, n - K)$
Alternative representation of the F-statistic

Minimization of the unrestricted sum of squared residuals:

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

Minimization of the restricted sum of squared residuals:

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

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
Introductory Econometrics

Duality of t-test and confidence interval

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

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

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

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

Confidence interval for $\beta_k$:

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

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 $\beta_k$ 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 $\alpha$ 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: $\sum y_i^2 = y'y$

Decomposition of $y'y$:

\[
y'y = (\hat{y} + e)'(\hat{y} + e)
= \hat{y}'\hat{y} + 2\hat{y}e + e'e
= \hat{y}'\hat{y} + e'e
\]

$\Rightarrow \quad R_{uc}^2 \equiv 1 - \frac{e'e}{y'y}$

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 ($x_{i1} = 1$)

\[
\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} e_i^2
\]

\[
\Rightarrow \quad R^2_c \equiv 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{SSR}{SST}
\]

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
Introductory Econometrics

**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 $Var(\varepsilon | \mathbf{X}) = \sigma^2 \mathbf{I}$: 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
\{z_n\}: sequence of random vectors

Convergence in probability:

A sequence \(\{z_n\}\) converges in probability to a constant \(\alpha\) if for any \(\varepsilon > 0\)

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

Short-hand we write: \(\text{plim } z_n = \alpha\) or \(z_n \to^p \alpha\) or \(z_n - \alpha \to^p 0\)

Extends to random vectors:

If \(\lim_{n \to \infty} P \left( |z_{kn} - \alpha_k| > \varepsilon \right) = 0 \ \forall \ k = 1, 2, \ldots, K\), then \(z_n \to^p \alpha\)

(element-wise convergence)
Modes of convergence - Convergence in mean square and distribution

Convergence in mean square:

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

Convergence in mean square implies convergence in probability.

Convergence in distribution:

\[
z_n \to^d z
\]

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 \( \bar{z}_n = \frac{1}{n} \sum_{i=1}^{n} z_i \)

we have:

\[
\bar{z}_n \xrightarrow{p} \mu \quad \text{or} \\
\lim_{n \to \infty} P(\left|\bar{z}_{kn} - \mu\right| > \varepsilon) = 0 \quad \text{or} \\
\text{plim} \quad \bar{z}_n = \mu
\]
Extensions of the Weak Law of Large Numbers (WLLN)

The WLLN holds for:

Extension (1): Multivariate Extension (sequence of random vectors \( \{ z_i \} \))

Extension (2): Relaxation of independence

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

Extension (4): Vector valued functions \( f(z_i) \)
Central Limit Theorems (Lindeberg-Levy)

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

\[
\sqrt{n}(\bar{z}_n - \mu) \xrightarrow{d} N(0, \sigma^2)
\]
or

\[
\bar{z}_n - \mu \xrightarrow{a} N\left(0, \frac{\sigma^2}{n}\right) \text{ or } \bar{z}_n \xrightarrow{a} N\left(\mu, \frac{\sigma^2}{n}\right)
\]

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

CLT also holds for multivariate extension: sequence of random vectors \( \{z_i\} \)
Useful lemmas of large sample theory

Lemma 1:

\( z_n \rightarrow p \alpha \) with \( a \) as a continuous function which does not depend on \( n \) then:

\[
a(z_n) \rightarrow p a(\alpha) \quad \text{or} \quad \text{plim} \ a(z_n) = a \left( \text{plim} \ (z_n) \right)
\]

Examples:

\[
x_n \rightarrow p \alpha \quad \Rightarrow \quad \ln(x_n) \rightarrow p \ln(\alpha)
\]

\[
x_n \rightarrow p \beta \quad \text{and} \quad y_n \rightarrow p \gamma \quad \Rightarrow \quad x_n + y_n \rightarrow p (\beta + \gamma)
\]

\[
Y_n \rightarrow p \Gamma \quad \Rightarrow \quad Y_n^{-1} \rightarrow p \Gamma^{-1}
\]
Useful lemmas of large sample theory (continued)

Lemma 2:

\[ z_n \xrightarrow{d} z \text{ then:} \]

\[ a(z_n) \xrightarrow{d} a(z) \]

Examples:

\[ z_n \xrightarrow{d} z, \quad z \sim N(0, 1) \quad \Rightarrow \quad z^2 \sim \chi^2(1) \]

\[ z_n \xrightarrow{d} N(0, 1) \]

\[ z^2 \xrightarrow{d} \chi^2(1) \]
Useful lemmas of large sample theory (continued)

Lemma 3:
\[ x_n \xrightarrow{d} x \quad \text{and} \quad y_n \xrightarrow{p} \alpha \quad \text{then:} \]
\[ x_n + y_n \xrightarrow{d} x + \alpha \]

Examples:
\[ x_n \xrightarrow{d} N(0, 1), \quad y_n \xrightarrow{p} \alpha \quad \Rightarrow \quad x_n + y_n \xrightarrow{d} N(\alpha, 1) \]
\[ x_n \xrightarrow{d} x, \quad y_n \xrightarrow{p} 0 \quad \Rightarrow \quad x_n + y_n \xrightarrow{d} x \]

Lemma 4:
\[ x_n \xrightarrow{d} x \quad \text{and} \quad y_n \xrightarrow{p} 0 \quad \text{then:} \]
\[ x_n \cdot y_n \xrightarrow{p} 0 \]
Useful lemmas of large sample theory (continued)

Lemma 5:
\( x_n \xrightarrow{d} x \) and \( A_n \xrightarrow{p} A \) then:

\( A_n \cdot x_n \xrightarrow{p} A \cdot x \)

Example:

\( x_n \xrightarrow{d} MVN(0, \Sigma) \)

\( A_n \cdot x_n \xrightarrow{d} MVN(0, A\Sigma A') \)

Lemma 6:
\( x_n \xrightarrow{d} x \) and \( A_n \xrightarrow{p} A \) then:

\( x'_n A_n^{-1} x_n \xrightarrow{d} x' A^{-1} x \)
Introductory Econometrics

Large sample assumptions for the OLS estimator

(2.1) Linearity: $y_i = x_i' \beta + \varepsilon_i \ \forall \ i = 1, 2, \ldots, n$

(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 $x_{ik} = 1 \ \Rightarrow \ E(\varepsilon_i) = 0 \ \Rightarrow \ Cov(x_{ik}, \varepsilon_i) = 0$

(2.4) Rank condition: $E(x_i x_i') \equiv \Sigma_{XX}$ is non-singular

See Hayashi pp. 109-113
Large sample distribution of the OLS estimator

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

\[
\widehat{b}_n = \left[ \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right]^{-1} \frac{1}{n} \sum_{i=1}^{n} x_i y_i'
\]

\( n \) indicates the dependence on the sample size

Under WLLN and lemma 1:

\[
b_n \xrightarrow{p} \beta
\]

\[
\sqrt{n}(b_n - \beta) \xrightarrow{d} MVN\left(0, A\text{var}(b)\right) \quad \text{or} \quad b \sim MVN\left(\beta, \frac{A\text{var}(b)}{n}\right)
\]

\( \Rightarrow \) \( b_n \) is consistent, asymptotically normal (CAN)
**Introductory Econometrics**

**How to estimate \( Avar(b) \)**

\[
Avar(b) = \Sigma_{xx}^{-1} E(g_i g_i') \Sigma_{xx}^{-1} \quad \text{with} \quad g_i = X_i \varepsilon_i
\]

\[
\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \to E(x_i x_i')
\]

Estimation of \( E(g_i g_i') \):
\[
\hat{S} = \frac{1}{n} \sum e_i^2 x_i x_i' \to E(g_i g_i')
\]

\[
\Rightarrow \quad \hat{Avar}(b) = \left[ \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right]^{-1} \hat{S} \left[ \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right]^{-1} \to_p \quad Avar(b) = E(x_i x_i')^{-1} E(g_i g_i') E(x_i x_i')^{-1}
\]
Developing a test statistic under the assumption of conditional homoskedasticity

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

\[
\hat{\text{Avar}}(\mathbf{b}) = \left[ \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \hat{\sigma}^2 \left[ \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right]^{-1}
\]

\[
= \hat{\sigma}^2 \left[ \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right]^{-1}
\]

with \( \hat{S} = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \)

Note: \( \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \) is a biased estimate for \( \sigma^2 \)
White standard errors

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

t-ratio

\[ t_k = \frac{b_k - \bar{\beta}_k}{\sqrt{\left[ \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right]^{-1} \frac{1}{n} \sum_{i=1}^{n} e_i^2 x_i x_i' \left[ \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right]^{-1}}} \sim N(0, 1) \]

Holds under \( H_0 : \beta_k = \bar{\beta}_k \)

F-ratio

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

Holds under \( H_0 : R\beta - r = 0 \); allows for nonlinear restrictions on \( \beta \)
We show that \( b_n = (X'X)^{-1}X'y \) is consistent

\[
b_n = \left[ \frac{1}{n} \sum_{i=1}^{n} x_ix_i' \right]^{-1} \frac{1}{n} \sum_{i=1}^{n} x_iy_i'
\]

\[
\Rightarrow b_n - \beta = \left[ \frac{1}{n} \sum x_ix_i' \right]^{-1} \frac{1}{n} \sum x_i\varepsilon_i
\]

sampling error

We show: \( b_n \xrightarrow{p} \beta \)

When sequence \( \{y_i, x_i\} \) allows application of WLLN

\[
\Rightarrow \frac{1}{n} \sum_{i=1}^{n} x_ix_i' \xrightarrow{p} E(x_ix_i')
\]

\[
\frac{1}{n} \sum_{i=1}^{n} x_i\varepsilon \xrightarrow{p} E(x_i\varepsilon_i) \xrightarrow{p} 0
\]
We show that \( b_n = (X'X)^{-1}X'y \) is consistent (continued)

Lemma 1 implies:

\[
\begin{align*}
    b_n - \beta &= \left[ \frac{1}{n} \sum x_i'x_i \right]^{-1} \left[ \frac{1}{n} \sum x_i \varepsilon_i \right] \\
    &\xrightarrow{p} E(x_ix_i')^{-1}E(x_i\varepsilon_i) \\
    &\xrightarrow{p} E(x_ix_i')^{-1} \cdot 0 = 0
\end{align*}
\]

\( b_n = (X'X)^{-1}X'y \) is consistent
We show that \( b_n = (X'X)^{-1}X'y \) is asymptotically normal.

Sequence \( \{g_i\} = \{x_i\varepsilon_i\} \) allows applying CLT for \( \frac{1}{n} \sum x_i\varepsilon_i = \bar{g} \)

\[
\sqrt{n}(\bar{g} - E(g_i)) \xrightarrow{d} MVN(0, \Sigma_{xx}^{-1}E(g_i g'_i)\Sigma_{xx}^{-1})
\]

\[
\sqrt{n}(b_n - \beta) = \left[\frac{1}{n} \sum x_i x'_i\right]^{-1} \sqrt{n} \bar{g}
\]

Applying lemma 5:

\[
A_n = \left[\frac{1}{n} \sum x_i x'_i\right]^{-1} \xrightarrow{p} A = \Sigma_{xx}^{-1}
\]

\[
x_n = \sqrt{n} \bar{g} \xrightarrow{d} x \xrightarrow{d} MVN(0, E(g_i g'_i))
\]

\[
\Rightarrow \sqrt{n}(b_n - \beta) \xrightarrow{d} MVN(0, \Sigma_{xx}^{-1}E(g_i g'_i)\Sigma_{xx}^{-1})
\]

\[
\Rightarrow b_n \text{ is CAN}
\]
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 \( \{z_1, z_2, z_3, \ldots \} \) or \( \{z_i\} \) with \( i = 1, 2, \ldots \)

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

⇒ 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 \( \frac{1}{T} \sum_{t=1}^{T} x_t \xrightarrow{p} E(x) \) hold?

Condition: Stationarity of the process
Definition of stationarity

Strict stationarity:

The joint distribution of $z_i, z_{i_1}, z_{i_2}, \ldots, z_{i_r}$ depends only on the relative position $i_1 - i, i_2 - i, \ldots, i_r - i$ but not on $i$ itself.

In other words: The joint distribution of $(z_i, z_{i_r})$ is the same as the joint distribution of $(z_j, z_{j_r})$ if $i - i_r = j - j_r$.

Weak stationarity:

- $E(z_i)$ does not depend on $i$
- $Cov(z_i, z_{i-j})$ depends on $j$ (distance), but not on $i$ (absolute position)
Ergodicity

A stationary process is also called ergodic if

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

$$E \left[ f(z_i, z_{i+1}, \ldots, z_{i+k}) \right] \cdot E \left[ g(z_{i+n}, z_{i+n+1}, \ldots, z_{i+n+l}) \right]$$

Ergodic Theorem:

Sequence \( \{z_i\} \) is stationary and ergodic with \( E(z_i) = \mu \), then

$$\overline{z}_n = \frac{1}{n} \sum_{i=1}^{n} z_i \xrightarrow{p} \mu$$
Introductory Econometrics

Martingale difference sequence

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

\[ \{g_i\} = \{x_i \varepsilon_i\} \]

\(\{g_i\}\) is a stationary and ergodic martingale difference sequence (m.d.s.):

\[ E(g_i|g_{i-1}, g_{i-2}, \ldots, g_{i-j}) = 0 \]

\[ \Rightarrow E(g_i) = 0 \]

Implications of m.d.s. when 1 \(\in\) \(x_i\):

\(\varepsilon_i\) and \(\varepsilon_{i-j}\) are uncorrelated, i.e. \(Cov(\varepsilon_i, \varepsilon_{i-j}) = 0\)
9. Generalized Least Squares

Hayashi p. 54-58
Assumptions of GLS

Linearity \( y_i = x_i'\beta + \varepsilon_i \)

Full rank: \( \text{rank}(X) = K \)

Strict exogeneity \( E(\varepsilon_i|X) = 0 \)

\[ \Rightarrow E(\varepsilon_i) = 0 \text{ and } Cov(\varepsilon_i, x_{ik}) = E(\varepsilon_i x_{ik}) = 0 \]

NOT assumed: \( Var(\varepsilon|X) = \sigma^2 I_n \)

Instead:

\[
Var(\varepsilon|X) = E(\varepsilon\varepsilon'|X) = \\
\begin{pmatrix}
Var(\varepsilon_1|X) & Cov(\varepsilon_1, \varepsilon_2|X) & \cdots & Cov(\varepsilon_1, \varepsilon_n|X) \\
Cov(\varepsilon_1, \varepsilon_2|X) & Var(\varepsilon_2|X) & \cdots & \\
Cov(\varepsilon_1, \varepsilon_3|X) & Cov(\varepsilon_2, \varepsilon_3|X) & \cdots & \\
\vdots & \vdots & \ddots & \\
Cov(\varepsilon_1, \varepsilon_n|X) & & & Var(\varepsilon_n|X)
\end{pmatrix}
\]

\[ \Rightarrow Var(\varepsilon|X) = E(\varepsilon\varepsilon'|X) = \sigma^2 V(X) \]
Deriving the GLS estimator

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

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

Transformation:

$$\tilde{y} = Cy$$

$$\tilde{X} = CX$$

$$\Rightarrow y = X\beta + \varepsilon$$

$$Cy = CX\beta + C\varepsilon$$

$$\tilde{y} = \tilde{X}\beta + \tilde{\varepsilon}$$
Least squares estimation of $\tilde{\beta}$ using transformed data

$$\hat{\beta}_{GLS} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$$

$$= (X'C'CX)^{-1}X'C'Cy$$

$$= (X'\frac{1}{\sigma^2}V^{-1}X)^{-1}X'\frac{1}{\sigma^2}V^{-1}y$$

$$= [X'[Var(\varepsilon|X)]^{-1}]^{-1}X'[Var(\varepsilon|X)]^{-1}y$$

GLS estimator is the best linear unbiased estimator (BLUE)

Problems:

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

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

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

\[ \Rightarrow C = \begin{pmatrix} \frac{1}{\sqrt{V_1(X)}} & 0 & \ldots & 0 \\ 0 & \frac{1}{\sqrt{V_2(X)}} & & \vdots \\ \vdots & \ddots & 0 & \vdots \\ 0 & \ldots & 0 & \frac{1}{\sqrt{V_n(X)}} \end{pmatrix} = \begin{pmatrix} \frac{1}{s_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{s_2} & & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \ldots & 0 & \frac{1}{s_n} \end{pmatrix} \]

\[ \Rightarrow \text{argmin} \sum_{i=1}^{n} \left( \frac{y_i}{s_i} - \hat{\beta}_1 s_i^{-1} - \hat{\beta}_2 \frac{x_{i2}}{s_i} \ldots - \hat{\beta}_K \frac{x_{iK}}{s_i} \right)^2 \]

Observations are weighted by standard deviation
10. Multicollinearity
Exact multicollinearity

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

\[ \text{rank}(X) \neq K: \text{No full rank} \]
\[ \Rightarrow \text{Assumption 1.3 or 2.4 is violated} \]
\[ (X'X)^{-1} \text{ does not exist} \]

Often economic variables are correlated to some degree

BLUE result is not affected
Large sample results are not affected

‡ relative results
‡ \( \text{Var}(b|X) \) 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: $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$

Regression of $y$ on $X_1 \Rightarrow X_2$ gets into the error term

$\Rightarrow$ Omitted variable bias

$$b_1 = (X'_1X_1)^{-1}X'_1y$$
$$= (X'_1X_1)^{-1}X'_1(X_1/\beta_1 + X_2/\beta_2 + \varepsilon)$$
$$= \beta_1 + (X'_1X_1)^{-1}X'_1X_2/\beta_2 + (X'_1X_1)^{-1}X'_1\varepsilon$$

OLS estimator is biased:

If $\beta_2 \neq 0 \Rightarrow (X'_1X_1)^{-1}X'_1X_2/\beta_2 \neq 0$

If $(X'_1X_1)^{-1}X'_1X_2 \neq 0 \Rightarrow (X'_1X_1)^{-1}X'_1X_2/\beta_2 \neq 0$
Endogeneity bias: Working example

Simultaneous equations model of market equilibrium (structural form):

\[ q^d_i = \alpha_0 + \alpha_1 p_i + u_i \]
\[ q^s_i = \beta_0 + \beta_1 p_i + v_i \]

Clear markets: \( q^d_i = q^s_i \)

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:

\[ \hat{\alpha}_1 = \frac{1}{n} \sum (q_i - \bar{q})(p_i - \bar{p}) \cdot \frac{1}{n} \sum (p_i - \bar{p})^2 \rightarrow \frac{\text{Cov}(p_i, q_i)}{\text{Var}(p_i)} \]

But here: \( \text{Cov}(p_i, q_i) = \alpha_1 \text{Var}(p_i) + \text{Cov}(p_i, u_i) \)

\[ \Rightarrow \frac{\text{Cov}(p_i, q_i)}{\text{Var}(p_i)} = \alpha_1 + \frac{\text{Cov}(p_i, u_i)}{\text{Var}(p_i)} \neq \alpha_1 \]

\[ \Rightarrow \text{OLS is not consistent} \]

Same holds for \( \beta_1 \)
Instruments for the market model

Properties of the instruments:

Uncorrelated with the errors, instruments are predetermined
Correlated with the endogenous regressors

\[ Cov(x_i, p_i) = \frac{\beta_2}{\alpha_1 - \beta_1} Var(x_i) \]
\[ Cov(x_i, u_i) = 0 \]

⇒ \( x_i \) an instrument for \( p_i \) ⇒ yields new reduced form

\[ p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1} x_i + \frac{\xi_i - u_i}{\alpha_1 - \beta_1} \]

\[ q_i = \frac{\alpha_1 \beta_0 + \alpha_0 \beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 \beta_1}{\alpha_1 - \beta_1} x_i + \frac{\alpha_1 \xi_i + \beta_1 u_i}{\alpha_1 - \beta_1} \]

\[ Cov(x_i, q_i) = \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1} \] by WLLN \( \hat{\alpha}_1 \rightarrow \alpha_1 \)
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 proportional 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

⇒ Solution: IV-estimators; here: \( x_i = 1 \)
12. IV estimation

Hayashi p. 186-1196
General solution to endogeneity problem: IV

Linear regression: \( y_i = z_i' \delta + \varepsilon_i \)

But the assumption of predetermined regressors does not hold: \( E(z_i \varepsilon_i) \neq 0 \)

\( \Rightarrow \) To get consistent estimators instrumental variables \( x_i \) are needed:

\[
\begin{bmatrix}
  x_{i1} \\
  x_{i2} \\
  \vdots \\
  x_{iK}
\end{bmatrix}
\]

\( x_i \) is correlated with endogenous regressor but uncorrelated with error term
Assumptions for IV-estimators

3.1 Linearity $y_i = z_i' \delta + \varepsilon_i$

3.2 Ergodic stationarity
- $K$ instruments $x_i$
- $L$ regressors $z_i$
- Data sequence $w_i \equiv \{y_i, z_i, x_i\}$ is stationary and ergodic

3.3 Orthogonality conditions: $E(x_i \varepsilon_i) = 0$

\[
E(x_{i1}(y_i - z_i' \beta)) = 0 \\
E(x_{i2}(y_i - z_i' \beta)) = 0 \\
\vdots \\
E(x_{iK}(y_i - z_i' \beta)) = 0
\]

$\Rightarrow E(x_i(y_i - z_i' \beta)) = 0$
Assumptions for IV-estimators (continued)

3.4 Rank condition for identification: \( \text{rank} (\Sigma_{XZ}) = L \) with

\[
E(x_i z'_i) = \begin{pmatrix}
E(x_{i1} z_{i1}) & \ldots & E(x_{i1} z_{iL}) \\
\vdots & \ddots & \vdots \\
E(x_{iK} z_{i1}) & \ldots & E(x_{iK} z_{iL})
\end{pmatrix} = \Sigma_{XZ}
\]

\( \Sigma_{XZ}^{-1} \) exists

Core question: Do moment conditions provide enough information to determine \( \delta \) uniquely?
**Deriving the IV-estimator**

\[ E(x_i \varepsilon_i) = E \left( x_i (y_i - z_i \delta) \right) = 0 \]

\[ E(x_i y_i) - E(x_i z_i' \delta) = 0 \]

\[ \delta = \left[ E(x_i z_i') \right]^{-1} E(x_i y_i) \]

\[ \hat{\delta}_{IV} = \left[ \frac{1}{n} \sum x_i z_i' \right]^{-1} \left[ \frac{1}{n} \sum x_i y_i \right] \rightarrow \delta \]

If \( K = L \) the rank condition implies that \( \Sigma_{XZ}^{-1} \) exists and the system is exactly identified.

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