1 Appendix A: Matrix Algebra

1.1 Definitions

- Matrix $\mathbf{A} = [a_{ik}] = [\mathbf{A}]_{ik}$
- Symmetric matrix: $a_{ik} = a_{ki}$ for all i and k
- Diagonal matrix: $a_{ij} \neq 0$ if $i = j$ but $a_{ij} = 0$ if $i \neq j$
- Scalar matrix: the diagonal matrix of $a_{ii} = a$.
- Identity matrix: the scalar matrix of $a = 1$
- Triangular matrix: $a_{ij} = 0$ if $j > i$
- Idempotent matrix: $\mathbf{A} = \mathbf{AA} = \mathbf{A}^2$
- Symmetric idempotent matrix: $\mathbf{A}'\mathbf{A} = \mathbf{A} = \mathbf{AA}$
- Orthogonal matrix: $\mathbf{A}^{-1} = \mathbf{A}'$
- Unitary matrix: $\mathbf{A}'\mathbf{A} = \mathbf{AA}' = \mathbf{I}$
- Trace of $\mathbf{A}$: $tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$, sum of diagonal terms.
  - $tr(\mathbf{ABC}) = tr(\mathbf{BCA}) = tr(\mathbf{CBA})$ if $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are symmetric.
  - $tr(c\mathbf{A}) = c[tr(\mathbf{A})]$
  - $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$

Matrix Addition:

- $\mathbf{A} + \mathbf{B} = [a_{ik} + b_{ik}]$
- $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
- $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$

Matrix Multiplication

- $\mathbf{AB} \neq \mathbf{BA}$:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$$
$$\begin{bmatrix} e & f \\ g & h \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ae+cf & be+df \\ ag+ch & bg+dh \end{bmatrix}$$

- $(\mathbf{AB})\,\mathbf{C} = \mathbf{A}\,(\mathbf{BC})$

- $\mathbf{A}\,(\mathbf{B}+\mathbf{C}) = \mathbf{AB} + \mathbf{AC}$

- $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$

Idempotent (projection) Matrix

$$\mathbf{y} = b\mathbf{x} + c\mathbf{z} + \mathbf{u}$$

where $\mathbf{y}, \mathbf{x}, \mathbf{z}$ and $\mathbf{u}$ are $T \times 1$ vectors, b and c are scalars. Let

$$\mathbf{M}_z = \left(I - \mathbf{z}\,(\mathbf{z}'\mathbf{z})^{-1}\,\mathbf{z}'\right)$$

then, M_z is an idempotent matrix.

$$\begin{aligned}
\mathbf{M}_z\mathbf{M}_z &= \left(I - \mathbf{z}\,(\mathbf{z}'\mathbf{z})^{-1}\,\mathbf{z}'\right)\left(I - \mathbf{z}\,(\mathbf{z}'\mathbf{z})^{-1}\,\mathbf{z}'\right) \\
&= I - \mathbf{z}\,(\mathbf{z}'\mathbf{z})^{-1}\,\mathbf{z}' - \mathbf{z}\,(\mathbf{z}'\mathbf{z})^{-1}\,\mathbf{z}' + \mathbf{z}\,(\mathbf{z}'\mathbf{z})^{-1}\,\mathbf{z}'\mathbf{z}\,(\mathbf{z}'\mathbf{z})^{-1}\,\mathbf{z}' \\
&= I - \mathbf{z}\,(\mathbf{z}'\mathbf{z})^{-1}\,\mathbf{z}' - \mathbf{z}\,(\mathbf{z}'\mathbf{z})^{-1}\,\mathbf{z}' + \mathbf{z}\,(\mathbf{z}'\mathbf{z})^{-1}\,\mathbf{z}' \\
&= I - \mathbf{z}\,(\mathbf{z}'\mathbf{z})^{-1}\,\mathbf{z}' = M_z
\end{aligned}$$

Further note that

$$\mathbf{M}_z\mathbf{z} = \left(I - \mathbf{z}\,(\mathbf{z}'\mathbf{z})^{-1}\,\mathbf{z}'\right)\mathbf{z} = 0$$

Hence we have

$$\begin{aligned}
\mathbf{M}_z\mathbf{y} &= b\mathbf{M}_z\mathbf{x} + c\mathbf{M}_z\mathbf{z} + \mathbf{M_z u} \\
&= b\mathbf{M}_z\mathbf{x} + \mathbf{M_z u}
\end{aligned}$$

Vector

- Lenth of a vector: Norm is defined as

$$||\mathbf{e}|| = \sqrt{\mathbf{e}'\mathbf{e}} = \left(\sum_{i=1}^{n} e_i^2\right)^{1/2}$$

- Orthogonal vectors: Two nonzero vectors $\mathbf{a}$ and $\mathbf{b}$ are orthogonal, written $\mathbf{a}\perp\mathbf{b}$, iff

$$\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a} = 0$$

Regression in a Matrix form

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u}$$

The OLS estimate is

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

The OLS residuals are

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\mathbf{b}} = \mathbf{X}\mathbf{b} - \mathbf{X}\hat{\mathbf{b}} + \mathbf{u} = \mathbf{u} - \mathbf{X}\left(\hat{\mathbf{b}} - \mathbf{b}\right)$$

Hence we have

$$\begin{aligned}
\mathbf{X}'\hat{\mathbf{u}} &= \mathbf{X}'\left(\mathbf{u} - \mathbf{X}\left(\hat{\mathbf{b}} - \mathbf{b}\right)\right) \\
&= \mathbf{X}'\mathbf{u} - \mathbf{X}'\mathbf{X}\left(\hat{\mathbf{b}} - \mathbf{b}\right) \\
&= \mathbf{X}'\mathbf{u} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\
&= \mathbf{0}
\end{aligned}$$

Matrix Inverse

- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- $\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} \end{bmatrix}$

- $\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1}$ =? (see p.966 A-74)

Kronecker Products Let $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} \end{bmatrix}$$

- $\mathbf{A}$ is $K \times L$ and $\mathbf{B}$ is $m \times n$. Then $\mathbf{A} \otimes \mathbf{B}$ is $(Km) \times (Ln)$
- $\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C}$
- $(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C}$
- $(k\mathbf{A}) \otimes \mathbf{B} = \mathbf{A} \otimes (k\mathbf{B}) = k\,(\mathbf{A} \otimes \mathbf{B})$ where k is a scalar
- $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$
- $(\mathbf{A} \otimes \mathbf{B})^{-1} = \left(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}\right)$
- $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$
- $\text{tr}(\mathbf{A} \otimes \mathbf{B}) = \text{tr}(\mathbf{A})\text{tr}(\mathbf{B})$
- $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$

1.2 Eigen values and Eigen vectors

Eigen Values Characteristic vectors = Eigen vectors, $\mathbf{c}$

Characteristic roots = Eigen values. λ

$$\begin{aligned} \mathbf{Ac} &= \lambda\mathbf{c} \\ \mathbf{Ac} - \lambda\mathbf{Ic} &= \mathbf{0} \\ |\mathbf{A} - \lambda\mathbf{I}|\,\mathbf{c} &= \mathbf{0} \end{aligned}$$

Example: Find eigen values of $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}$$

$$|\mathbf{A} - \lambda \mathbf{I}| = \begin{vmatrix} 1-\lambda & 3 \\ 0 & 2-\lambda \end{vmatrix} = 0$$

Solutions:

$$\lambda = 1, 2$$

Eigen Vector: The characteristic vectors of a symmetric matrix are orthogonal. That is,

$$\mathbf{C}'\mathbf{C} = \mathbf{I}$$

where $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_K]$. Alternatively $\mathbf{C}$ is a unitary matrix.

Let

$$\mathbf{\Lambda} = diag(\lambda_1, ..., \lambda_K) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & & \\ 0 & 0 & \cdots & \lambda_K \end{bmatrix}$$

Then we have

$$\mathbf{A}\mathbf{c}_k = \lambda_k \mathbf{c}_k$$

or

$$\begin{aligned} \mathbf{AC} &= \mathbf{C\Lambda} \\ \mathbf{C}'\mathbf{AC} &= \mathbf{C}'\mathbf{C\Lambda} = \mathbf{\Lambda} = diag(\lambda_1, ..., \lambda_K) \end{aligned}$$

Alternatively we have

$$\mathbf{A} = \mathbf{C\Lambda C}'$$

which is the spectral decomposition of $\mathbf{A}$

Some Facts: Prove the followings

1. $tr(\mathbf{A}) = tr(\mathbf{\Lambda})$

$$tr(\mathbf{A}) = tr(\mathbf{C\Lambda C'}) = tr(\mathbf{\Lambda CC'}) = tr(\mathbf{\Lambda I}) = tr(\mathbf{\Lambda})$$

   The trace of a matrix equals the sum of its eigen values.

2. $|\mathbf{A}| = |\mathbf{\Lambda}|$

3. $\mathbf{AA} = \mathbf{A}^2 = \mathbf{C\Lambda}^2\mathbf{C}'$

4. $\mathbf{A}^{-1} = \mathbf{C\Lambda}^{-1}\mathbf{C}'$

5. Suppose that $\mathbf{A}$ is a nonsigular symmetric matrix. Then

$$\mathbf{A}^{1/2} = \mathbf{C\Lambda}^{1/2}\mathbf{C}'$$

6. Consider a matrix $\mathbf{P}$ such that

$$\mathbf{P'P} = \mathbf{A}^{-1}$$

   then

$$\mathbf{P} = \mathbf{\Lambda}^{-1/2}\mathbf{C}'$$

Matrix Decomposition LU decomposition (Cholesky Decomposition)

$$\mathbf{A} = \mathbf{LU}$$

where $\mathbf{L}$ is lower triangular and $\mathbf{U}$ is upper triangular matrix. $\mathbf{L} = \mathbf{U}'$

Example:

$$\begin{bmatrix} a & c \\ c & b \end{bmatrix} = \begin{bmatrix} e & f \\ 0 & g \end{bmatrix} \begin{bmatrix} e & 0 \\ f & g \end{bmatrix} = \begin{bmatrix} f^2 + e^2 & gf \\ gf & g^2 \end{bmatrix}$$

Hence the solution is given by

$$g = \sqrt{b}, f = c/\sqrt{b}, a = ?$$

Spectral (Eigen) Decomposition

$$\mathbf{A} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}'$$

Schur Decomposition

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{U}'$$

where $\mathbf{U}$ is an orthogonal matrix and $\mathbf{S}$ is a upper triangular matrix.

Quadratic forms Let A be a symmetric matrix. Then all eigen values of A are positive (negative), then A is a positive (negative) definite matrix. If A has both negative and positive eigen values, then A is indefinite.

1.3 Matrix Algebra

$$\begin{aligned}
\frac{\partial\left(\mathbf{A}\mathbf{x}\right)}{\partial\mathbf{x}} &= \mathbf{A}, \ \frac{\partial\left(\mathbf{A}\mathbf{x}\right)}{\partial\mathbf{x}'} = \mathbf{A}' \\
\frac{\partial\left(\mathbf{x}'\mathbf{A}\mathbf{x}\right)}{\partial\mathbf{x}} &= 2\mathbf{A}\mathbf{x} \\
\frac{\partial\left(\mathbf{x}'\mathbf{A}\mathbf{x}\right)}{\partial\mathbf{A}} &= \mathbf{x}\mathbf{x}'
\end{aligned}$$

1.4 Sample Questions:

Part I: Calculation $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Q1: Find eigen values of A

Q2: Find the lower triangular matrix of A

Q3: $\mathbf{A} \otimes \mathbf{B}$

Q4: $(\mathbf{A} \otimes \mathbf{B})^{-1}$

Q5: tr(A)

Part II: Matrix Algebra Consider the following regression

$$y_i = a + bx_{1i} + cx_{2i} + u_i \tag{1}$$

Q6: If you wrote (9) as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{2}$$

Define y,X, β, and u

Q7: Consider the following problem

$$\arg\min_{\beta} L = \mathbf{u}'\mathbf{u}$$

Show the first derivertives of L function.

Q8: Show the solution satisfies $\boldsymbol{\beta} = (X'X)^{-1} \mathbf{X}'\mathbf{y}$

2 Probability and Distribution Theory

2.1 Probability Distributions

$$f(x) = \text{Prob}(X = x)$$

1. $0 \leq \text{Prob}(X = x) \leq 1$,
2. $\sum_x f(x) = 1$ (discrete case), $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (continuous case)

Cumulative Distribution Function

$$F(x) = \begin{cases} \sum_{X \leq x} f(X) = \text{Prob}(X \leq x) : \text{ discrete} \\ \int_{-\infty}^{x} f(x)\,dx = \text{Prob}(X \leq x) : \text{ continuous} \end{cases}$$

1. $0 \leq F(x) \leq 1$
2. If $x > y$, then $F(x) \geq F(y)$
3. $F(+\infty) = 1$
4. $F(-\infty) = 0$

Expectations of a Random Variable Mean or expected value of a random variable is

$$E[x] = \begin{cases} \sum_x x f(x) \text{ discrete} \\ \int_x x f(x)\,dx \text{ continuous} \end{cases}$$

Median: used when the distribution is not symmetric

Mode: the value of x at which $f(x)$ take its maximum

Functional expectation Let $g(x)$ be a function of x. Then

$$E[g(x)] = \begin{cases} \sum_x g(x) f(x) & \text{discrete} \\ \int_x g(x) f(x)\, dx & \text{continuous} \end{cases}$$

Variance

$$\begin{aligned} V(x) &= E(x-\mu)^2 \\ &= \begin{cases} \sum_x (x-\mu)^2 f(x) & \text{discrete} \\ \int_x (x-\mu)^2 f(x)\, dx & \text{continuous} \end{cases} \end{aligned}$$

Note that

$$\begin{aligned} V(x) &= E(x-\mu)^2 = E\left(x^2 - 2x\mu + \mu^2\right) \\ &= E\left(x^2\right) - \mu^2 \end{aligned}$$

Now we consider third and fourth central moments

$$\begin{aligned} \text{Skewness} &: \quad E(x-\mu)^3 \\ \text{Kurtosis} &: \quad E(x-\mu)^4 \end{aligned}$$

Skewness is a measure of the asymmetry of a distribution. For symmetric distribution, we have

$$f(x-\mu) = f(x+\mu)$$

and

$$E(x-\mu)^3 = 0$$

2.2 Some Specific Probabilty Distributions

2.2.1 Normal Distribution

$$f\left(x|\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

or we note

$$x \sim N\left(\mu, \sigma^2\right)$$

Properties:

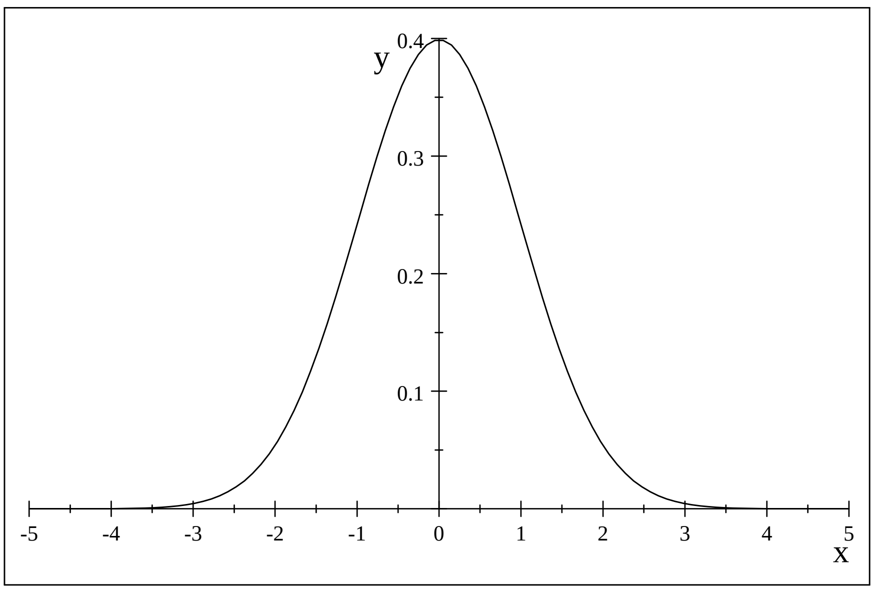
1. Addition and Multiplication

$$x \sim N\left(\mu, \sigma^2\right), \; a + bx \sim N\left(a + b\mu, b^2\sigma^2\right)$$

2. Standard normal function

$$\begin{aligned} x &\sim N(0,1) \\ f(x|0,1) &= \phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \end{aligned}$$

$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$



Chi-squared, t and F distributions

$$f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k-2}{2}} \exp\left(-\frac{x}{2}\right) 1[x \geq 0]$$

1. If $x \sim N(0,1)$, then $x^2 \sim \chi_1^2$ chi-squared with one degress of freedom.

2. If $x_1, ..., x_n$ are n independent χ_1^2 variables, then

$$\sum_{i=1}^{n} x_i \sim \chi_n^2$$

3. If $x_1, ..., x_n$ are n independent $N(0,1)$ variables, then

$$\sum_{i=1}^{n} x_i^2 \sim \chi_n^2$$

4. If $x_1, ..., x_n$ are n independent $N(0,\sigma^2)$ variables, then

$$\sum_{i=1}^{n} \frac{x_i^2}{\sigma^2} \sim \chi_n^2$$

5. If x_1 and x_2 are independent χ_n^2 and χ_m^2 variables, then

$$x_1 + x_2 \sim \chi_{n+m}^2$$

6. If x_1 and x_2 are independent χ_n^2 and χ_m^2 variables, then

$$\frac{x_1/n}{x_2/m} \sim F(n,m)$$

7. If z is a $N(0,1)$ variable and x is χ_n^2 and is independent of z, then the ratio

$$t_n = \frac{z}{\sqrt{x/n}}$$

   and it has the density function given by

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}}$$

   where $v = n-1$ and $\Gamma(.)$ is the gamma function

$$\Gamma(p) = \int_0^{\infty} t^{p-1} e^{-t} dt.$$

8. $t_n \rightarrow N(0,1)$ as $n \rightarrow \infty$

9. If $x \sim t_n$, then $x^2 \sim F(1,n)$

10. Noncentral χ^2 distribution: If $x \sim N(\mu, \sigma^2)$, then $(x/\sigma)^2$ has a noncentral χ_1^2 distribution.

11. If x and z have a joint normal distribution, then $\mathbf{w} = (x, z)'$ has

$$\mathbf{w} \sim N(0, \mathbf{\Sigma})$$

where

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xz} \\ \sigma_{xz} & \sigma_z^2 \end{bmatrix}$$

12. If $\mathbf{w} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$ where $\mathbf{w}$ has J elements, then $\mathbf{w}'\mathbf{\Sigma}^{-1}\mathbf{w}$ has a noncentral χ_J^2. The noncentral parameter is $\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}/2$.

2.2.2 Other Distributions

Lognormal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{1}{2}\left[\frac{\ln x - \mu}{\sigma}\right]^2\right)$$

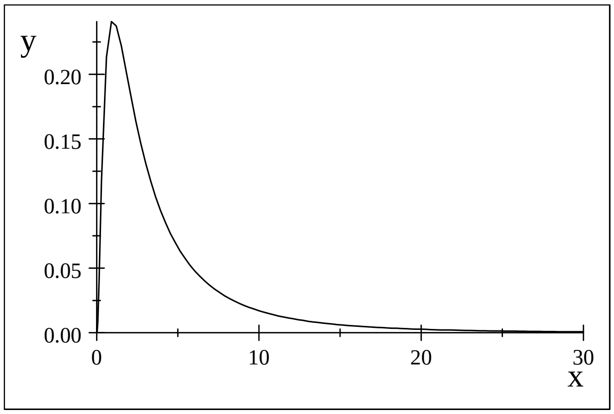
Note that

$$\begin{aligned} E(x) &= \exp\left(\mu + \frac{\sigma^2}{2}\right) \\ V(x) &= e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) \end{aligned}$$

Hence

$$\begin{aligned} \mu &= \ln E(x) - \frac{1}{2}\ln\left(1 + \frac{V(x)}{E(x)^2}\right) \\ \sigma^2 &= \ln\left(1 + \frac{V(x)}{E(x)^2}\right) \\ \mathrm{Mode}(x) &= e^{\mu-\sigma^2} \\ \mathrm{Median}(x) &= e^{\mu} \end{aligned}$$

$\frac{1}{\sqrt{2\pi}x}\exp\left(-\frac{1}{2}[\ln x - 1]^2\right)$



- Properties

1. If $x \sim N(\mu, \sigma^2)$, then $\exp(x) \sim LN(\mu, \sigma^2)$

2. If $x \sim LN(\mu, \sigma^2)$, then $\ln(x) \sim N(\mu, \sigma^2)$

3. If $x \sim LN(\mu, \sigma^2)$, then $y = x + c$ is a shifted LN of x. $E(y) = E(x) + c$, $V(y) = V(x + c) = V(x)$

4. If $x \sim LN(\mu, \sigma^2)$, then $y = ax$ is also LN. $y \sim LN(\ln a + \mu, \sigma^2)$

5. If $x \sim LN(\mu, \sigma^2)$, then $y = 1/x$ is also LN. $y \sim LN(-\mu, \sigma^2)$

6. If $x \sim LN(\mu_1, \sigma_1^2)$, $y \sim LN(\mu_2, \sigma_2^2)$ and they are independent, then

$$xy \sim LN\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$$

Gamma Distribution

$$f(x) = \frac{\lambda^p}{\Gamma(p)} \exp(-\lambda x) x^{p-1} \text{ for } x \geq 0, \lambda > 0, p > 0$$

where

$$\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt.$$

Note that if p is a positive integer, then

$$\Gamma(p) = (p-1)!$$

When $p = 1$, gamma distribution becomes exponential distribution

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) \text{ for } x \geq 0, \\ 0 \text{ for } x < 0 \end{cases}$$

When $p = \frac{n}{2}, \lambda = 1/2$, gamma dist. $= \chi^2$ dist.

When p is a positive integer, gamma dist. is called Erlang family.

Beta distribution For a variable constrained between 0 and $c > 0$, the beta distribution has its density as

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{x}{c}\right)^{\alpha-1} \left(1 - \frac{x}{c}\right)^{\beta-1} \frac{1}{c}$$

Usually x's range becomes $x \in (0, 1)$ that is $c = 1$.

1. symmetric if $\alpha = \beta$
2. $\alpha = 1, \beta = 1$, becomes $U(0,1)$
3. $\alpha < 1, \beta < 1$, becomes $U - shape$
4. $\alpha = 1, \beta > 2$, strictly convex
5. $\alpha = 1, \beta = 2$, straight line
6. $\alpha = 1, 1 < \beta < 2$, strictly concave
7. Mean: $\frac{c\alpha}{\alpha+\beta}$, $\frac{\alpha}{\alpha+\beta}$ for $c = 1$
8. Variance: $\frac{c^2\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$, $\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$ for $c = 1$

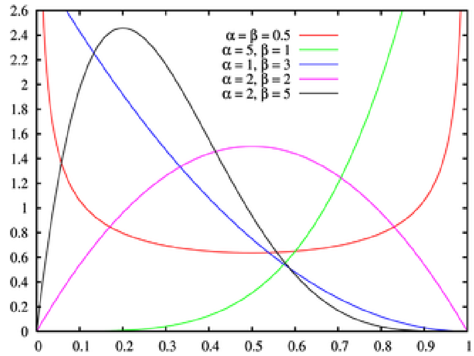


Figure 1:

Logistic Distribution

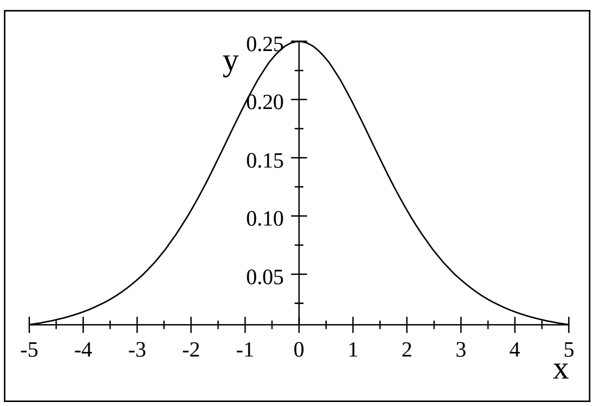
$$f(x) = \frac{e^{-(x-\mu)/s}}{s\left(1+e^{-(x-\mu)/s}\right)^2},\ \ s>0$$

$$F(x) = \frac{1}{1+e^{-(x-\mu)/s}}$$

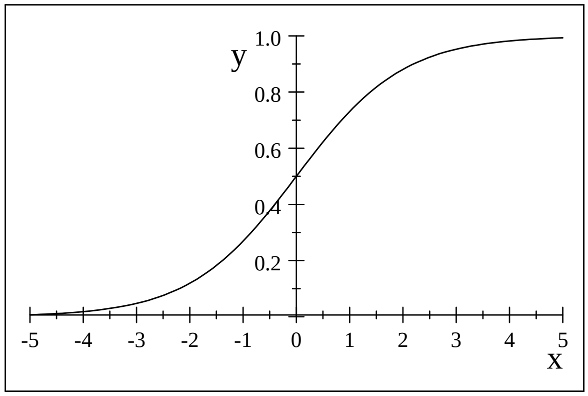
When $s=1$ and $\mu=0$, we have

$$f(x) = \frac{e^{-x}}{\left(1+e^{-x}\right)^2},\ \ F(x) = \frac{1}{1+e^{-x}}$$

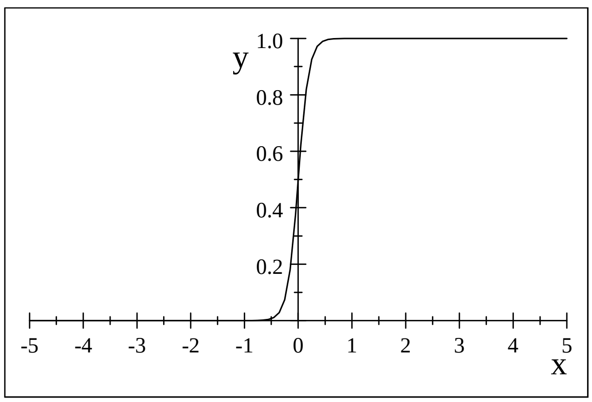
$\frac{e^{-x}}{(1+e^{-x})^2}$



$\frac{1}{1+e^{-x}}$



$\frac{1}{1+e^{-x/0.1}}$



Exponential distribution

$$f(x) = \lambda e^{-\lambda x}$$

Weibull Distribution

$$f(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} \text{ for } x \geq 0 \\ 0 \text{ for } x < 0 \end{cases}$$

When $k = 1$, Weibell becomes the exponential distribution.

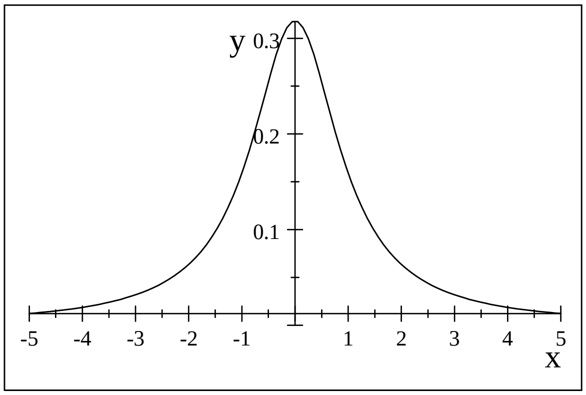
Cauchy Distribution

$$f(x) = \frac{1}{\pi}\left[\frac{\gamma}{(x-x_0)^2+\gamma^2}\right]$$

where x_0 is the location parameter, γ is the scale parameter. The standard Cauchy distribution is the case where $x_0 = 0$ and $\gamma = 1$.

$$f(x) = \frac{1}{\pi(1+x^2)}$$

$\frac{1}{\pi(1+x^2)}$



Note that the t distribution with $v = 1$ becomes a standard Cauchy distribution. Also note that the mean and variance of the Cauchy distribution don't exist.

2.3 Representations of A Probability Distribution

Survival Function

$$S(x) = 1 - F(x) = \text{Prob}[X \geq x]$$

where X is a continuous random variable.

Hazard Function (Failure rate)

$$h(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{1-F(x)}$$

Let $f(t) = \lambda e^{-\lambda t}$ (exponential density function), then we have

$$h(t) = \frac{f(t)}{S(t)} = \lambda$$

which implies that the harzard rate is a constant with respect to time. However for Weibull distribution or log normal distribution, the harzard function is not a constant any more.

Moment Generating Function (mdf) The mgf of a random variable x is

$$M_x(t) = E\left(e^{tx}\right), \quad \text{for } t \in R$$

Note that mgf is an alternate definition of probability distribution. Hence there is one for one relationship between the pdf and mgf. However mgf does not exist sometimes. For example, the mgf for the Cauchy distribution is not able to be defined.

Characteristic Function (cf) Alternatively, the following characteristic function is used frequently in Finance to define probability function. Even when the mdf does not exist, cf always exist.

$$\phi(t) = E\left(e^{itx}\right), \quad i^2 = -1$$

For example, the cf for the Cauchy distribution is $\exp\left(x_0 it - \gamma\left|t\right|\right)$.

Cumulants The cumulants κ_n of a random variable x are defined by the cumulant generating function which is the logarithm of the mgf.

$$g(t) = \log\left[E\left(e^{tx}\right)\right]$$

Then, the cumulants are gievn by

$$\begin{aligned}
\kappa_1 &= \mu = g'(0) \\
\kappa_2 &= \sigma^2 = g''(0) \\
\kappa_n &= g^{(n)}(0)
\end{aligned}$$

2.4 Joint Distributions

The joint distribution for x and y denoted $f(x,y)$ is defined as

$$\mathrm{Prob}\left(a \le x \le b, c \le y \le d\right) = \begin{cases} \sum_a^b \sum_c^d f(x,y) \\ \int_a^b \int_c^d f(x,y)\,dxdy \end{cases}$$

Consider the following bivariate normal distribution as an example.

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\left[\varepsilon_x^2 + \varepsilon_y^2 - 2\rho\varepsilon_x\varepsilon_y\right] / \left(1-\rho^2\right)\right)$$

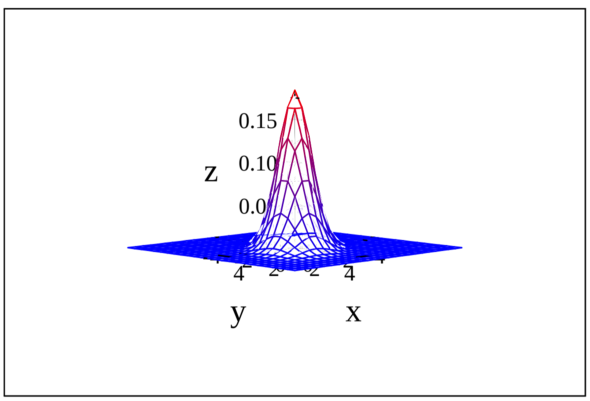
where

$$\varepsilon_x = \frac{x-\mu_x}{\sigma_x}, \quad \varepsilon_y = \frac{y-\mu_y}{\sigma_y}, \;\; \rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

Suppose that $\sigma_x = \sigma_y = 1, \mu_x = \mu_y = 0, \sigma_{xy} = 0.5$. Then we have

$$f(x,y) = \frac{1}{2\pi\sqrt{1-0.5^2}} \exp\left(-\frac{1}{2}\left[x^2 + y^2 - xy\right] / \left(1-0.5^2\right)\right)$$

$\frac{1}{2\pi\sqrt{1-0.5^2}} \exp\left(-\frac{1}{2}\left(x^2+y^2-xy\right)/\left(1-0.5^2\right)\right)$



We denote

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}\right)$$

Marginal distribution It is defined as

$$\mathrm{Prob}\left(x=x_0\right)=\sum_y \mathrm{Prob}\left(x=x_0|y=y_0\right)\mathrm{Prob}\left(y=y_0\right)=f_x\left(x\right)=\begin{cases}\sum_y f\left(x,y\right)\\ \int_y f\left(x,s\right)ds\end{cases}$$

Note that

$$f\left(x,y\right)=f_x\left(x\right)f_y\left(y\right)\text{ iff } x \text{ and } y \text{ are independent}$$

Also note that if x and y are independent, then

$$F\left(x,y\right)=F_x\left(x\right)F_y\left(y\right)$$

alternatively

$$\mathrm{Prob}\left(x\leq x_o,y\leq y_o\right)=\mathrm{Prob}\left(x\leq x_o\right)\mathrm{Prob}\left(y\leq y_o\right)$$

For a bivariate normal distribution case, the marginal distribution is given by

$$\begin{aligned} f_x\left(x\right) &= N\left(\mu_x,\sigma_x^2\right)\\ f_y\left(y\right) &= N\left(\mu_y,\sigma_y^2\right)\end{aligned}$$

Expectations in a joint distribution Mean:

$$E\left(x\right)=\begin{cases}\sum_x xf_x\left(x\right)=\sum_x x\sum_y f\left(x,y\right)\\ \int_x xf_x\left(x\right)dx=\int_x\int_y xf\left(x,y\right)dydx\end{cases}$$

Variance: See B-50.

Covariance and Correlation

$$\begin{aligned} Cov\left[x,y\right] &= E\left[\left(x-\mu_x\right)\left(y-\mu_y\right)\right]=\sigma_{xy}\\ Corr\left(x,y\right) &= \frac{\sigma_{xy}}{\sigma_x\sigma_y}\end{aligned}$$

$$V\left(x+y\right)=V\left(x\right)+V(y)+2Cov\left(x,y\right)$$

2.5 Conditioning in a bivariate distribution

$$f(y|x) = \frac{f(x,y)}{f_x(x)}$$

For a bivariate normal distribution case, the conditional distribution is given by

$$f(y|x) = N\left(\alpha + \beta x, \sigma_y^2 \left(1 - \rho^2\right)\right)$$

where $\alpha = \mu_y - \beta\mu_x$, $\beta = \sigma_{xy}/\sigma_x^2$.

If $\rho = 0$, then y and x are independent.

Regression: The Conditional Mean The conditional mean is the mean of the conditional distribution which is defined as

$$E(y|x) = \begin{cases} \sum_y y f(y|x) \\ \int_y y f(y|x)\, dy \end{cases}$$

The conditional mean function $E(y|x)$ is called the regression of y on x.

$$\begin{aligned} y &= E(y|x) + (y - E(y|x)) \\ &= E(y|x) + \varepsilon \end{aligned}$$

Example:

$$y = a + bx + \varepsilon$$

Then

$$E(y|x) = a + bx.$$

Conditional Variance

$$\begin{aligned} V(y|x) &= E\left[(y - E(y|x))^2 |x\right] \\ &= E\left(y^2|x\right) - E(y|x)^2 \end{aligned}$$

2.6 The Multivariate Normal Distribution

Let $\mathbf{x} = (x_1, ..., x_n)'$ and have a multivariate normal distribution. Then we have

$$f(\mathbf{x}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

1. If $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$\mathbf{A}\mathbf{x} + \mathbf{b} \sim N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$$

2. If $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim \chi^2_n$$

3. If $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$\boldsymbol{\Sigma}^{-1/2} (\mathbf{x} - \boldsymbol{\mu}) \sim N(0, \mathbf{I}_n)$$

2.7 Sample Questions

Q1: Write down the definitions of skewness and kurtosis. What is the value of skewness for the symmetric distribution.

Q2: Let $x_i \sim N(0, \sigma^2)$ for $i = 1, ..., n$. Further assume that x_i is independent each other.Then

1. $x_i^2 \sim$
2. $\sum_{i=1}^{n} x_i^2 \sim$
3. $\sum_{i=1}^{n} \frac{x_i^2}{\sigma^2} \sim$
4. $\frac{x_1^2}{x_2^2} \sim$
5. $\frac{x_1^2+x_2^2}{x_3^2} \sim$
6. $\frac{x_1}{x_2^2} \sim$
7. $\frac{x_1}{x_2^2+x_3^2} \sim$

Q3: Write down the standard normal density

Q4: Let $x \sim LN(\mu, \sigma^2)$.

1. $\ln(x) \sim$
2. Prove that $y = ax \sim LN(\ln a + \mu, \sigma^2)$.
3. Prove that $y = 1/x \sim LN(-\mu, \sigma^2)$.

Q5: Write down the density function of the Gamma distribution

1. Write down the values of p and λ when Gamma $= \chi^2$
2. Write down the values of p and λ when Gamma $=$ exponential distribution

Q6: Write down the density function of the logistic distribution.

Q7: Write down the density function of Cauchy distribution. Write down the value of v when Cauchy=t distribution

Q8: Write down the definition of Moment Generating and Characteristic function.

Q9: Suppose that $\mathbf{x} = (x_1, x_2, x_3)'$ and $\mathbf{x} \sim N(\mu, \Sigma)$

1. Write down the normal density in this case.

2. $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} \sim$

3. $\mathbf{z} = \mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{c}) \sim$ where $\mathbf{c} \neq \boldsymbol{\mu}$.

3 Estimation and Inference

3.1 Definitions

1. Random variable and constant: A random variable is believed to change over time across individual. Constant is believed not to change either dimension. It becomes an issue in the panel data.

   $$x_{it} = a_i + x_{it}^o$$

   Here we decompose x_{it} $(i = 1, ..., N;\ t = 1, ..., T)$ into its mean (time invariant) and time varying components. Now is a_i random or constant. According to the definition of random variables, a_i can be a constant since it does not change over time. However, if a_i has a pdf, then it becomes a random variable.

2. IID: independent, identically distributed: Consider the following sequence

   $$\mathbf{x} = (x_1, x_2, x_2) = (1, 2, 3)$$

   Now we are asking if each number is a random variable or constant. If they are random, then we have to ask the pdf of each number. Suppose that

   $$x_i \sim N\left(0, \sigma_i^2\right),$$

   Now we have to know that x_i is an independent event. If they are independent, then next we have to know σ_i^2 is identical or not. Typical assumption is IID.

3. Mean:

   $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

4. Standard error (deviation)

   $$s_x = \left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{1/2}$$

5. Covariance

   $$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

6. Correlation

$$\gamma_{xy} = \frac{s_{xy}}{s_x s_y}$$

Population values and estimates

$$y_i = bx_i + u_i$$

$$u_i \sim iidN\left(\mu, \sigma^2\right)$$

1. Estimate: it is a statistic computed from a sample $\left(\hat{b}\right)$. Sample mean is the statistic for the population mean (μ).

2. Standard deviation and error: σ is the standard deviation and s_x is the standard error of the population.

3. regression error and residual: u_i is the error, $\hat{u}_i$ is the residual

4. Estimator: a rule or method for using the data to estimate the parameter. "OLS estimator is consistent" should read "the estimation method using OLS is consistently estimated a parameter.

5. Asymptotic = Approximated. Asymptotic theory = approximation property. We are interested in how an approximation works as $n \to \infty$.

Estimation in the Finite Sample

1. Unbiased: An estimator of a parameter θ is unbiased if the mean of its sampling distribution is θ.

$$E\left(\hat{\theta} - \theta\right) = 0 \text{ for all } n.$$

2. Efficient: An unbiased estimator $\hat{\theta}_1$ is more efficient than another unbiased estimator $\hat{\theta}_2$ is the sampling variance of $\hat{\theta}_1$ is less than that of $\hat{\theta}_2$.

$$V\left(\hat{\theta}_1\right) < V\left(\hat{\theta}_2\right)$$

3. Mean Squared Error:

$$\begin{aligned} MSE\left(\hat{\theta}\right) &= E\left[\left(\hat{\theta}-\theta\right)^2\right] \\ &= E\left[\left(\hat{\theta}-E\hat{\theta}+E\hat{\theta}-\theta\right)^2\right] \\ &= V\left(\hat{\theta}\right)+\left[E\left(\hat{\theta}-\theta\right)\right]^2 \end{aligned}$$

4. Likelihood Function: rewrite

$$u_i = y_i - bx_i$$

and consider the joint density of u_i. If u_i are independent, then

$$\begin{aligned} f\left(u_1,...,u_n|b\right) &= f\left(u_1|b\right)f\left(u_2|b\right)...f\left(u_n|b\right) \\ &= \prod_{i=1}^{n} f\left(u_i|b\right) = L\left(b|x_i,...,x_n\right) \end{aligned}$$

The function $L\left(b|\mathbf{u}\right)$ is called the likelihood function for b given the data $\mathbf{u}$.

5. Cramer-Rao Lower Bound: Under regularity condition, the variance of an unbiased estimator of a parameter θ will always be at least as large as

$$\left[I\left(\theta\right)\right]^{-1} = \left(-E\left[\frac{\partial^2 \ln L\left(\theta\right)}{\partial\theta^2}\right]\right)^{-1} = \left(E\left[\frac{\partial \ln L\left(\theta\right)}{\partial\theta}\right]^2\right)^{-1}$$

where the quantity $I\left(\theta\right)$ is the information number for the sample.

4 Large Sample Distribution Theory

Definition and Theorem (Consistency and Convergence in Probability)

1. Convergence in probability: The random variable x_n converges in probability to a constant c if

$$\lim_{n\to\infty} \text{Prob}\left(|x_n-c|>\varepsilon\right)=0 \text{ for any positive } \varepsilon.$$

We denote

$$x_n \to^p c, \text{ or } \text{plim}_{n\to\infty} x_n = c$$

Carefully look at the subscript 'n'. This means x_n is dependent on the size of n. For an example, the sample mean, $n^{-1}\sum_{i=1}^{n} x_i$ is a function of n.

2. Almost sure convergence:

$$\text{Prob}\left(\lim_{n\to\infty} x_n = c\right) = 1$$

Note that almost sure convergence is stronger than convergence in probability. We denote

$$x_n \to^{a.s.} c$$

3. Convergence in the $r-$th mean

$$\lim_{n\to\infty} E\left(|x_n - c|^r\right) = 0$$

and denote it as

$$x_n \to^{L^r} c$$

When $r = 2$, we say convergence in quardratic mean.

4. Consistent Estimator: An estimator $\hat{\theta}_n$ of a parameter θ is a consistent estimator of θ iff

$$\text{plim}_{n\to\infty}\hat{\theta}_n = \theta$$

5. Khinchine's weak law of large number: If x_i is a random sample from a distribution with finite mean $E(x_i) = \mu$, then

$$\text{plim}_{n\to\infty}\bar{x}_n = \text{plim}_{n\to\infty} n^{-1}\sum_{i=1}^{n} x_i = \mu$$

6. Chebychev's weak law of large number: If x_i is a sample of observations such that $E(x_i) = \mu_i < \infty$, $V(x_i) = \sigma_i^2 < \infty$, $\bar{\sigma}_n^2/n = n^{-2}\sum_{i=1}^{n}\sigma_i^2 \to 0$ as $n \to \infty$, then

$$\text{plim}_{n\to\infty}\left(\bar{x}_n - \bar{\mu}_n\right) = 0$$

where $\bar{\mu}_n = n^{-1}\sum_{i=1}^{n}\mu_i$.

7. Kolmogorov's Strong LLN: If x_i is a sequence of independently distributed random variables such that $E(x_i) = \mu_i < \infty$ and $V(x_i) = \sigma_i^2 < \infty$ such that $\sum_{s=1}^{\infty} \sigma_s^2/s^2 < \infty$ as $n \to \infty$ then

$$\bar{x}_n - \bar{\mu}_n \to^{a.s.} 0$$

8. (Corollary of 7) If x_i is a sequence of iid variables such that $E(x_i) = \mu < \infty$, and $E|x_i| < \infty$, then

$$\bar{x}_n - \mu \to^{a.s.} 0$$

9. Markov's Strong LLN: If x_i is a sequence of independent random variables with $E(x_i) = \mu_i < \infty$ and if for some $\delta > 0$, $\sum_{i=1}^{\infty} E\left[|x_i - \mu_i|^{1+\delta}\right]/i^{1+\delta} < \infty$, then

$$\bar{x}_n - \bar{\mu}_n \to^{a.s.} 0$$

Properties of Probability Limits

1. If x_n and y_n are random variables with $\text{plim} x_n = b$ and $\text{plim} y_n = c$, then

$$\begin{aligned} \text{plim}(x_n + y_n) &= b + c \\ \text{plim} x_n y_n &= bc \\ \text{plim} \frac{x_n}{y_n} &= \frac{b}{c} \text{ if } c \neq 0 \end{aligned}$$

2. $\mathbf{W}_n$ is a matrix whose elements are random variables and if $\text{plim}\mathbf{W}_n = \Omega$, then

$$\text{plim}\mathbf{W}_n^{-1} = \Omega^{-1}$$

3. If $\mathbf{X}_n$ and $\mathbf{Y}_n$ are random matrices with $\text{plim}\mathbf{X}_n = \mathbf{B}$ and $\text{plim}\mathbf{Y}_n = \mathbf{C}$, then

$$\text{plim}\mathbf{X}_n\mathbf{Y}_n = \mathbf{BC}$$

Convergence in Distribution

1. x_n converges in distribution to a random variable x with cdf $F(x)$ if

$$\lim_{n\to\infty} |F(x_n) - F(x)| = 0 \text{ at all continuity points of } F(x)$$

   In this case, $F(x)$ is the limiting distribution of x_n, and this is written

$$x_n \to^d x$$

2. Cramer-Wold Device: If $\mathbf{x}_n \to^d \mathbf{x}$, then

$$\mathbf{c}'\mathbf{x}_n \to \mathbf{c}'\mathbf{x}$$

   where $\mathbf{c} \in R$

3. Lindeberg-Levy CLT (Central limit theorem): If $x_1, ..., x_n$ are a random sample from a probability distribution with finite mean μ and finite variance σ^2, then it sample mean, $\bar{x}_n = n^{-1}\sum_{i=1}^n x_i$ have the following limiting distribution

$$\sqrt{n}(\bar{x}_n - \mu) \to^d N\left(0, \sigma^2\right)$$

4. Lindegerg-Feller CLT: Suppose that $x_1, ..., x_n$ are a random sample from a probability distribution with finite mean μ_i and finite variance σ_i^2. Let

$$\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n \sigma_i^2, \quad \bar{\mu}_n = \frac{1}{n}\sum_{i=1}^n \mu_i$$

   where $\lim_{n\to\infty} \max(\sigma_i) / (n\bar{\sigma}_n) = 0$. Further assume that $\lim_{n\to\infty} \bar{\sigma}_n^2 = \bar{\sigma}^2 < \infty$, then it sample mean, $\bar{x}_n = n^{-1}\sum_{i=1}^n x_i$ have the following limiting distribution

$$\sqrt{n}(\bar{x}_n - \bar{\mu}_n) \to^d N\left(0, \bar{\sigma}^2\right)$$

   or

$$\frac{\sqrt{n}(\bar{x}_n - \bar{\mu}_n)}{\bar{\sigma}} \to^d N(0, 1)$$

5. Liapounov CLT: Suppose that $\{x_i\}$ is a sequnece of independent random variables with finite mean μ_i and finite positive variance σ_i^2 such that $E\left(|x_i - \mu_i|^{2+\delta}\right) < \infty$ for some $\delta > 0$. If $\bar{\sigma}_n$ is positive and finite for all n sufficiently large, then

$$\frac{\sqrt{n}\,(\bar{x}_n - \bar{\mu}_n)}{\bar{\sigma}_n} \to^d N(0,1)$$

6. Multivariate Lindeberg-Feller CLT:

$$\sqrt{n}\,(\bar{\mathbf{x}}_n - \bar{\boldsymbol{\mu}}_n) \to^d N(0,\mathbf{Q})$$

where $V(\mathbf{x}_i) = \mathbf{Q}_i$ and we assume that $\lim \bar{\mathbf{Q}}_n = \mathbf{Q}$

7. Asymptotic Covariance Matrix: Suppose that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \to^d N(0,\mathbf{V})$$

then its asymptotic covariance matrix is defined as

$$\text{Asy. Var}\left(\hat{\boldsymbol{\theta}}_n\right) = \frac{1}{n}\mathbf{V}$$

Order of A Sequence

1. A sequence c_n is of order n^δ, denoted $O\left(n^\delta\right)$, iff

$$\text{plim}_{n\to\infty}\frac{c_n}{n^\delta} = c < \infty$$

   (a) $c_n = 1 = O(1)$

   (b) $c_n = n^2 = O(n^2)$: $\text{plim} n^2/n^2 = 1$.

   (c) $c_n = 1/(n+10) = O(n^{-1})$: $\text{plim}(n+10)^{-1}/n^{-1} = 1$

2. A sequence c_n is of order less than n^δ iff

$$\text{plim}_{n\to\infty}\frac{c_n}{n^\delta} = 0$$

   (a) $c_n = 0 = o(1)$: $\text{plim} 0/1 = 0$.

   (b) $c_n = O\left(n^{-1/2}\right)$, then $c_n = o(1)$

Order in Probability

1. A sequence random variable x_n is $O_p(g(n))$ if there exists some N_ε such that $\varepsilon > 0$ and all $n > N_\varepsilon$,

$$\text{Prob}\left(\left|\frac{f_n}{g(n)}\right| < c\right) > 1 - \varepsilon$$

where c is a finite constant

   (a) If $x_n \sim N(0, \sigma^2)$, then $x_n = O_p(1)$. Since given ε, there is always some c such that

$$\text{Prob}(|x_n| < c) > 1 - \varepsilon$$

   (b) $O_p(n^a) O_p(n^b) = O_p(n^{a+b})$

   (c) If $\sqrt{n}(\bar{x}_n - \bar{\mu}_n) \rightarrow^d N(0, \bar{\sigma}^2)$, then $(\bar{x}_n - \bar{\mu}_n) = O_p(n^{-1/2})$ but $\bar{x}_n = O_p(1)$

2. The notation $x_n = o_p(g_n)$ means

$$\frac{x_n}{g_n} \rightarrow^p 0$$

   (a) If $\sqrt{n}\bar{x}_n \rightarrow^d N(0, \bar{\sigma}^2)$, then $\bar{x}_n = O_p(n^{-1/2})$ and $\bar{x}_n = o_p(1)$

   (b) $o_p(n^a) o_p(n^b) = o_p^{a+b}$

Sample Qestions

Part I: Consider the following model

$$\begin{aligned} \text{M1} &: y_i = bx_i + u_i, \quad i = 1, ..., n \\ \text{M2} &: y_i = a + bx_i + u_i \end{aligned}$$

Suppose that

$$Ex_n u_1 = c < \infty \text{ but } Ex_i u_i = 0 \text{ for all } i$$

Q1: Show the OLS estimator $\hat{b}$ in M1 is unbiased and consistent

Q2: Show the OLS estimator $\hat{b}$ is biased but consistent

Q3: Suppose that $u_i \sim iidN(0,1)$. Derive the limiting distribution of $\hat{b}$ in M1

Q4: Suppose that $u_i \sim iidN(0,\sigma_i^2)$. Derive the limiting distribution of $\hat{b}$ in M2

Part II: Consider the following model

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u}$$

Q5: Obtain the limiting distribution of $\boldsymbol{\theta} = \mathbf{x}'\mathbf{u}$

Q6: Obtain the limiting distribution of $\hat{\mathbf{b}}$

Q7: Suppose that $u_i \sim N(0,\sigma_i^2)$. Find the asymptotic variance of $\hat{\mathbf{b}}$

5 Chapters 1 through 4: The Classical Assumptions

1. Linear

$$\mathbf{y} = \mathbf{Xb} + \mathbf{u} < \infty$$

2. $\mathbf{X}$ is a nonstochastic and finite $n \times K$ matrix

3. $\mathbf{X}'\mathbf{X}$ is nonsingular for all $n \geq K$

4. $E(\mathbf{u}) = 0$

5. $\mathbf{u} \sim N(0, \sigma^2 \mathbf{I})$.

When X is stochastic 4. Exogeneity of the independent variables: $E(\mathbf{u}|\mathbf{X}) = 0$

5-1. Homoscedasticity and no-autocorrelation.

5-2. Normal distribution.

Properties: A. Existence: Given 1,2,3, $\hat{\boldsymbol{\beta}}$ exists for all $n \geq k$ and it unique

B. Unbiasedness: Given 1-4,

$$E\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$$

C. Normality: Given 1-5,

$$\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$$

D. Gauss-Markov Theorem: OLS estimator is the minimum variance linear unbiased estimator (whether $\mathbf{X}$ is stochastic or nonstochastic).

Linear Model Consider two models.

$$\begin{aligned} y_i &= a + bx_i + u_i \\ y_i &= a + bx_i + cx_i^2 + u_i \end{aligned}$$

Consider log, semilog, level:

$$\begin{aligned} y_i &= a + bx_i + u_i \\ \ln y_i &= a + b \ln x_i + e_i \\ \ln y_i &= a + bx_i + v_i \end{aligned}$$

5.1 Brainstorm I: Sample Mean

Notation:

$$y_i = \text{ earning at time } t$$

Classification: Male, Female, Skilled Worker, Non-skilled worker.

Question 1: How to test the difference between male and female earning.

$$\begin{aligned} y_1 &= \text{ sample mean of male earning} \\ y_2 &= \text{ sample mean of female earning} \end{aligned}$$

Question 2: How to explain the earning difference between male and female. By using skill data.

$$\begin{aligned} y_3 &= \text{ sample mean of skilled worker} \\ y_4 &= \text{ sample mean of nonskilled worker} \end{aligned}$$

If so how?

Question 3: Deriving the limiting distributions for Q1 and Q2 as $n \to \infty$

Question 4: Form a null hypothesis to test if male and female earning difference.

5.2 Brainstorm II: Trend Regression

Notation: The true model is given by

$$y_t = t + \varepsilon_t, \quad \varepsilon_t \sim iidN\left(0, \sigma^2\right)$$

Now consider two regressions

$$\begin{aligned} y_t &= b_1 t + u_t \\ y_t &= b_2 \sqrt{t} + e_t \end{aligned}$$

Question 1: Deriving the limiting distribution of $\hat{b}_1$

Question 2: Write down e_t as a function of t, ε_t and $\sqrt{t}$.

Question 3: Deriving the limiting distribution of $\hat{b}_2$

5.3 Brainstorm I Continue: Dummy Regression

Notation: y_i = earning. S_i = Decision for Ph.D. program. G_i = Decision for taking Econometric class

Consider the following decision tree.

If $S_i = 0$, then the values for G_i does not matter.

Question 1: Construct a dummy regression for the first example of Brainstorm I

Question 2: Construct a dummy regression for the current example

Question 3: Derive the limiting distribution for Q2.

5.4 Assignment II: 2 Extra Credits

Basic 0: Sample Mean Suppose that x_i is i.i.d. $N(\mu, \sigma^2)$.

1. Let $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$. Show that $\hat{\sigma}^2$ is biased but consistent. Obtain the unbiased estimator.

2. Let $z_i = x_i - x_1$ for $i = 2, ..., n$. Obtain the unbiased variance for z_i.

3. Let $w_i = x_i - 2x_1 + x_2$ for $i = 3, ..., n$. Obtain the unbiased variance for w_i.

4. Let $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2$ be the unbiased variance for Q1,2,3. Find the smallest variance.

Basic I: Single Regressor Consider the following model

$$y_i = bx_i + u_i$$

We assume that all classical assumptions hold.

1. Show the OLS estimator $\hat{b}$ is unbiased.

2. Show the OLS estimator is minimizing the following quadratic loss
$$\sum_{i=1}^{n} u_i^2$$

3. Show that
$$\frac{1}{n}\sum_{i=1}^{n} x_i \hat{u}_i = 0$$

Basic II: Multiple Regressors Consider the following model

$$y_i = \mathbf{X}_i \boldsymbol{\beta} + u_i = x_{1i}\beta_1 + x_{2i}\beta_2 + u_i$$

or equivalently

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{u}$$

We assume the classical assumptions hold.

1. Show the OLS estimator $\hat{\beta}$ is unbiased.

2. Show the OLS estimator is minimizing the following quadratic loss

$$\mathbf{u}'\mathbf{u}$$

3. Show that

$$\mathbf{X}'\hat{\mathbf{u}} = 0$$

4. Show that

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

5. Define $\mathbf{M}_1$ and $\mathbf{M}_2$, and show the followings

$$\begin{aligned}\hat{\beta}_1 &= (\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y} \\ \hat{\beta}_2 &= (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y}\end{aligned}$$

6. Suppose that $\beta_2 = 0$. Consider the following two regressions

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{e} \quad (3)$$

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{u} \quad (4)$$

   (a) Show that R^2 in (3) is smaller than that in (4).

   (b) Write down the relationship between R^2 and $\bar{R}^2$ in general.

   (c) Suppose that $\beta_2 \neq 0$ and t ratio for $\hat{\beta}_2$ is greater than 1. Show that $\bar{R}^2$ in (4) is greater than R^2 in (3).

7. Consider (3) as the true model. $e_i \sim iidN(0, \sigma^2)$. Let $s^2 = \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-1}$.

   (a) Show that $\hat{\beta}_1$ and $\hat{s}^2$ be independent.

   (b) Show that

$$\frac{\hat{\beta}_1 - \beta}{\sqrt{s^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}}} \sim t_1$$

8. Consider (4) as the true model. Let $x_{2i} = x_{1i} + v_i$. x_{1i} is independent on v_i. Also y_i is independent on v_i.

   (a) Find plim$\hat{\beta}_2$.

   (b) Let $v_i \sim iid\,(0, n^{-\alpha})$ for $\alpha > 0$. Find plim$\hat{\beta}_1$ and plim$\hat{\beta}_2$

9. Consider (4) as the true model. $\beta_2 \neq 0$, and x_{1i} is independent on x_{2i}. Suppose that you are interested in testing

$$H_0 : \gamma = \frac{\beta_1}{\beta_2} = 0$$

   Derive the limiting distribution $\hat{\gamma} = \hat{\beta}_1/\hat{\beta}_2$

5.5 Answer and Additional Notes

True Model (when x_i is nonstochastic)

$$y_i^* = \mu_y + y_i, \quad x_i^* = \mu_x + x_i$$

$$y_i = \beta x_i + u_i$$

Regression

$$\begin{aligned} y_i^* &= \mu_y + \beta x_i + u_i = \mu_y - \beta\mu_x + \beta\left(\mu_x + x_i\right) + u_i \\ &= \alpha + \beta x_i^* + u_i \end{aligned} \tag{5}$$

1. How to obtain $\hat{\alpha}$ and $\hat{\beta}$ (OLS estimators)

Take the sample average

$$\frac{1}{n}\sum y_i^* = \alpha + \beta\frac{1}{n}\sum x_i^* + \frac{1}{n}\sum u_i^* \tag{6}$$

(2) - (1) yields

$$\tilde{y}_i^* = \beta\tilde{x}_i^* + \tilde{u}_i$$

or equivalently

$$\tilde{y}_i = \beta\tilde{x}_i + \tilde{u}_i$$

since

$$\tilde{y}_i^* = y_i - \frac{1}{n}\sum y_i = \tilde{y}_i$$

$$\hat{\beta} = \beta + \frac{\sum \tilde{x}_i\tilde{u}_i}{\sum \tilde{x}_i^2}$$

$$\hat{\beta} - \beta = \frac{\sum \tilde{x}_i\tilde{u}_i}{\sum \tilde{x}_i^2} = \frac{\frac{1}{n}\sum \tilde{x}_i\tilde{u}_i}{\frac{1}{n}\sum \tilde{x}_i^2}$$

Since we assume x_i is nonstochastic, we have

$$E\left(\hat{\beta} - \beta\right) = E\frac{\frac{1}{n}\sum \tilde{x}_i\tilde{u}_i}{\frac{1}{n}\sum \tilde{x}_i^2} = \frac{\frac{1}{n}\sum \tilde{x}_i E\tilde{u}_i}{\frac{1}{n}\sum \tilde{x}_i^2} = 0$$

$$E\left(\hat{\beta} - \beta\right)^2 = E\left[\frac{\frac{1}{n}\sum \tilde{x}_i\tilde{u}_i}{\frac{1}{n}\sum \tilde{x}_i^2}\right]^2 = \frac{E\frac{1}{n^2}\left(\sum \tilde{x}_i\tilde{u}_i\right)^2}{\left[\frac{1}{n}\sum \tilde{x}_i^2\right]^2} \tag{7}$$

Note that

$$\left(\sum \tilde{x}_i\tilde{u}_i\right)^2 = \left(\tilde{x}_1\tilde{u}_1 + ... + \tilde{x}_n\tilde{u}_n\right)^2 = \left(\tilde{x}_1^2\tilde{u}_1^2 + ... + \tilde{x}_n^2\tilde{u}_n^2\right) + 2\left(\tilde{x}_1\tilde{u}_1\tilde{x}_2\tilde{u}_2 + ... + \tilde{x}_n\tilde{u}_n\tilde{x}_{n-1}\tilde{u}_{n-1}\right)$$

Hence

$$\begin{aligned} E\left(\sum \tilde{x}_i\tilde{u}_i\right)^2 &= E\left(\tilde{x}_1^2\tilde{u}_1^2 + ... + \tilde{x}_n^2\tilde{u}_n^2\right) + 2E\left(\tilde{x}_1\tilde{u}_1\tilde{x}_2\tilde{u}_2 + ... + \tilde{x}_n\tilde{u}_n\tilde{x}_{n-1}\tilde{u}_{n-1}\right) \\ &= \left(\tilde{x}_1^2E\tilde{u}_1^2 + ... + \tilde{x}_n^2E\tilde{u}_n^2\right) + 2\left(\tilde{x}_1\tilde{x}_2E\tilde{u}_1\tilde{u}_2 + ... + \tilde{x}_n\tilde{x}_{n-1}E\tilde{u}_n\tilde{u}_{n-1}\right) \end{aligned}$$

We will assume

$$\begin{aligned} E\left(u_iu_j\right) &= 0 \text{ for } i \neq j : \text{ independent} \\ Eu_i^2 &= \sigma_u^2 : \text{identical} \end{aligned}$$

Then we have

$$\begin{aligned} E\tilde{u}_i^2 &= E\left(u_i - \frac{1}{n}\sum u_i\right)^2 = E\left(u_i^2 - \frac{2}{n}u_i\sum u_i + \frac{1}{n^2}\left(\sum u_i\right)^2\right) \\ &= \sigma_u^2 - \frac{2}{n}E\left(u_iu_1 + ... + u_i^2 + u_iu_{i+1} + ... + u_iu_n\right) \\ &\quad + E\frac{1}{n^2}\left(\sum u_i^2 + 2\left(u_1u_2 + ... + u_nu_{n-1}\right)\right) \\ &= \sigma_u^2 - \frac{2}{n}\left(Eu_iu_1 + ... + Eu_i^2 + Eu_iu_{i+1} + ... + Eu_iu_n\right) \\ &\quad + \frac{1}{n^2}\left(\sum Eu_i^2 + 2E\left(u_1u_2 + ... + u_nu_{n-1}\right)\right) \\ &= \sigma_u^2 - \frac{2}{n}\sigma_u^2 + \frac{1}{n}\sigma_u^2 = \sigma_u^2\left(1 - \frac{1}{n}\right) = \frac{n-1}{n}\sigma_u^2 \end{aligned} \tag{8}$$

since

$$Eu_iu_1 = 0 \text{ if } i \neq 1, \text{ and } Eu_iu_{i+1} = 0 \text{ for all } i.$$

Also note that

$$\begin{aligned} E\tilde{u}_i\tilde{u}_{i+1} &= E\left(u_i - \frac{1}{n}\sum u_i\right)\left(u_{i+1} - \frac{1}{n}\sum u_i\right) \\ &= E\left(u_iu_{i+1} - \frac{1}{n}u_{i+1}\sum u_i - \frac{1}{n}u_i\sum u_{i+1} + \frac{1}{n^2}\left(\sum u_i\right)^2\right) \\ &= 0 - \frac{1}{n}\sigma_u^2 - \frac{1}{n}\sigma_u^2 + \frac{1}{n^2}\left(n \cdot \sigma_u^2\right) = -\frac{1}{n}\sigma_u^2 = O\left(n^{-1}\right) \end{aligned}$$

Now, we have

$$
\begin{aligned}
E\left(\sum \tilde{x}_i\tilde{u}_i\right)^2 &= \left(\tilde{x}_1^2E\tilde{u}_1^2 + ... + \tilde{x}_n^2E\tilde{u}_n^2\right) + 2\left(\tilde{x}_1\tilde{x}_2E\tilde{u}_1\tilde{u}_2 + ... + \tilde{x}_n\tilde{x}_{n-1}E\tilde{u}_n\tilde{u}_{n-1}\right) \\
&= \frac{n-1}{n}\sigma_u^2\sum_{i=1}^n \tilde{x}_i^2 - 2\frac{1}{n}\sigma_u^2\left(\tilde{x}_1\tilde{x}_2 + ... + \tilde{x}_n\tilde{x}_{n-1}\right) \\
&= \sigma_u^2\sum_{i=1}^n \tilde{x}_i^2 - \frac{1}{n}\sigma_u^2\left[\sum_{i=1}^n \tilde{x}_i^2 - 2\left(\tilde{x}_1\tilde{x}_2 + ... + \tilde{x}_n\tilde{x}_{n-1}\right)\right] \\
&= \sigma_u^2\sum_{i=1}^n \tilde{x}_i^2 - \frac{1}{n}\sigma_u^2\left(\sum_{i=1}^n \tilde{x}_i\right)^2 \\
&= \sigma_u^2\sum_{i=1}^n\left(\tilde{x}_i - \frac{1}{n}\sum_{i=1}^n \tilde{x}_i\right)^2 = \sigma_u^2\sum_{i=1}^n\left(x_i - \frac{1}{n}\sum_{i=1}^n x_i\right)^2 \\
&= \sigma_u^2\sum_{i=1}^n \tilde{x}_i^2
\end{aligned}
$$

since

$$
\begin{aligned}
\tilde{x}_i - \frac{1}{n}\sum_{i=1}^n \tilde{x}_i &= x_i - \frac{1}{n}\sum_{i=1}^n x_i - \frac{1}{n}\sum_{i=1}^n\left(x_i - \frac{1}{n}\sum_{i=1}^n x_i\right) \\
&= x_i - \frac{1}{n}\sum_{i=1}^n x_i - \frac{1}{n}\sum_{i=1}^n x_i + \frac{1}{n}\sum_{i=1}^n x_i \\
&= x_i - \frac{1}{n}\sum_{i=1}^n x_i = \tilde{x}_i.
\end{aligned}
$$

Now from (3), we have

$$
\begin{aligned}
E\left(\hat{\beta} - \beta\right)^2 &= E\left[\frac{\frac{1}{n}\sum \tilde{x}_i\tilde{u}_i}{\frac{1}{n}\sum \tilde{x}_i^2}\right]^2 = \frac{E\frac{1}{n^2}\left(\sum \tilde{x}_i\tilde{u}_i\right)^2}{\left[\frac{1}{n}\sum \tilde{x}_i^2\right]^2} = \frac{\frac{1}{n^2}\sigma_u^2\sum_{i=1}^n \tilde{x}_i^2}{\left[\frac{1}{n}\sum \tilde{x}_i^2\right]^2} \\
&= \frac{\frac{1}{n}\sigma_u^2\left[\frac{1}{n}\sum_{i=1}^n \tilde{x}_i^2\right]}{\left[\frac{1}{n}\sum \tilde{x}_i^2\right]^2} = \frac{\sigma_u^2}{\sum \tilde{x}_i^2}
\end{aligned}
$$

Note that

$$
\lim_{n\to\infty}\frac{\sigma_u^2}{\sum \tilde{x}_i^2} = 0 \text{ if } \frac{1}{n}\sum \tilde{x}_i^2 = c
$$

In order to have a finite variance, we need to have

$$
\sqrt{n}\left(\hat{\beta} - \beta\right) = \frac{\frac{1}{\sqrt{n}}\sum \tilde{x}_i\tilde{u}_i}{\frac{1}{n}\sum \tilde{x}_i^2}
$$

Now it is easy to show that (by using LL CLT)

$$\sqrt{n}\left(\hat{\beta}-\beta\right) \rightarrow^{d} N\left(0, \frac{\sigma_u^2}{\frac{1}{n}\sum \tilde{x}_i^2}\right)$$

or

$$\frac{\hat{\beta}-\beta}{\sqrt{\sigma_u^2/\sum \tilde{x}_i^2}} \rightarrow^{d} N(0,1).$$

To Students:

Q1. Now you do more simple model

$$y_i = \beta x_i + u_i \tag{9}$$

and get the limiting distribution of $\hat{\beta}$. Here we assume $\mu_y = \mu_x = 0$.

Q2. (Example of nonstochastic x_i) Consider

$$y_i = a + u_i. \tag{10}$$

This is a regression of y_i on 1. That is, we let $\alpha = \beta$, and $x_i = 1$ for all i in (9), then we have (10). Find the limiting distribution of $\hat{a}$.

When x_i is stochastic We don't work with expectation term here. Instead of this, we take probability limit (since we are obtaining the limiting distribution, so we are caring about consistency only. The previous case, both two unbiaseness and consistency becomes identical problem since x_i was nonstochastic.)

$$\text{plim}_{n\to\infty}\left(\hat{\beta}_n-\beta\right) = \frac{\text{plim}_{n\to\infty}\frac{1}{n}\sum \tilde{x}_i\tilde{u}_i}{\text{plim}_{n\to\infty}\frac{1}{n}\sum \tilde{x}_i^2} = \frac{\text{plim}_{n\to\infty}\frac{1}{n}\sum \tilde{x}_i\tilde{u}_i}{\text{plim}_{n\to\infty}\frac{1}{n}\sum \tilde{x}_i^2} = 0$$

Note that

$$E\tilde{u}_i^2 = \frac{n-1}{n}\sigma_u^2, \quad E\tilde{u}_i\tilde{u}_{i+1} = -\frac{1}{n}\sigma_u^2$$

Similarly, we have

$$E\tilde{x}_i^2 = \frac{n-1}{n}\sigma_x^2, \quad E\tilde{x}_i\tilde{x}_{i+1} = -\frac{1}{n}\sigma_x^2$$

Hence

$$
\begin{aligned}
E\left(\sum \tilde{x}_i\tilde{u}_i\right)^2 &= E\left(\tilde{x}_1^2\tilde{u}_1^2+...+\tilde{x}_n^2\tilde{u}_n^2\right)+2E\left(\tilde{x}_1\tilde{x}_2\tilde{u}_1\tilde{u}_2+...+\tilde{x}_n\tilde{x}_{n-1}\tilde{u}_n\tilde{u}_{n-1}\right)\\
&= \left(E\tilde{x}_1^2E\tilde{u}_1^2+...+E\tilde{x}_n^2E\tilde{u}_n^2\right)+2\left(E\tilde{x}_1\tilde{x}_2E\tilde{u}_1\tilde{u}_2+...+E\tilde{x}_n\tilde{x}_{n-1}E\tilde{u}_n\tilde{u}_{n-1}\right)\\
&= n\left(\frac{n-1}{n}\sigma_x^2\right)\left(\frac{n-1}{n}\sigma_u^2\right)-n\left(n-1\right)\left(\frac{1}{n}\sigma_x^2\right)\left(\frac{1}{n}\sigma_u^2\right)\\
&= n\sigma_x^2\sigma_u^2\frac{\left(n-1\right)^2}{n^2}-\frac{n\left(n-1\right)}{n^2}\sigma_x^2\sigma_u^2=n\sigma_x^2\sigma_u^2\left[\frac{\left(n-1\right)^2-\left(n-1\right)}{n^2}\right]\\
&= n\sigma_x^2\sigma_u^2\left[1-\frac{3n+2}{n^2}\right]=n\sigma_x^2\sigma_u^2\left[1-\frac{3}{n}\right]=n\sigma_x^2\sigma_u^2+O\left(n^{-1}\right)
\end{aligned}
$$

Also note that

$$
\begin{aligned}
\text{plim}_{n\to\infty}\frac{1}{n}\sum\left(x_i-\frac{1}{n}\sum x_i\right)^2 &= \text{plim}_{n\to\infty}\frac{1}{n}\sum x_i^2-\text{plim}_{n\to\infty}\left(\frac{1}{n}\sum x_i\right)^2\\
&= \sigma_x^2-\frac{1}{n}\sigma_x^2=\sigma_x^2+O\left(n^{-1}\right)
\end{aligned}
$$

Then we have

$$
\begin{aligned}
\frac{1}{n}\sum \tilde{x}_i\tilde{u}_i &\to {}^d N\left(0,\frac{\sigma_x^2\sigma_u^2}{n}+O\left(n^{-1}\right)\right)\\
\frac{1}{\sqrt{n}}\sum \tilde{x}_i\tilde{u}_i &\to {}^d N\left(0,\sigma_x^2\sigma_u^2\right)
\end{aligned}
$$

From Cramer-Wold Device, we have

$$
\frac{\frac{1}{\sqrt{n}}\sum \tilde{x}_i\tilde{u}_i}{\frac{1}{n}\sum \tilde{x}_i^2}\to^d N\left(0,\frac{\sigma_u^2}{\sigma_x^2}\right)
$$

Usually in textbooks, we don't follow the above derivation. Simply others use the conditional expectation. Let $\mathbf{x}=(x_1,...,x_n)'$. Then we consider

$$
E\left(\hat{\beta}-\beta|\mathbf{x}\right)
$$

and

$$
E\left(\hat{\beta}-\beta|\mathbf{x}\right)^2
$$

which is equivalent to treat x_i like a nonstochastic variable. Asymptotically we note that

$$\sqrt{n}\left(\hat{\beta}-\beta|\mathbf{x}\right) \rightarrow^{d} N\left(0, \sigma_u^2 \mathbf{Q}_x^{-1}\right)$$

where

$$\mathbf{Q}_x = \lim_{n\rightarrow\infty}\left(\frac{\mathbf{x}'\mathbf{x}}{n}\right)^{-1}$$

Dummy Variable Regression: Let's go back to our orignial example

$$y_i = a + \beta S_i + u_i$$

where

$$S_i = 0 \text{ for } i = female, \quad 1 \text{ for } i = male.$$

Suppose that n_1 = total number of female = total number of male. Then

$$n = n_1 + n_1 = 2n_1, \text{ or } n_1 = \frac{n}{2}$$

$$\hat{\beta} = \beta + \frac{\sum \tilde{S}_i \tilde{u}_i}{\sum \tilde{S}_i^2}$$

Treat as if S_i is nonrandom. Then we have

$$E \sum \tilde{S}_i \tilde{u}_i = \sum \tilde{S}_i E \tilde{u}_i = 0$$

Next

$$\sum \tilde{S}_i^2 = \sum \left(S_i - \frac{1}{n} \sum_{i=1}^{n} S_i \right)^2 = \sum S_i^2 - \frac{1}{n} \left(\sum S_i \right)^2$$

Note

$$S_i^2 = S_i = \begin{cases} 0 \text{ if female} \\ 1 \text{ if male} \end{cases}$$

so that

$$\sum S_i^2 = \sum S_i = n_1 : \text{ total number of female or male}$$

Hence we have

$$\sum \tilde{S}_i^2 = \sum S_i^2 - \frac{1}{n} \left(\sum S_i \right)^2 = n_1 - \frac{1}{n} n_1^2 = \frac{n}{2} - \frac{1}{n} \frac{n^2}{4} = \frac{2n - n}{4} = \frac{1}{4} n \tag{11}$$

Next, find the limiting distribution of

$$\frac{1}{\sqrt{n}} \sum \tilde{S}_i \tilde{u}_i.$$

We know

$$\text{plim} \frac{1}{n} \sum \tilde{S}_i \tilde{u}_i = 0$$

also we know

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum \tilde{S}_i\tilde{u}_i &= \frac{1}{\sqrt{n}}\sum\left(S_i - \frac{1}{n}\sum_{i=1}^{n} S_i\right)\tilde{u}_i = \frac{1}{\sqrt{n}}\sum S_i\tilde{u}_i - \frac{1}{\sqrt{n}}\sum\left(\frac{1}{n}\sum_{i=1}^{n} S_i\right)\tilde{u}_i \\
&= \frac{1}{\sqrt{n}}\sum S_i\tilde{u}_i - \frac{1}{\sqrt{n}}\left(\frac{1}{n}\sum_{i=1}^{n} S_i\right)\sum \tilde{u}_i \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n_1}\tilde{u}_i - \frac{n_1}{n}\frac{1}{\sqrt{n}}\sum\tilde{u}_i
\end{aligned}$$

since

$$\begin{aligned}
S_i\tilde{u}_i &= 0 \text{ if } i \text{ is female.} \\
&= \tilde{u}_i \text{ if } i \text{ is male}
\end{aligned}$$

Hence

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum \tilde{S}_i\tilde{u}_i &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n_1}\tilde{u}_i - \frac{n_1}{n}\frac{1}{\sqrt{n}}\sum\tilde{u}_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n_1}\tilde{u}_i - \frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{u}_i \\
&= \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n_1}\tilde{u}_i - \frac{1}{2}\sum_{i=1}^{n}\tilde{u}_i\right) = \frac{1}{\sqrt{n}}\left(\frac{1}{2}\sum_{i=1}^{n_1}\tilde{u}_i - \frac{1}{2}\sum_{i=n_1+1}^{n}\tilde{u}_i\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{u}_i^*
\end{aligned}$$

where

$$\tilde{u}_i^* = \begin{cases} \frac{1}{2}\tilde{u}_i \text{ if } i \text{ is male} \\ -\frac{1}{2}\tilde{u}_i \text{ if } i \text{ is female} \end{cases}$$

Note that the stochastic properties of $\tilde{u}_i^*$ is the same as $\frac{1}{2}\tilde{u}_i$ as long as u_i is independent and identically distributed.

Next, we know already the value of $E\left(\tilde{u}_i\right)^2$ from (8).

$$\begin{aligned}
E\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{u}_i^*\right)^2 &= \frac{1}{n}E\left(\sum_{i=1}^{n}\left(u_i^*\right)^2 - \frac{1}{n}\left(\sum_{i=1}^{n}u_i\right)^2\right) \\
&= \frac{1}{n}\left(\frac{n}{4}\sigma_u^2 - \frac{1}{n}\frac{n}{4}\sigma_u^2\right) \\
&= \frac{1}{4}\sigma_u^2 - \frac{1}{4n}\sigma_u^2 = \frac{1}{4}\sigma_u^2 + O\left(n^{-1}\right)
\end{aligned}$$

Hence

$$\frac{1}{\sqrt{n}} \sum \tilde{S}_i \tilde{u}_i \to^d N\left(0, \frac{1}{4}\sigma_u^2\right)$$

and

$$n\left(\hat{\beta} - \beta\right) = \frac{\frac{1}{\sqrt{n}} \sum \tilde{S}_i \tilde{u}_i}{\frac{1}{n} \sum \tilde{S}_i^2} = \frac{\frac{1}{\sqrt{n}} \sum \tilde{S}_i \tilde{u}_i}{\frac{1}{n} \frac{n}{4}} = \frac{4}{\sqrt{n}} \sum \tilde{S}_i \tilde{u}_i$$

since $\sum \tilde{S}_i^2 = n/4$ from (11). Therefore finally we have

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \to^d N\left(0, 4\sigma_u^2\right),$$

or

$$\sqrt{n} \frac{\hat{\beta} - \beta}{\sqrt{4\sigma_u^2}} \to^d N(0, 1)$$

Note that the asymptotic variance of $\hat{\beta}$ is

$$\text{Asy Var}\left(\hat{\beta}\right) = \frac{4\sigma_u^2}{n}$$

Now consider the quantity of

$$\sigma_u^2 \left(\sum \tilde{S}_i^2\right)^{-1} = \sigma_u^2 \frac{4}{n} = \text{Asy Var}\left(\hat{\beta}\right)$$

Two Dependent Dummies Consider two models

$$\begin{aligned} y_i &= \alpha + \beta S_i + \gamma U_i + \varepsilon_i \\ y_i &= \alpha + \beta S_i + u_i \end{aligned}$$

Now

$$u_i = \gamma U_i + \varepsilon_i$$

where

$$U_i = \begin{cases} 0 \text{ if } i \text{ is non-skilled} \\ 1 \text{ if } i \text{ is skilled} \end{cases}$$

Further note that

$$\frac{1}{n} \sum S_i U_i \neq 0$$

Consider the following 'pay-off' matrix where n_{ij} indiciates the total number of observations.

	Unskilled	Skilled	Total
Female	n_{11}	n_{12}	$n_{11}+n_{12}$
Male	n_{21}	n_{22}	$n_{21}+n_{22}$
Total	$n_{11}+n_{21}$	$n_{12}+n_{22}$	n

Assume the total number of female = that of male.

$$n_{11}+n_{12}=n_{21}+n_{22}$$

Further consider the following assumptions.

$$n_{11}=2n_{21}$$

$$2n_{12}=n_{22}$$

Then we have

	Unskilled	Skilled	Total
Female	$2n_{11}$	n_{22}	$2n_{11}+n_{22}$
Male	n_{11}	$2n_{22}$	$n_{11}+2n_{22}$
Total	$3n_{11}$	$3n_{22}$	n

and the probability matrix becomes

	Unskilled	Skilled	Total
Female	$2n_{11}/n$	n_{22}/n	$\frac{2n_{11}+n_{22}}{n}$
Male	n_{11}/n	$2n_{22}/n$	$\frac{n_{11}+2n_{22}}{n}$
Total	$3n_{11}/n$	$3n_{22}/n$	1

Note that

$$\frac{2n_{11}+n_{22}}{n}=\frac{n_{11}+2n_{22}}{n}=\frac{1}{2}\iff n_{22}=n_{11}=n_0=\frac{1}{6}$$

so that finally we have

	Unskilled	Skilled	Total
Female	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$
Male	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$
Total	$\frac{1}{2}$	$\frac{1}{2}$	1

Consider the following expectation

Expected earning	
Female & unskilled; $S_i = U_i = 0$	$E(y_i) = \alpha$
Female & skilled; $S_i = 0$, $U_i = 1$	$E(y_i) = \alpha + \gamma$
Male & unskilled; $S_i = 1$, $U_i = 0$	$E(y_i) = \alpha + \beta$
Male & skilled; $S_i = 1$, $U_i = 1$	$E(y_i) = \alpha + \beta + \gamma$

	Unskilled	Skilled	Total
Female	α	$\alpha + \gamma$	$\frac{2}{3}\alpha + \frac{1}{3}(\alpha + \gamma)$
Male	$\alpha + \beta$	$\alpha + \beta + \gamma$	$\frac{1}{3}(\alpha + \beta) + \frac{2}{3}(\alpha + \beta + \gamma)$
Total	$\frac{2}{3}\alpha + \frac{1}{3}(\alpha + \beta)$	$\frac{1}{3}(\alpha + \gamma) + \frac{2}{3}(\alpha + \beta + \gamma)$	

Now let $\beta = 0$ but $\gamma > 0$. Then we have

	Unskilled	Skilled	Total
Female	α	$\alpha + \gamma$	$\frac{2}{3}\alpha + \frac{1}{3}(\alpha + \gamma) = \alpha + \frac{1}{3}\gamma$
Male	α	$\alpha + \gamma$	$\frac{1}{3}\alpha + \frac{2}{3}(\alpha + \gamma) = \alpha + \frac{2}{3}\gamma$
Total	α	$\alpha + \gamma$	

Hence unskilled worker's earning is lower than skilled workers, and female earning is lower than male earning because of $\gamma > 0$ not of $\beta > 0$.

We can do the above analysis (very tedious) but rather also can do the following regression analysis to test if $\beta = 0$ but $\gamma \neq 0$

$$y_i = \alpha + \beta S_i + \gamma U_i + \varepsilon_i \tag{12}$$

Construct the following null hypothesis

$$\begin{aligned} H_0^1 &: \beta = 0 \\ H_0^2 &: \gamma = 0 \\ H_0^3 &: \beta = \gamma = 0 \end{aligned}$$

Further note that when we run

$$y_i = \alpha + \beta S_i + u_i \tag{13}$$

the OLS estimator $\hat{\beta}$ may not be zero. But the OLS estimator $\hat{\beta}$ in (12) could be zero. Why?

Omitted Variable If $E(S_i u_i) \neq 0$, or in other words, $E(S_i U_i) \neq 0$, then the OLS estimator $\hat{\beta}$ in (13) becomes inconsistent. (To students: Prove it)

Cross Dummies Suppose that among unskilled workers, there is no gender earning difference. However among skilled workers, there is gender earning difference. How to test?

	Expected earning
Female & unskilled; $S_i = U_i = 0$	$E(y_i) = \alpha$
Female & skilled; $S_i = 0$, $U_i = 1$	$E(y_i) = \alpha + \gamma$
Male & unskilled; $S_i = 1$, $U_i = 0$	$E(y_i) = \alpha + \beta$
Male & skilled; $S_i = 1$, $U_i = 1$	$E(y_i) = \alpha + \beta + \gamma + \delta$

In this case, we will have

$$y_i = \alpha + \beta S_i + \gamma U_i + \delta S_i U_i + e_i$$

Then test $\beta = 0$ but $\delta \neq 0$.

What is the meaning of $\delta = \gamma = 0$ but $\beta \neq 0$?

What is the meaning of $\delta = 0$ and $\beta = 0$?

What is the meaning of $\gamma = 0$ but $\delta \neq 0$ and $\beta \neq 0$?

Sequential Dummies Consider the following learning choice:

	Expected earning
Ph.D & taking Econometrics III	$E(y_i) = \alpha + \beta + \gamma$
Ph.D & not taking Econometrics III	$E(y_i) = \alpha + \beta$
No Ph.D	$E(y_i) = \alpha$

Construct dummy variable regression:

Another Example of Nonstochastic Regressor

$$y_t = \beta t + u_t, \quad u_t \sim iid\left(0, \sigma^2\right)$$

Q1: Find the limiting distribution of β.

$$\hat{\beta} - \beta = \frac{\sum t u_t}{\sum t^2}$$

Consider

$$E \sum t u_t = 0$$

$$\begin{aligned} E\left(\sum t u_t\right)^2 &= E\left(u_1 + 2u_2 + ... + T u_T\right)^2 \\ &= E\left(u_1^2 + 2^2 u_2^2 + ... + T^2 u_T^2\right) + 2E\left(2u_1 u_2 + ...\right) \\ &= \sigma^2\left(1 + 4 + ... + T^2\right) = \sigma^2 \sum t^2 \end{aligned}$$

Note that

$$\sum_{t=1}^{T} t^2 = \frac{1}{6} T\left(2T+1\right)\left(T+1\right)$$

Hence we have

$$\sum t u_t \to^d N\left(0, \sigma^2 \frac{T\left(2T+1\right)\left(T+1\right)}{6}\right)$$

and

$$\frac{\sum t u_t}{\sum t^2} = \frac{\sum t u_t}{\frac{1}{6} T\left(2T+1\right)\left(T+1\right)} \to^d N\left(0, \sigma^2 \frac{6}{T\left(2T+1\right)\left(T+1\right)}\right)$$

Now fine δ_T which makes

$$\delta_T \frac{\sum t u_t}{\sum t^2} \to^d N\left(0, \sigma^2\right)$$

The answer is

$$\delta_T = \sqrt{\frac{T\left(2T+1\right)\left(T+1\right)}{6}} = \sqrt{\frac{T^3}{3} + \frac{T^2}{6} + O\left(T\right)} = \sqrt{\frac{1}{3}} T^{3/2} + O\left(T\right)$$

Hence we have

$$T^{3/2}\left(\hat{\beta} - \beta\right) \to^d N\left(0, \frac{\sigma^2}{3}\right)$$

Linear and Nonlinear Restrictions (Chapter 5) Consider the following regression

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u} = \mathbf{x}_1 b_1 + \mathbf{x}_2 b_2 + \mathbf{u}$$

Then in general we have

$$\sqrt{n}\left(\hat{\mathbf{b}} - \mathbf{b}\right) \to^d N\left(0, \mathbf{\Sigma}_{\mathbf{b}}\right)$$

or

$$\sqrt{n}\begin{pmatrix} \hat{b}_1 - b_1 \\ \hat{b}_2 - b_2 \end{pmatrix} \to^d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{bmatrix}\right).$$

Next consider the following linear restriction

$$\alpha_0 \hat{b}_1 + \alpha_1 \hat{b}_2 = \alpha_2$$

Alternatively we may let

$$f\left(\hat{b}_1, \hat{b}_2\right) = \alpha_0 \hat{b}_1 + \alpha_1 \hat{b}_2 + \alpha_2 := \hat{\gamma} \tag{14}$$

Taking Taylor expansion around their true values yields

$$\begin{aligned} f\left(\hat{b}_1, \hat{b}_2\right) &= f\left(b_1, b_2\right) + \frac{\partial f\left(b_1, b_2\right)}{\partial b_1}\left(\hat{b}_1 - b_1\right) + \frac{\partial f\left(b_1, b_2\right)}{\partial b_2}\left(\hat{b}_2 - b_2\right) \\ &\quad + \frac{1}{2}\frac{\partial^2 f\left(b_1, b_2\right)}{\partial b_1^2}\left(\hat{b}_1 - b_1\right)^2 + \frac{1}{2}\frac{\partial^2 f\left(b_1, b_2\right)}{\partial b_2^2}\left(\hat{b}_2 - b_2\right)^2 \\ &\quad + \frac{1}{2}\frac{\partial^2 f\left(b_1, b_2\right)}{\partial b_1 \partial b_2}\left(\hat{b}_1 - b_1\right)\left(\hat{b}_2 - b_2\right) + \cdots . \end{aligned}$$

Note that from (14), we have

$$\frac{\partial f\left(b_1, b_2\right)}{\partial b_1} = \alpha_0, \quad \frac{\partial f\left(b_1, b_2\right)}{\partial b_2} = \alpha_1, \quad \frac{\partial^2 f\left(b_1, b_2\right)}{\partial b_1^2} = \frac{\partial^2 f\left(b_1, b_2\right)}{\partial b_2^2} = \frac{\partial^2 f\left(b_1, b_2\right)}{\partial b_1 \partial b_2} = 0, \tag{15}$$

so that

$$\begin{aligned} \hat{\gamma} &= \alpha_0 b_1 + \alpha_1 b_2 + \alpha_2 + \alpha_0\left(\hat{b}_1 - b_1\right) + \alpha_1\left(\hat{b}_2 - b_2\right) \\ &= \alpha_0 b_1 + \alpha_1 b_2 + \alpha_2 + \left(\alpha_0, \alpha_1\right)\begin{bmatrix} \left(\hat{b}_1 - b_1\right) \\ \left(\hat{b}_2 - b_2\right) \end{bmatrix} \end{aligned}$$

or

$$\begin{aligned}
\sqrt{n}\left(\hat{\gamma}-\gamma\right) &= \left[\begin{array}{cc} \alpha_0 & \alpha_1 \end{array}\right] \left[\begin{array}{c} \sqrt{n}\left(\hat{b}_1 - b_1\right) \\ \sqrt{n}\left(\hat{b}_2 - b_2\right) \end{array}\right] \\
&\rightarrow {}^d N\left(0, \left[\begin{array}{cc} \alpha_0 & \alpha_1 \end{array}\right] \left[\begin{array}{cc} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{array}\right] \left[\begin{array}{c} \alpha_0 \\ \alpha_1 \end{array}\right]\right) \\
&= N\left(0, \sigma_\gamma^2\right)
\end{aligned}$$

Finally we have

$$\frac{\sqrt{n}\left(\hat{\gamma}-\gamma\right)}{\sqrt{\sigma_\gamma^2}} \rightarrow^d N(0,1)$$

Now compare this limiting result with Greene. (page 84) Let

$$\mathbf{R} = \left[\begin{array}{cc} \alpha_0 & \alpha_1 \end{array}\right] \text{ and } \mathbf{q} = \alpha_2$$

Then we have

$$W = (\mathbf{Rb}-\mathbf{q})' \left[\sigma^2 \mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right]^{-1} (\mathbf{Rb}-\mathbf{q}) \rightarrow^d \chi_1^2$$

Note that $\sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \mathbf{\Sigma_b}$ in our notation.

Next we consider a nonlinear restriction. Usually for a nonlinear restriction case, the second and cross terms in (14) are not equal to zero but become small terms. To see this, consider

$$f\left(\hat{b}_1, \hat{b}_2\right) - f\left(b_1, b_2\right) = \frac{\partial f\left(b_1, b_2\right)}{\partial b_1}\left(\hat{b}_1 - b_1\right) + \frac{\partial f\left(b_1, b_2\right)}{\partial b_2}\left(\hat{b}_2 - b_2\right) + R_n$$

or

$$\hat{\gamma} - \gamma = \frac{\partial f\left(b_1, b_2\right)}{\partial b_1}\left(\hat{b}_1 - b_1\right) + \frac{\partial f\left(b_1, b_2\right)}{\partial b_2}\left(\hat{b}_2 - b_2\right) + R_n$$

where R_n is the remainder term. Note that $\sqrt{n}\left(\hat{b}_1 - b_1\right)$ is $O_p(1)$, so that $\left(\hat{b}_1 - b_1\right)$ is $O_p\left(n^{-1/2}\right)$, $\left(\hat{b}_1 - b_1\right)^2$ is $O_p\left(n^{-1}\right)$ and $\left(\hat{b}_1 - b_1\right)\left(\hat{b}_2 - b_2\right) = O_p\left(n^{-1}\right)$. Hence we have

$$R_n = O_p\left(n^{-1}\right)$$

Therefore,

$$\sqrt{n}\left(\hat{\gamma}-\gamma\right) = \frac{\partial f\left(b_1, b_2\right)}{\partial b_1}\sqrt{n}\left(\hat{b}_1 - b_1\right) + \frac{\partial f\left(b_1, b_2\right)}{\partial b_2}\sqrt{n}\left(\hat{b}_2 - b_2\right) + O_p\left(\frac{1}{\sqrt{n}}\right),$$

and

$$\begin{aligned}
\sqrt{n}\left(\hat{\gamma}-\gamma\right) &= \left[\begin{array}{cc} \frac{\partial f(b_1,b_2)}{\partial b_1} & \frac{\partial f(b_1,b_2)}{\partial b_2} \end{array}\right]\left[\begin{array}{c} \sqrt{n}\left(\hat{b}_1-b_1\right) \\ \sqrt{n}\left(\hat{b}_2-b_2\right) \end{array}\right]+O_p\left(\frac{1}{\sqrt{n}}\right) \\
&\rightarrow {}^d N\left(0,\left[\begin{array}{cc} \frac{\partial f(b_1,b_2)}{\partial b_1} & \frac{\partial f(b_1,b_2)}{\partial b_2} \end{array}\right]\left[\begin{array}{cc} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{array}\right]\left[\begin{array}{c} \frac{\partial f(b_1,b_2)}{\partial b_1} \\ \frac{\partial f(b_1,b_2)}{\partial b_2} \end{array}\right]\right)
\end{aligned}$$

Example: Suppose that you want to test if

$$\frac{\hat{b}_1}{\hat{b}_2}=0.$$

Then we have

$$\frac{\partial f\left(b_1,b_2\right)}{\partial b_1}=\frac{1}{b_2},\quad \frac{\partial f\left(b_1,b_2\right)}{\partial b_2}=-\frac{b_1}{b_2^2}$$

6 Time Series Models (Ref: Chap 19 & 21)

Consider the following regression

$$y_t = bx_t + u_t, \quad t = 1, ..., T$$

where

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid\left(0, \sigma^2\right)$$

Let's assume that $E(x_t u_s) = E(x_t x_s) = 0$ **for all** t **and** s. Q1: Find the limiting distribution of $\hat{b}$.

We know

$$\hat{b} - b = \frac{\sum x_t u_t}{\sum x_t^2},$$

and

$$E\left(\sum x_t u_t\right) = 0.$$

Consider

$$E\left(\sum x_t u_t\right)^2.$$

Observe this

$$\left(\sum x_t u_t\right)^2 = (x_1 u_1 + ... + x_T u_T)^2 = \sum x_t^2 u_t^2 + 2\left(x_1 u_1 x_2 u_2 + ... + x_{T-1} u_{T-1} x_T u_T\right) \quad (16)$$

Now

$$E \sum x_t^2 u_t^2 = \sum E x_t^2 E u_t^2$$

To calculate Eu_t^2, consider the folllowings

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad u_{t-1} = \rho u_{t-2} + \varepsilon_{t-1}$$

so that

$$\begin{aligned} u_t &= \rho^2 u_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t \\ &= \rho^3 u_{t-3} + \rho^2 \varepsilon_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t \\ &\vdots \\ &= \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + ... = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j} \end{aligned}$$

Next,

$$u_t^2 = \left(\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + ...\right)^2 = \sum_{j=0}^{\infty} \rho^{2j}\varepsilon_{t-j}^2 + \text{cross product terms}$$

so that

$$Eu_t^2 = E\left(\varepsilon_t^2 + \rho^2\varepsilon_{t-1}^2 + \rho^4\varepsilon_{t-2}^2 + ...\right) + E\left(\text{cross}\right)$$

Since

$$E\varepsilon_t\varepsilon_s = 0 \text{ for } t \neq s,\ E\left(\text{cross}\right) = 0.$$

Hence we have

$$\begin{aligned} Eu_t^2 &= E\left(\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + ...\right)^2 \\ &= E\left(\varepsilon_t^2 + \rho^2\varepsilon_{t-1}^2 + \rho^4\varepsilon_{t-2}^2 + ...\right) \\ &= \sigma^2\left(1 + \rho^2 + \rho^4 + ...\right) \end{aligned}$$

Note that

$$\begin{aligned} 1 + \rho^2 + \rho^4 + ... &= \frac{1}{1-\rho^2}, \\ 1 + \rho + \rho^2 + ... &= \frac{1}{1-\rho}, \\ 1 + \rho + \rho^2 + ... + \rho^T &= \frac{1-\rho^{T+1}}{1-\rho}. \end{aligned}$$

Finally

$$Eu_t^2 = \frac{\sigma^2}{1-\rho^2} = \sigma_u^2.$$

Also note that

$$\begin{aligned} Eu_tu_{t-1} &= E\left(\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + ...\right)\left(\varepsilon_{t-1} + \rho\varepsilon_{t-2} + \rho^2\varepsilon_{t-3} + ...\right) \\ &= \sigma^2\left(\rho + \rho^3 + ...\right) = \sigma^2\rho\left(1 + \rho^2 + ...\right) = \frac{\sigma^2}{1-\rho^2}\rho \\ &= \rho\left(Eu_t^2\right), \end{aligned}$$

and

$$\begin{aligned} Eu_tu_{t-2} &= E\left(\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + ...\right)\left(\varepsilon_{t-2} + \rho\varepsilon_{t-3} + \rho^2\varepsilon_{t-4} + ...\right) \\ &= \sigma^2\left(\rho^2 + \rho^4 + ...\right) = \sigma^2\rho^2\left(1 + \rho^2 + ...\right) = \frac{\sigma^2}{1-\rho^2}\rho^2 \\ &= \rho^2\left(Eu_t^2\right). \end{aligned}$$

In general

$$Eu_tu_{t-k} = Eu_{t+k}u_t = \rho^k Eu_t^2 = \rho^k \sigma_u^2$$

Then we have

$$E\left(\sum x_t u_t\right)^2 = \sum Ex_t^2 Eu_t^2 + 2E\left(x_1u_1x_2u_2 + ... + x_{T-1}u_{T-1}x_Tu_T\right) = T\sigma_x^2\sigma_u^2$$

Hence there is no much difference.

Let's assume that $E(x_tu_s) = 0$ for all t and s but $E(x_tx_s) = \rho^{t-s}E(x_t^2)$. Then we have

$$\begin{aligned} Ex_sx_tu_su_t &= \rho^{t-s}E\left(x_t^2\right)\rho^{t-s}E\left(u_t^2\right) \\ &= \rho^{2(t-s)}\sigma_x^2\sigma_u^2 \end{aligned}$$

Consider the cross product term carefully

$$\begin{aligned} \text{Cross Term} &= x_1u_1\left(x_2u_2 + ... + x_Tu_T\right) \\ &\quad +x_2u_2\left(x_1u_1 + x_3u_3... + x_Tu_T\right) \\ &\quad +... \\ &\quad +x_Tu_T\left(x_1u_1 + ... + x_{T-1}u_{T-1}\right) \end{aligned}$$

Hence

$$\begin{aligned} E\text{Cross Term} &= Ex_1u_1\left(x_2u_2 + ... + x_Tu_T\right) \\ &\quad +Ex_2u_2\left(x_1u_1 + x_3u_3... + x_Tu_T\right) \\ &\quad +... \\ &\quad +Ex_Tu_T\left(x_1u_1 + ... + x_{T-1}u_{T-1}\right) \end{aligned}$$

where

$$\begin{aligned} Ex_1u_1\left(x_2u_2 + ... + x_Tu_T\right) &= \rho^2\sigma_x^2\sigma_u^2 + \rho^4\sigma_x^2\sigma_u^2 + ... + \rho^{2(T-1)}\sigma_x^2\sigma_u^2 \\ &= \sigma_x^2\sigma_u^2\rho^2\left(1 + \rho^2 + ... + \rho^{2(T-2)}\right) = \sigma_x^2\sigma_u^2\rho^2\frac{1-\rho^{2(T-1)}}{1-\rho^2}, \end{aligned}$$

$$\begin{aligned} Ex_2u_2\left(x_1u_1+x_3u_3...+x_Tu_T\right) &= \rho^2\sigma_x^2\sigma_u^2+\rho^2\sigma_x^2\sigma_u^2+...+\rho^{2(T-2)}\sigma_x^2\sigma_u^2 \\ &= \sigma_x^2\sigma_u^2\rho^2\left(2+\rho^2+...+\rho^{2(T-3)}\right) \\ &= \sigma_x^2\sigma_u^2\rho^2\frac{1-\rho^{2(T-2)}}{1-\rho^2}+\sigma_x^2\sigma_u^2\rho^2, \end{aligned}$$

$$\begin{aligned} & Ex_3u_3\left(x_1u_1+x_2u_2+x_4u_4+...+x_Tu_T\right) \\ =\; & \rho^4\sigma_x^2\sigma_u^2+\rho^2\sigma_x^2\sigma_u^2+\rho^2\sigma_x^2\sigma_u^2+...+\rho^{2(T-3)}\sigma_x^2\sigma_u^2 \\ =\; & \sigma_x^2\sigma_u^2\rho^2\left(\rho^2+2+\rho^2+...+\rho^{2(T-4)}\right) \\ =\; & \sigma_x^2\sigma_u^2\rho^2\frac{1-\rho^{2(T-3)}}{1-\rho^2}+\sigma_x^2\sigma_u^2\rho^2\left(1+\rho^2\right), \end{aligned}$$

$$Ex_Tu_T\left(x_1u_1+...+x_{T-1}u_{T-1}\right)=\rho^{2T-2}\sigma_x^2\sigma_u^2+...+\rho^2\sigma_x^2\sigma_u^2=\sigma_x^2\sigma_u^2\rho^2\frac{1-\rho^{2(T-1)}}{1-\rho^2}$$

Hence the total sum becomes

$$2\sigma_x^2\sigma_u^2\rho^2\sum_{i=1}^{T}\frac{1-\rho^{2(T-i)}}{1-\rho^2}=2\frac{\sigma_x^2\sigma_u^2\rho^2}{1-\rho^2}T+O\left(1\right)$$

Or

$$\begin{aligned} E\left(\sum x_tu_t\right)^2 &= \sum Ex_t^2Eu_t^2+2E\left(x_1u_1x_2u_2+...+x_{T-1}u_{T-1}x_Tu_T\right) \\ &= T\sigma_x^2\sigma_u^2+2\frac{\sigma_x^2\sigma_u^2\rho^2}{1-\rho^2}T+O\left(1\right) \\ &= T\sigma_x^2\sigma_u^2\left(\frac{1+\rho^2}{1-\rho^2}\right)+O\left(1\right) \end{aligned}$$

Finally we have

$$\sqrt{T}\left(\hat{b}-b\right)\to^d N\left(0,\omega_b^2\right)$$

where

$$\omega_b^2=\frac{\sigma_x^2\sigma_u^2\left(\frac{1+\rho^2}{1-\rho^2}\right)}{\sigma_x^2\sigma_x^2}=\left(\frac{1+\rho^2}{1-\rho^2}\right)\frac{\sigma_u^2}{\sigma_x^2}\geq\frac{\sigma_u^2}{\sigma_x^2}$$

In other words, the typical limiting distribution such as

$$\sqrt{T}\left(\hat{b}-b\right)\to^d N\left(0,\sigma_u^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$$

does not work here.

6.1 Definitions

1. Strong Stationary: A time-series process, $\{z_t\}_{t=-\infty}^{t=\infty}$ is stronlgy stationary if the joint probability distribution of any set of k obersvations in the sequence $\{z_t, ..., z_{t+k}\}$ is the same regardless of the origin t, in the time scale.

2. Weak Stationary: $\{z_t\}$ is weakly stationary if (i) $E(z_t)$ is finite, (ii) $Cov(z_t, z_{t-k})$ is a finite function only of k and model parameters. (In other words, it should not be time varying)

3. Ergodicity: A strongly stationary time series process is ergodic if

$$\begin{aligned} & \lim_{k\to\infty} |E[f(z_t, z_{t+1}, ..., z_{t+a}) g(z_{t+k}, z_{t+k+1}, ..., z_{t+k+b})]| \\ = & \ |Ef(z_t, z_{t+1}, ..., z_{t+a})| \, |Eg(z_{t+k}, z_{t+k+1}, ..., z_{t+k+b})| \end{aligned}$$

   (a) Example: Let $z_t = \rho z_{t-1} + u_t, \ \ u_t \sim iid(0,1)$

$$\lim_{k\to\infty} |E(z_t z_{t+k})| = \lim_{k\to\infty} \left|\rho^k \sigma_z^2\right| = 0 = |Ez_t| \, |Ez_{t+k}|$$

4. The Ergodic Theorem: If z_t is strongly stationary and ergodic and $E|z_t|$ is a finite constant, then $\bar{z}_T = T^{-1} \sum z_t \to^{a.s.} \mu = E(z_t)$.

5. Martingale Sequence: z_t is a martingale sequence if

$$E(z_t | z_{t-1}, z_{t-2}, ...) = z_{t-1}$$

   (a) Example: $z_t = z_{t-1} + u_t, \ \ E(z_t | z_{t-1}, z_{t-2}, ...) = z_{t-1}$

6. Martingale Difference Sequence: z_t is a martingale difference sequence if

$$E(z_t | z_{t-1}, z_{t-2}, ...) = 0$$

7. White Noise process: stationary but not-autocorrelated process.

6.2 Long Run Variance

Q1: Consider $u_t = \rho u_{t-1} + e_t,\ \ e_t$ is a white noise process with a finite variance of σ^2. Find the limiting distribution of the sample mean of u_t.

$$\mu_T = \frac{1}{T}\sum_{t=1}^{T} u_t$$

Mean: $E\mu_T = 0$.

Variance:

$$\begin{aligned}
E\left(\frac{1}{T}\sum_{t=1}^{T}u_t\right)^2 &= \frac{1}{T^2}E\left(u_1+...+u_T\right)\left(u_1+...+u_T\right) = \frac{1}{T^2}\left[\sum_{t=1}^{T}Eu_t^2 + 2E\sum_{t=1}^{T-1}\sum_{s=t}^{T}u_t u_s\right] \\
&= \frac{1}{T^2}\left[T\sigma_u^2 + 2\sum_{t=1}^{T-1}\sum_{s=t}^{T}\rho^{t-s}\sigma_u^2\right] = \frac{1}{T^2}\sigma_u^2\left[T + 2\frac{\rho}{1-\rho}T + O\left(1\right)\right] \\
&= \frac{1}{T}\frac{\sigma^2}{1-\rho^2}\left(1+2\frac{\rho}{1-\rho} + O\left(T^{-1}\right)\right) = \frac{1}{T}\frac{\sigma^2}{1-\rho^2}\left(\frac{1+\rho}{1-\rho} + O\left(T^{-1}\right)\right) \\
&= \frac{1}{T}\frac{\sigma^2}{\left(1-\rho\right)^2} + O\left(T^{-2}\right) := \frac{1}{T}\omega^2 \quad \text{omega}
\end{aligned}$$

Hence we have

$$d\mu_T \to^d N\left(0, \frac{1}{T}\omega^2\right)$$

or

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t \to^d N\left(0, \frac{\sigma^2}{\left(1-\rho\right)^2}\right) \tag{17}$$

We call ω^2 long run variance of u_t.

6.3 Estimation of Long Run Variance (HAC Estimation)

How to estimate the long run variance of u_t in (17) then? The unknowns are σ^2 and ρ. How many observations do we have? T. So it is easy to estimate it.

Now what if the parametric structure is unknown. Let say u_t follows $AR(T)$ or $ARMA(p,q)$ where p and q are unknown? Is it possible to estimate ω^2? No. The total number of unknowns becomes $\frac{T(T-1)}{2}+1$. The first term is the sum of cross product terms and the last term, 1, is the unknown variance term (diagonal term). If variance is time varying, then it becomes $\frac{T(T-1)}{2}+T$. Simply impossible to estimate the long run variance in this case.

Therefore we are imposing regularity: Ergodic and stationary process. And then we assume that

$$E(u_t u_{t-k}) \simeq 0 \text{ for a large } k.$$

Alternatively let say

$$\begin{aligned} E\frac{1}{T}\sum_{t=1}^{T-1}\sum_{s=t}^{T} u_t u_s &= E\frac{1}{T}\sum_{t=1}^{T-1}\sum_{s=t}^{t+k} u_t u_s + E\frac{1}{T}\sum_{t=1}^{T-1}\sum_{s=t+k+1}^{T} u_t u_s \\ &= E\frac{1}{T}\sum_{t=1}^{T-1}\sum_{s=t}^{t+k} u_t u_s + o_p(1) \end{aligned} \tag{18}$$

In this case, we don't need to estimate the second term.

Newey and West Estimator Let

$$\omega^2 = \omega_0^2 + \sum_{j=1}^{\infty}\left(\omega_j^2 + \omega_{-j}^2\right)$$

where

$$\omega_j^2 = E u_t u_{t-j}$$

Then we can apply the above concept in (18), so we have

$$\hat{\omega}^2 = \hat{\omega}_0^2 + \sum_{j=1}^{k}\left(\hat{\omega}_j^2 + \hat{\omega}_{-j}^2\right)$$

According to Andrews (1991), we can modify the estimator further in an elegant way

$$\hat{\omega}^2 = \hat{\omega}_0^2 + \sum_{j=1}^{k} w_j \left(\hat{\omega}_j^2 + \hat{\omega}_{-j}^2 \right)$$

where w_j is some optimal weight. Newey and West (1992) suggest

$$w_j = 1 - \frac{j}{k+1}, \quad k = int\left(T^{1/3}\right).$$

We call such weight Bartlett kernel weight. They show that this type of estimator becomes consistent.

Parametric Version: Andrews and Monahan's Prewhitening HAC estimator Let e_t is a stationary and ergodic process. Then we may have

$$u_t = \rho u_{t-1} + e_t$$

and

$$E\left(\frac{1}{\sqrt{T}} \sum u_t\right)^2 = \frac{\omega_e^2}{(1-\rho)^2}$$

where ω_e^2 is the long run variance of e_t. Now we estimate $\hat{\rho}$ and replace this. That is,

$$\hat{\omega}_u^2 = \frac{\hat{\omega}_e^2}{(1-\hat{\rho})^2}.$$

Conversion to Matrix Form Consider

$$y_t = \mathbf{X}_t' \mathbf{b} + u_t$$

$$\sqrt{T}\left(\hat{\mathbf{b}} - \mathbf{b}\right) \to^d N\left(0, \mathbf{V}_{\hat{b}}\right)$$

where

$$\mathbf{V}_{\hat{b}} = \left(\frac{1}{T} \sum \mathbf{X}_t \mathbf{X}_t'\right)^{-1} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} E u_s \mathbf{X}_s \left(u_t \mathbf{X}_t\right)' \left(\frac{1}{T} \sum \mathbf{X}_t \mathbf{X}_t'\right)^{-1}.$$

Now let

$$\boldsymbol{\xi}_t = u_t \cdot \mathbf{X}_t = \left(u_t x_{1t}, u_t x_{2t}, ..., u_t x_{kt}\right)$$

Then

$$\Omega^2 = \Omega_0 + \Omega_j + \Omega_{-j}$$

$$\hat{\Omega}^2 = \hat{\Omega}_0 + \sum_{j=1}^{k} w_j \left(\hat{\Omega}_j + \hat{\Omega}_{-j} \right)$$

Then we have

$$\hat{\mathbf{V}}_{\hat{b}} = \left(\frac{1}{T} \sum \mathbf{X}_t \mathbf{X}_t' \right)^{-1} \hat{\Omega}^2 \left(\frac{1}{T} \sum \mathbf{X}_t \mathbf{X}_t' \right)^{-1}.$$

Alternative Approach Let assume

$$y_t = \mathbf{X}_t'\mathbf{b} + u_t, \quad u_t = \sum_{j=1}^{p} \rho_j u_{t-j} + e_t$$

Then

$$\begin{aligned} \rho_1 y_{t-1} &= \rho_1 \mathbf{X}_{t-1}'\mathbf{b} + \rho_1 u_{t-1} \\ &\vdots \\ \rho_p y_{t-p} &= \rho_p \mathbf{X}_{t-p}'\mathbf{b} + \rho_p u_{t-p} \end{aligned}$$

Now subtract $\rho y_{t-1}, ...\rho_p y_{t-p}$ from y_t.

$$\begin{aligned} y_t &= \mathbf{X}_t'\mathbf{b} - \sum_{j=1}^{p} \rho_j \mathbf{X}_{t-j}'\mathbf{b} + \sum_{j=1}^{p} \rho_j y_{t-j} + \mathbf{b}u_t - \sum_{j=1}^{p} \rho_j u_{t-j} \\ &= \mathbf{X}_t'\mathbf{b} - \sum_{j=1}^{p} \rho_j \mathbf{X}_{t-j}'\mathbf{b} + \sum_{j=1}^{p} \rho_j y_{t-j} + e_t = \mathbf{Z}_t\boldsymbol{\gamma} + e_t \end{aligned}$$

where $Z_t = \left(\mathbf{X}_t, \mathbf{X}_{t-1}, ..., \mathbf{X}_{t-p}, y_{t-1}, ..., y_{t-p}\right)$. Let rewrite it as

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e},$$

and then we have

$$\sqrt{T}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) \to^d N\left(0, \sigma_e^2 Q_Z^{-1}\right).$$

where

$$Q_Z = \operatorname{plim}_{T\to\infty} \frac{\mathbf{Z}'\mathbf{Z}}{T}$$

Conventional Approach (Generalized Least Squares GLS: Chapter 8) Suppose that we know the AR order. Let say AR(1). Then we have

$$u_t = \rho u_{t-1} + e_t$$

so that

$$\begin{aligned} E\mathbf{u}\mathbf{u}' &= \Omega_{T\times T} = \sigma_e^2 \begin{bmatrix} \frac{1}{1-\rho^2} & \frac{\rho}{1-\rho^2} & \cdots & \frac{\rho^{T-1}}{1-\rho^2} \\ \frac{\rho}{1-\rho^2} & \frac{1}{1-\rho^2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \frac{\rho}{1-\rho^2} \\ \frac{\rho^{T-1}}{1-\rho^2} & \cdots & \frac{\rho}{1-\rho^2} & \frac{1}{1-\rho^2} \end{bmatrix} \\ &= \frac{\sigma_e^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \rho \\ \rho^{T-1} & \cdots & \rho & 1 \end{bmatrix} \end{aligned}$$

Now we know

$$\Omega = \mathbf{C}\mathbf{\Lambda}\mathbf{C}'$$

where $\mathbf{C}'\mathbf{C} = \mathbf{I}$, and

$$\begin{aligned} \Omega^{-1} &= \mathbf{C}\mathbf{\Lambda}^{-1}\mathbf{C}' \\ &= \mathbf{P}'\mathbf{P} \end{aligned}$$

where

$$\mathbf{P} = \mathbf{\Lambda}^{-1/2}\mathbf{C}'$$

Next consider the following transformation

$$\mathbf{Py} = \mathbf{PXb} + \mathbf{Pu}$$

or

$$\mathbf{y}^* = \mathbf{X}^*\mathbf{b} + \mathbf{u}^* \tag{19}$$

Now define the GLS estimator

$$\hat{\mathbf{b}}_{gls} = \left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^*$$

or alternatively we can say

$$\mathbf{X}^{*\prime}\mathbf{X}^{*} = \mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{X} = \mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}$$

and

$$\mathbf{X}^{*\prime}\mathbf{y}^{*} = \mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{y} = \mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$$

so that we have

$$\hat{\mathbf{b}}_{gls} = \left(\mathbf{X}'\Omega^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y}$$

and find its limiting distribution.

First note that

$$E\mathbf{u}^{*}\mathbf{u}^{*\prime} = \mathbf{P}E\mathbf{u}\mathbf{u}'\mathbf{P}' = \mathbf{P}\boldsymbol{\Omega}\mathbf{P}' = \boldsymbol{\Lambda}^{-1/2}\mathbf{C}'\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}'\mathbf{C}\boldsymbol{\Lambda}^{-1/2} = \mathbf{I}$$

Hence the limiting distribution of $\hat{\mathbf{b}}_{gls}$ is given by

$$\sqrt{n}\left(\hat{\mathbf{b}}_{gls} - \mathbf{b}\right) \to^{d} N\left(0, \left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^{*}}{n}\right)^{-1}\right)$$

or

$$\sqrt{n}\left(\hat{\mathbf{b}}_{gls} - \mathbf{b}\right) \to^{d} N\left(0, \left(\frac{\mathbf{X}'\Omega^{-1}\mathbf{X}}{n}\right)^{-1}\right)$$

Feasible GLS Replace Ω by $\hat{\Omega}$.

$$\hat{\mathbf{b}}_{fgls} = \left(\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\Omega}^{-1}\mathbf{y}$$

7 Heteroskedasticity (Chapter 8 Continue)

Now we allow heterogenous variance for each i or t. That is,

$$Eu_i^2 = \sigma_i^2 \neq \sigma_j^2 = Eu_j^2$$

However we assume that

$$Eu_i u_j = 0.$$

Then we have

$$E\mathbf{u}\mathbf{u}' = \Omega_{T\times T} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n^2 \end{bmatrix}$$

Note that

$$\mathbf{X}'E\mathbf{u}\mathbf{u}'\mathbf{X} = \mathbf{X}'\boldsymbol{\Omega}\mathbf{X} \neq \mathbf{X}'\mathbf{X}.$$

But in this case, we have

$$\begin{aligned} [\mathbf{X}_1, ..., \mathbf{X}_n] \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n^2 \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \\ \mathbf{X}_n \end{bmatrix} &= \sigma_1^2\mathbf{X}_1'\mathbf{X}_1 + \sigma_2^2\mathbf{X}_2'\mathbf{X}_2 + ... + \sigma_n^2\mathbf{X}_n'\mathbf{X}_n \\ &= \sum_{i=1}^n \sigma_i^2 \mathbf{X}_i'\mathbf{X}_i \end{aligned}$$

Therefore we have

$$\sqrt{n}\left(\hat{\mathbf{b}} - \mathbf{b}\right) \to^d N\left(0, \mathbf{V}_b\right)$$

where

$$\begin{aligned} \mathbf{V}_b &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\Omega\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\sum_{i=1}^n \sigma_i^2\mathbf{X}_i'\mathbf{X}_i\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1} \end{aligned}$$

If we replace σ_i^2 but $\hat{\sigma}_i^2 = \hat{u}_i^2$, then we call this estimator 'White' heteroskedasticity consistent estimator.

8 Instrumental Variables

Consider the following data generating process

$$y_i = ax_i + u_i$$

where

$$u_i = \beta x_i + e_i$$

we assume that $E(x_i e_j) = 0$ for all i and j.

Now we have

$$\hat{a} = a + (x'x)^{-1} x'u = a + \beta + (x'x)^{-1} x'e$$

so that

$$E(\hat{a} - a|x) = \beta \neq 0.$$

We say that x is endogeneous in this case. Note that the concept of endogeneity is in general somewhat different. We will explain it later in Chapter 20.

8.1 Can we know if $\beta = 0$ or not?

1. Hausman Test: testing for exogeneity. We will study it later.

2. u is unknown. How do we know if x is correlated with u ?

3. Known case: Lagged dependent variable.

   $$y_t = a + y_t^o, \; y_t^o = \rho y_{t-1}^o + u_t$$

   so that

   $$y_t = a(1-\rho) + \rho y_{t-1} + u_t$$

   Then we can rewrite it as

   $$\tilde{y}_t = \rho \tilde{y}_{t-1} + \tilde{u}_t.$$

   Note that $E(\tilde{y}_{t-1}\tilde{u}_t) \neq 0$. However as $t \to \infty$, this bias goes away at the $O_p(T^{-1})$ rate.

4. Measurement error: True model

$$y_i = \alpha x_i + u_i \tag{20}$$

But we observe $x_i^* = x_i + e_i$. So you run

$$y_i = \alpha x_i^* + v_i.$$

From (20), we haev

$$y_i = \alpha (x_i + e_i) - \alpha e_i + u_i = \alpha x_i^* + v_i$$

Now $E(v_i x_i^*) = E(u_i - \alpha e_i)(x_i + e_i) \neq 0$.

8.2 Solution I

Including control variables.

$$y_i = \alpha x_i + \mathbf{w}_i' \boldsymbol{\gamma} + v_i$$

where $\mathbf{w}_i = (w_{1i}, ..., w_{ki})'$. Now $\mathbf{w}_i$ becomes a proxy variable for u_i.

Problem: We don't know how many control variables should be included.

8.3 Solution II

Construct instrumental variable, z_i such that

$$E(x_i z_i) \neq 0$$

but

$$E(z_i u_i) = 0.$$

Then construct IV estimator

$$\begin{aligned} \hat{\alpha}_{IV} &= (\mathbf{z}'\mathbf{x})^{-1} \mathbf{z}'\mathbf{y} \\ &= (\mathbf{z}'\mathbf{x})^{-1} \mathbf{z}' (\mathbf{x}\boldsymbol{\alpha} + \mathbf{u}) \\ &= \alpha + (\mathbf{z}'\mathbf{x})^{-1} \mathbf{z}'\mathbf{u} \end{aligned}$$

Next,

$$\hat{\alpha}_{IV} - \alpha = \left(\mathbf{z}'\mathbf{x}\right)^{-1}\mathbf{z}'\mathbf{u}$$

and

$$\begin{aligned}\operatorname{plim}\left(\hat{\alpha}_{IV} - \alpha\right) &= \operatorname{plim}\left(\frac{\mathbf{z}'\mathbf{x}}{n}\right)^{-1}\operatorname{plim}\frac{\mathbf{z}'\mathbf{u}}{n} \\ &= Q_{zx} \cdot 0 = 0\end{aligned}$$

so that $\hat{\alpha}_{IV}$ is a consistent estimator of α.

Asymptotic variance:

$$E\left(\hat{\alpha}_{IV} - \alpha\right)\left(\hat{\alpha}_{IV} - \alpha\right)' = E\left[\left(\mathbf{z}'\mathbf{x}\right)^{-1}\mathbf{z}'\mathbf{u}\mathbf{u}'\mathbf{z}\left(\mathbf{z}'\mathbf{x}\right)^{-1}\right]$$

If z and x are non-stochastic, we have

$$E\left(\hat{\alpha}_{IV} - \alpha\right)\left(\hat{\alpha}_{IV} - \alpha\right)' = \left(z'x\right)^{-1}z'\Omega_u z\left(z'x\right)^{-1}$$

8.3.1 Getting into details: Measurement error

$$y_i = \alpha x_i^* + v_i, \quad x_i^* = x_i + e_i, \quad v_i = -\alpha e_i + u_i$$

Find a variable such that

$$z_i = \beta x_i + m_i$$

but

$$Em_i e_i = 0 \text{ and } Em_i u_i = 0.$$

Then z_i is the right instrumental variable.

How to find such a good IV then? Ask GOD.

9 Method of Moments (Chap 15)

Consider moment conditions such that

$$E\left(\xi_t - \mu\right) = 0$$

where ξ_t is a random variable and μ is the unknown mean of ξ_t. The parameter of interest, here, is μ. Consider the following minimum criteria given by

$$\arg\min_{\mu} V_T = \arg\overset{2}{\min_{\mu}} \frac{1}{T}\sum_{t=1}^{T}\left(\xi_t - \mu\right)$$

which becomes the minimum variance of ξ_t with respect to μ. Of course, the simple solution becomes the sample mean for μ since we have

$$\frac{\partial V_T}{\partial \mu} = -2\frac{1}{T}\sum_{t=1}^{T}\left(\xi_t - \mu\right) = 0, \quad \Longrightarrow \frac{1}{T}\sum_{t=1}^{T}\xi_t = \mu$$

The above case is the simple example of the method of moment(s).

Now consider more moments such that

$$\begin{aligned}
E\left(\xi_t - \mu\right) &= 0 \\
E\left[\left(\xi_t - \mu\right)^2 - \gamma_0\right] &= 0 \\
E\left[\left(\xi_t - \mu\right)\left(\xi_{t-1} - \mu\right) - \gamma_1\right] &= 0 \\
E\left[\left(\xi_t - \mu\right)\left(\xi_{t-2} - \mu\right) - \gamma_2\right] &= 0
\end{aligned}$$

Then we have the four unknowns: $\mu, \gamma_0, \gamma_1, \gamma_2$. We have four sample moments such that

$$\frac{1}{T}\sum_{t=1}^{T}\xi_t, \frac{1}{T}\sum_{t=1}^{T}\xi_t^2, \frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_{t-1}, \frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_{t-2}$$

so that we can solve this numerically.

However, we want to impose further restriction. Suppose that we assume ξ_t follows AR(1) process. Then we have

$$\gamma_1 = \rho\gamma_0, \quad \gamma_2 = \rho\gamma_0$$

so that the total number of unknowns is reducing to three (γ_0, ρ, μ). We can increase more cross moment conditions also. Let $\psi_T = \left(\frac{1}{T}\sum_{t=1}^T \xi_t, \frac{1}{T}\sum_{t=1}^T \xi_t^2, \frac{1}{T}\sum_{t=1}^T \xi_t\xi_{t-1}, \frac{1}{T}\sum_{t=1}^T \xi_t\xi_{t-2}\right)'$. Then we have

$$E\frac{1}{T}\sum_{t=1}^{T}(\xi_t-\mu)^2 = E\frac{1}{T}\sum_{t=1}^{T}\xi_t^2-\mu^2 = \gamma_0$$

so that

$$E\frac{1}{T}\sum_{t=1}^{T}\xi_t^2 = \gamma_0-\mu^2$$

Also note that

$$E\frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_{t-1} = \rho\gamma_0-\mu^2, \text{ and so on.}$$

Hence we may consider the following estimation

$$\arg\min_{\mu,\rho,\gamma_0}\left[\psi_T-\psi(\theta)\right]'\left[\psi_T-\psi(\theta)\right]. \tag{21}$$

where θ is the parameters of interest (true parameters, μ, γ_0, ρ). The resulting estimator is called 'method of moments estimator'. Note that MM estimator is a kind of minimum distance estimators.

In general, MM estimator can be used in many cases. However, this method has one weakness. Suppose that the second moment is relatively huge than the first moment. Since V_T function assigns the same weight across moments, the minimum problem in (21) tries to minimize the second moment rather than the first and second moment both. Hence we need to design the optimal weighted method of moments, which becomes generalized method of moments (GMM).

To understand the nature of GMM, we have to study the asymptotic properties of MM estimator. (in order to find the optimal weighting matrix). Now to get the asymptotic distribution of $\hat{\theta}$, we need a Taylor expansion.

$$\psi_T = \psi(\theta)+\frac{\partial\psi_T(\theta)}{\partial\theta'}\left(\hat{\theta}-\theta\right)+O_p\left(\frac{1}{T}\right)$$

so that we have

$$\sqrt{T}\left(\hat{\theta}-\theta\right) = \sqrt{T}\left[\psi_T-\psi(\theta)\right]G(\theta)^{-1}+O_p\left(\frac{1}{\sqrt{T}}\right)$$

where $G_T(\theta) = \frac{\partial \psi_T(\theta)}{\partial \theta'}$. Note that we know that

$$\sqrt{T}\left[\psi_T - \psi(\theta)\right] \to^d N(0, \Phi)$$

Hence we have

$$\sqrt{T}\left(\hat{\theta} - \theta\right) \to^d N\left(0, G(\theta)^{-1}\Phi G(\theta)'^{-1}\right)$$

where $G_T(\theta) \to^p G(\theta)$.

9.1 GMM

First consider infeasible generalized version of method of moments.

$$\arg\min_{\mu,\rho,\gamma_0} \left[\psi_T - \psi(\theta)\right]' \Phi^{-1} \left[\psi_T - \psi(\theta)\right].$$

where Φ is true unknown weighting matrix. Now feasible version becomes

$$\arg\min_{\mu,\rho,\gamma_0} \left[\psi_T - \psi(\theta)\right]' \mathbf{W}_T \left[\psi_T - \psi(\theta)\right] = \arg\min_{\mu,\rho,\gamma_0} G_T(\theta)' \mathbf{W}_T G_T(\theta)$$

where $\mathbf{W}_T$ is a consistent estimator of Φ^{-1}. Let

$$V_T = \left[\psi_T - \psi(\theta)\right]' \mathbf{W}_T \left[\psi_T - \psi(\theta)\right]$$

Then GMM estimator satisfies

$$\frac{\partial V_T\left(\hat{\theta}_{GMM}\right)}{\partial \hat{\theta}_{GMM}} = 2G_T\left(\hat{\theta}_{GMM}\right)' \mathbf{W}_T \left[\psi_T - \psi\left(\hat{\theta}_{GMM}\right)\right] = 0$$

so that we have

$$\psi\left(\hat{\theta}_{GMM}\right) = \psi_T(\theta) + G_T(\theta)\left(\hat{\theta}_{GMM} - \theta\right) + O_p\left(\frac{1}{T}\right)$$

Thus

$$\begin{aligned} & G_T\left(\hat{\theta}_{GMM}\right)' \mathbf{W}_T \left[\psi_T - \psi\left(\hat{\theta}_{GMM}\right)\right] \\ = \; & G_T\left(\hat{\theta}_{GMM}\right)' \mathbf{W}_T \left[\psi_T - \psi\left(\hat{\theta}_{GMM}\right)\right] + G_T\left(\hat{\theta}_{GMM}\right)' \mathbf{W}_T G_T(\theta)\left(\hat{\theta}_{GMM} - \theta\right) = 0 \end{aligned}$$

Hence

$$\left(\hat{\theta}_{GMM}-\theta\right)=-\left\{G_T\left(\hat{\theta}_{GMM}\right)'\mathbf{W}_T G_T\left(\theta\right)\right\}^{-1}G_T\left(\hat{\theta}_{GMM}\right)'\mathbf{W}_T\left[\psi_T-\psi\left(\hat{\theta}_{GMM}\right)\right]$$

and

$$\sqrt{T}\left(\hat{\theta}_{GMM}-\theta\right)\rightarrow^{d}N\left(0,V\right)$$

where

$$V=\frac{1}{T}\left\{G'\mathbf{W}G\right\}^{-1}G'\mathbf{W}\mathbf{\Phi}\mathbf{W}G\left\{G'\mathbf{W}G\right\}^{-1}.$$

When $W=\Phi^{-1}$, then we have

$$V=\frac{1}{T}\left\{G'\mathbf{\Phi}^{-1}G\right\}^{-1}G'\mathbf{\Phi}^{-1}G\left\{G'\mathbf{\Phi}^{-1}G\right\}^{-1}=\frac{1}{T}\left\{G'\mathbf{\Phi}^{-1}G\right\}^{-1}.$$

10 Panel Data Analysis (Chapter 9)

Latent Data Generating Process

$$y_{it} = \mu_{yi} + \lambda_{yt} + y_{it}^o$$

where

$$\begin{aligned} \mu_{yi} &= \text{ time invariant individual characteristics} \\ \lambda_{yt} &= \text{ cross sectional invariant common factor} \\ y_{it}^o &= \text{ idiosyncratic term} \end{aligned}$$

10.1 Economic, Financial, or Social Theory:

1. Time series approach: Long T but small N : Finance and macroeconomics

$$y_{it} = a_i + b_i x_{it} + u_{it}$$

   (a) Heterogeneity (coefficients, especially constant term) becomes an important issue.

   (b) Pooling regression coefficient b : Testing heterogeneity becomes an issue.

   (c) Cross section dependence becomes an issue

2. Cross sectional approach: Small T but large N : microeconomics, political science.

$$y_{it} = a + bx_{it} + e_{it}, \quad e_{it} = a_i + u_{it}$$

   (a) Heterogeneity (variance of e_{it} is correlated with x_{it}) becomes an issue

   (b) Use usually random effects model. Why?

   (c) Cross section dependence does not matter much.

10.2 Cross Section & Time Series Regressions

1. Cross Section Regression (applied microeconomics, typically labor, health, demography etc.)

(a) Usually try to explain the different averages: Examples; gender wage difference, race wage difference etc. Use survey data.

(b) Typical regression setting: Let y_i be the ith individual wage (or income) at a particular time (survey year)

$$y_i = a + b_1 \text{gender}_i + b_2 \text{region}_i + b_3 \text{age}_i + b_5 \text{edu}_i + ... + e_i$$

i. Explanatory variables: discrete variables. In other words, dummies.

ii. Nonlinear versus linear: Approximation around $\mathbf{x}_0$

$$\begin{aligned} y_i &= f(\mathbf{x}_i) \simeq f(\mathbf{x}_0) + \frac{\partial f}{\partial \mathbf{x}_{0i}} (\mathbf{x}_i - \mathbf{x}_0) + \frac{1}{2} \frac{\partial^2 f}{\partial \mathbf{x}_{0i}^2} (\mathbf{x}_i - \mathbf{x}_0)^2 + ... \\ &= a + b_1 x_{1i} + b_2 x_{2i} + c_1 x_{1i}^2 + c_2 x_{2i}^2 + c_3 x_{1i} x_{2i} + error: \text{ for two regressors case} \end{aligned}$$

Hence without including the second moments, the regression suffers from misspecification

(c) Don't run cross section regression to examine time series behavior

$$y_{it} = \alpha_{yi} + y_{it}^o, \quad x_{it} = \alpha_{xi} + x_{it}^o$$

Assume

$$\begin{aligned} \alpha_{yi} &= b\alpha_{xi} + e_i: \text{ mean relation} \\ y_{it}^o &= \gamma x_{it}^o + \varepsilon_{it}: \text{ time series relation} \end{aligned}$$

In fact,

$$b \neq \gamma$$

2. Time series regression (Finance, International Economics, Macroeconomic etc)

(a) Examining parities (PPP, UIP, CIP, Fisher Hypothesis, etc). Dynamic stability becomes the main issue.

(b) Cointegration among nonstationary variables becomes an issue.

(c) Ignore time invariant variables such as means.

11 Pooling Panel and Random Effects (Estimation: Micro Panel)

Model

$$y_{it} = a + bx_{it} + \varepsilon_{it}$$

1. Why pooling?

   (a) Economic theory must hold for all individuals.

   (b) More data: either more cross sectional or time series observations. Pooling means more 'efficient' and 'powerful' (will explain later)

2. Why not pooling?

   (a) Account for individual heterogeneity. So at least we have to allow some level heterogeneity such as

   $$y_{it} = a_i + bx_{it} + u_{it}$$

   (b) How to handle for a_i then? Either fixed or random effects.

   (c) What if a_i is observable? like gender, edu, age etc. You may want to include them. How?

11.1 Random Effects

Model:

$$y_{it} = a + bx_{it} + e_{it}, \quad e_{it} = a_i - a + u_{it} = \mu_i + u_{it}$$

Assumption:

A1 $\mathrm{E}(\mu_i x_{it}) = 0$ for all i

A2 $\mathrm{E}(\mu_i u_{it}) = 0$ for all i

Under A1 and A2, note that the pooled OLS becomes consistent but not efficient. The consistency (here we are assuming $N, T \to \infty$ or $N \to \infty$ for any T) requires that

$$\text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} x_{it} e_{it} = 0.$$

Indeed under A1 and A2, we can prove that POLS estimator satisfies the above condition. However, the regression errors are not i.i.d. anymore.

$$e_{i1} e_{i2} = \mu_i^2 + u_{i1} u_{i2} + \mu_i u_{i1} + \mu_i u_{i2}$$

Taking expectation yields

$$\begin{aligned} \mathrm{E} e_{i1} e_{i2} &= \mathrm{E}\mu_i^2 + \mathrm{E} u_{i1} u_{i2} + \mathrm{E}\mu_i u_{i1} + \mathrm{E}\mu_i u_{i2} \\ &= \sigma_\mu^2 \text{ if } \mathrm{E} u_{i1} u_{i2} = 0 \text{ (no serial corr.)} \end{aligned}$$

where we assume $\mathrm{E}(\mu_i u_{it}) = 0$. Also note that

$$\begin{aligned} \mathrm{E} e_{i1} e_{i1} &= \mathrm{E}\mu_i^2 + \mathrm{E} u_{i1} u_{i1} + 2\mathrm{E}\mu_i u_{i1} \\ &= \sigma_\mu^2 + \sigma_u^2 \end{aligned}$$

In this case, pooled GLS estimator becomes efficient and consistent. Here is how to obtain the feasible GLS estimator

1. Run

$$y_{it} = a + b x_{it} + e_{it}$$

and get the pooled OLS residuals $\hat{e}_{it}$. Let $\hat{b}_{\text{pols}}$ and $\hat{a}_{\text{pols}}$ be the POLS estimates for b and a.

2. Construct

$$\begin{aligned}
\hat{\sigma}_e^2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}-\hat{a}_{\text{pols}}-\hat{b}_{\text{pols}}x_{it}\right)^2 = \frac{1}{NT}\sum_{i=1}^{T}\sum_{t=1}^{T}\hat{e}_{it}^2 \\
\hat{\mu}_i &= \frac{1}{T}\sum_{t=1}^{T}\hat{e}_{it}, \quad \hat{u}_{it}=\hat{e}_{it}-\hat{\mu}_i \\
\hat{\sigma}_\mu^2 &= \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\mu}_i-\frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_i\right)^2 \\
\hat{\sigma}_u^2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{u}_{it}-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}\right)^2
\end{aligned}$$

Note if T is small, $(T-1)$ should be used for the above calculation.

3. Construct the sample covariance matrix

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_\mu^2+\hat{\sigma}_u^2 & \hat{\sigma}_\mu^2 & \cdots & \hat{\sigma}_\mu^2 \\ \hat{\sigma}_\mu^2 & \hat{\sigma}_\mu^2+\hat{\sigma}_u^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_\mu^2 & \hat{\sigma}_\mu^2 & \ldots & \hat{\sigma}_\mu^2+\hat{\sigma}_u^2 \end{bmatrix} \tag{22}$$

and then construct the feasible GLS estimator given by

$$\hat{b}_{\text{fgls}} = \left(\sum_{i=1}^{N}X_i'\hat{\Omega}^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N}X_i'\hat{\Omega}^{-1}Y_i\right)$$

where $X_i=[x_{1i},...,x_{Ti}]'$ and $Y_i=[y_{1i},...,y_{Ti}]'$

Remark 1: (Inconsistency relies on A1) If A1 does not hold (usually A1 does not hold), that is, if individual characteristics are correlated with regressors, then POLS estimator becomes inconsistent. Also the random effects estimator (FGLS) is also inconsistent. Because of this reason, many researchers in practice don't run the random effects model (or FGLS estimator). We will study the alternative estimation method in the below (fixed effects model).

Remark 2: (Including Observed Individual Effects) Even when A1 does not hold, if μ_i is observable, then the observed μ_i can be entered the regression as a regressor. That is,

$$y_{it} = a + \gamma_1 \mu_{1i} + \gamma_2 \mu_{2i} + ... + bx_{it} + u_{it}$$

We will study this model later (after studying fixed effects model) in detail.

12 Fixed Effects (Estimation: Micro Panel)

12.1 Eyeball Approach: Works well.

You need to draw some graphs (for your dissertation or journal article) why? looks good, and give more direct information. Try to draw one nice graph which explains main theme of the paper.

12.1.1 Single explanatory variables

Target: Want to explain the relationship between y_{it} and x_{it}. Plot y_{it} on x_{it}. Use **different** color for each i.

1. See if there is one unique relationship between y_{it} and x_{it} across i.
2. <insert a graph here> fixed effects (positive and positive)
3. <insert a graph here> fixed effects (positive but negative)
4. <insert a graph here> heterogeneity (positive but negative)
5. <insert a graph here> projected graph. (demean)

Demean:

$$
\begin{aligned}
y_{it} &= a_i + bx_{it} + u_{it} \\
\frac{1}{T}\sum_{t=1}^{T} y_{it} &= a_i + b\frac{1}{T}\sum_{t=1}^{T} x_{it} + \frac{1}{T}\sum_{t=1}^{T} u_{it} \\
y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} &= b\left(x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it}\right) + u_{it} - \frac{1}{T}\sum_{t=1}^{T} u_{it} \\
\tilde{y}_{it} &= b\tilde{x}_{it} + \tilde{u}_{it}
\end{aligned}
$$

12.1.2 More than two variables

$$y_{it} = a_i + bx_{it} + cz_{it} + u_{it}$$

1. Don't plot either $\tilde{y}_{it}$ on $\tilde{x}_{it}$ or $\tilde{y}_{it}$ on $\tilde{z}_{it}$: Why?

2. running $\tilde{y}_{it}$ on $\tilde{x}_{it}$ implies

$$\tilde{y}_{it} = b\tilde{x}_{it} + \tilde{e}_{it},\ \ \tilde{e}_{it} = c\tilde{z}_{it} + \tilde{u}_{it}$$

   If $E(\tilde{x}_{it}\tilde{z}_{it}) \neq 0$, then $\hat{b}$ becomes inconsistent. Worst case: $b = 0$ but $E(\tilde{x}_{it}\tilde{z}_{it}) \neq 0$, then $\hat{b} \neq 0$.

3. Solution: Run

$$\tilde{y}_{it} = a_1\tilde{z}_{it} + \tilde{y}_{it}^{+},\quad \tilde{x}_{it} = a_2\tilde{z}_{it} + \tilde{x}_{it}^{+}$$

   and get residuals $\tilde{y}_{it}^{+}$ and $\tilde{x}_{it}^{+}$. Plot them. Similarly, Run

$$\tilde{y}_{it} = b_1\tilde{x}_{it} + \tilde{y}_{it}^{*},\quad \tilde{z}_{it} = b_2\tilde{x}_{it} + \tilde{z}_{it}^{*}$$

   and plot $\tilde{y}_{it}^{*}$ on $\tilde{z}_{it}^{*}$

4. Mathematically, it is a projection approach. $I - Z(Z'Z)^{-1}Z' = M_z$ or M_x matrix.

12.2 Common Time Effects:

$$y_{it} = a_i + \lambda_t + bx_{it} + u_{it}$$

Allows time dummies also. How to estimate $\hat{b}$?

1. Eliminate fixed effects by demeaning over t.

$$y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} = \lambda_t - \frac{1}{T}\sum_{t=1}^{T}\lambda_t + b\left(x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it}\right) + u_{it} - \frac{1}{T}\sum_{t=1}^{T} u_{it}$$

   Still you have λ_t terms.

2. Rewrite this as

$$\tilde{y}_{it} = \tilde{\lambda}_t + b\tilde{x}_{it} + \tilde{u}_{it} \tag{23}$$

Take cross sectional mean

$$\frac{1}{N}\sum_{i=1}^{N}\tilde{y}_{it} = \tilde{\lambda}_t + b\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_{it} + \frac{1}{N}\sum_{i=1}^{N}\tilde{u}_{it} \tag{24}$$

3. subtract (24) from (23).

$$\tilde{y}_{it} - \frac{1}{N}\sum_{i=1}^{N}\tilde{y}_{it} = b\left(\tilde{x}_{it} - \frac{1}{N}\sum_{i=1}^{N}\tilde{x}_{it}\right) + \left(\tilde{u}_{it} - \frac{1}{N}\sum_{i=1}^{N}\tilde{u}_{it}\right)$$

4. Finally evaluate

$$y_{it}^{\dagger} = \tilde{y}_{it} - \frac{1}{N}\sum_{i=1}^{N}\tilde{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T}y_{it} - \frac{1}{N}\sum_{i=1}^{N}y_{it} + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}y_{it} \tag{25}$$

we call it 'within transformation'.

Note: Fixed effects estimator is called either 'Least Squares Dummies Variable (LSDV)' estimator or 'Within Group' estimator.

Questions Consider the following data generating process

$$y_{it} = \mu_{y,i} + y_{it}^{o}, \quad x_{it} = \mu_{x,i} + x_{it}^{o} \tag{26}$$

where

$$\mu_{y,i} = a + b\mu_{x,i} + \epsilon_i \tag{27}$$

$$y_{it}^{o} = \alpha_i + \beta x_{it}^{o} + u_{it}^{o} \tag{28}$$

1. Suppose that you run the following cross section regression for $t = 1$.

$$y_{i1} = c_1 + \gamma_1 x_{i1} + \varepsilon_{i1} \tag{29}$$

   Prove that the OLS estimate becomes inconsistent generally. That is,

$$\text{plim}_{N\to\infty}\hat{\gamma}_1 \neq b$$

2. Rather than running (29), you run the following cross sectional regression with time series average.

$$\bar{y}_i = c + \gamma\bar{x}_i + \bar{\varepsilon}_i \tag{30}$$

   where

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$$

   Derive the limiting distribution of $\hat{\gamma}$ in (30). Is the convergence rate equal to $\sqrt{NT}$ or $\sqrt{N}$?

Part II (POLS): Consider the following DGP

$$y_{it} = a_i + y_{it}^{o}, \quad y_{it}^{o} = \rho y_{it-1}^{o} + u_{it}, \quad u_{it} \sim iid\left(0, \sigma^2\right)$$

1. You run the POLS given by

$$y_{it} = a + \rho y_{it-1} + e_{it}$$

   Prove that when $\rho < 1$, the POLS estimator becomes inconsistent. Derive the exact bias.

Part III (Dynamic Panel Regression I) Consider the following DGP

$$y_{it} = a_i + y_{it}^o, \quad y_{it}^o = \rho y_{it-1}^o + u_{it}, \quad u_{it} \sim iid\left(0, \sigma^2\right)$$

Derive Nickell bias when $\rho = 1$

12.3 Dynamic Panel Regression

Read: Bertrand, M., E. Duflo and S. Mullainathan, 2004, How much should we trust differences-in-differences estimates?, *Quarterly Journal of Economics*, 249–275.

Model:

$$y_{it} = a_i + \lambda_t + bx_{it} + u_{it} \tag{31}$$

Now the regression error follows

$$u_{it} = \rho u_{it-1} + \varepsilon_{it}$$

Remark 1: As long as x_{it} is exogenous, the LSDV estimator in (31) becomes consistent. However, the statistical inference (in other words, $t-$value for $\hat{b}$) becomes an issue (in other words, the critical value for $\hat{t}_b$ must be different than the ordinary critical value). We will suggest the solution for the statistical inference later. (see section 3)

Remark 2: If T is large, then more efficient estimator can be obtain by running dynamic panel regression.

Let's transform (31) as

$$\rho y_{it-1} = a_i \rho + \rho\lambda_{t-1} + b\rho x_{it-1} + \rho u_{it-1} \tag{32}$$

and next subtract (32) from (31). Then we have

$$y_{it} = a_i (1 - \rho) + \rho\lambda_t + \rho\lambda_{t-1} + \rho y_{it-1} + bx_{it} - b\rho x_{it-1} + \varepsilon_{it}$$

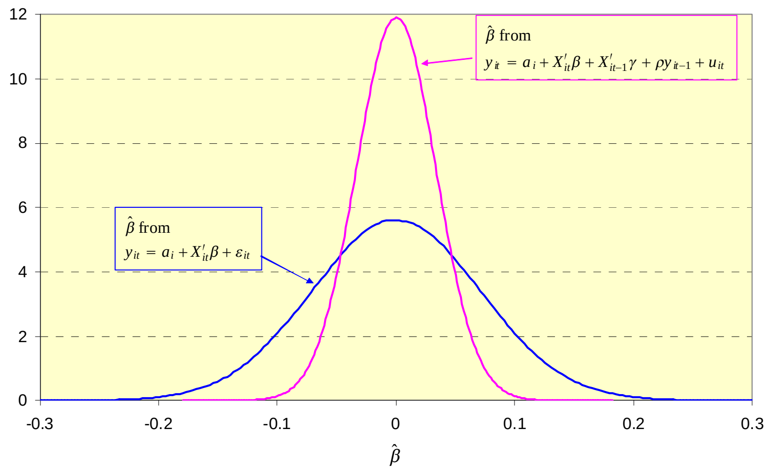
or

$$y_{it} = \alpha_i + \theta_t + \rho y_{it-1} + bx_{it} + \gamma x_{it-1} + \varepsilon_{it}. \tag{33}$$

By using within transformation, we can run

$$y_{it}^{\dagger} = \rho y_{it-1}^{\dagger} + bx_{it}^{\dagger} + \gamma x_{it-1}^{\dagger} + \varepsilon_{it}^{\dagger}$$

See (25) the definition of ‘†’.



Remark 3 (Consistency for ρ and γ): The LSDV estimators for ρ and γ are inconsistent but the LSDV estimator β becomes consistent. So the parameter of interest is here assumed to be β. Since the estimators for ρ and γ are inconsistent, the statistical inference for β should be carefully constructed. In the next section, we will study how to obtain robust statistical inference regardless of error term structures.

13 Pooling Panel and Random Effects (Testing: Micro Panel)

13.1 Bench Mark Model: Strongly Exogenous Single Regressor with Fixed Effects

Model

$$y_{it} = a_i + bx_{it} + u_{it} \tag{34}$$

Assumptions

1. $\mathrm{E}x_{it}u_{js} = 0$ for all i, j, s, t.

Here we are interested in testing the null hypothesis of $H_0 : b = 0$. To test this null hypothesis, we need a statistic. Usually we use a formal t statistic defined by

$$t_{\hat{b}} = \frac{\hat{b}}{\sqrt{Var\left(\hat{b}\right)}}$$

where $Var\left(\hat{b}\right)$ stands for the sample variance of the point estimate $\hat{b}$ which depends on the parametric assumptions for the regression errors.

In the below, we will study various hypotheses testings and statistics. Before that, I will address why the panel data is useful (and powerful) compared with either cross sectional or time series regressions.

13.1.1 More T or More N?

General statistical panel theory states that the panel gain comes from the use of more data. However, this statement is not quite right. One may have either a lengthy time series or cross section data. However whenever one uses a panel data, s/he can use either a short time series across some individuals, or a small individual over somewhat large time series data. For example, many empirical growth regressions have been based on cross sectional studies due to the data limitation. Even though PWT provides more than 150 countries panel data, it is often very hard to obtain a full set of panel data for all 150 countries. Here we consider which data sets (larger T or N) we should use to increase panel gain.

To attack this issue, we first consider the rate of convergence concept. Consider the following simple regression

$$y_s = bx_s + u_s, \text{ for } s = i \text{ or } t, \text{ and } s = 1,,,.S$$

where we assume the strong exogeneity of x_s. Typical limiting distribution theory says

$$\begin{aligned}
\hat{b} &= \frac{\frac{1}{S}\sum_{s=1}^{S} x_s y_s}{\frac{1}{S}\sum_{s=1}^{S} x_s^2} = b + \frac{\frac{1}{S}\sum_{s=1}^{S} x_s u_s}{\frac{1}{S}\sum_{s=1}^{S} x_s^2}, \\
\hat{b} - b &= \left(\frac{1}{\sqrt{S}}\right) \frac{\frac{1}{\sqrt{S}}\sum_{s=1}^{S} x_s u_s}{\frac{1}{S}\sum_{s=1}^{S} x_s^2} := \left(\frac{1}{\sqrt{S}}\right) \frac{A_S}{B_S}, \text{ let say}
\end{aligned}$$

We may assume that

$$A_S \Longrightarrow^d N\left(0, \Omega_A^2\right), \quad B_S \longrightarrow^p Q_B \text{ as } S \to \infty$$

where '$\Longrightarrow^d$' stands for convergence in distribution and '$\longrightarrow^p$' means convergence in probability. Then we finally have (following by Cramer's theorem)

$$\sqrt{S}\left(\hat{b} - b\right) \Longrightarrow^d N\left(0, Q_B^{-1}\Omega_A^2 Q_B^{-1}\right)$$

Alternatively

$$\frac{\sqrt{S}\left(\hat{b} - b\right)}{\sqrt{Q_B^{-1}\Omega_A^2 Q_B^{-1}}} \Longrightarrow^d N\left(0, 1\right)$$

Meanwhile the testing hypothesis is given by

$$\begin{aligned}
H_0 &: b_s = 0, \text{ usually.} \\
H_A &: b_s \neq 0
\end{aligned}$$

Then we have

$$\frac{\sqrt{S}\hat{b}}{\sqrt{Q_B^{-1}\Omega_A^2 Q_B^{-1}}} \Longrightarrow^d N\left(\frac{\sqrt{S}b}{\sqrt{Q_B^{-1}\Omega_A^2 Q_B^{-1}}}, 1\right)$$

so that the power of the test (how frequently a test can reject the null hypothesis when the alternative is true) is getting larger if

1. true value of $|b|$ is getting larger,
2. Variance of b is getting smaller,
3. the number of observations, S, is getting larger.

Among them, the last item, 3, is only thing we can control for. We don't know the true value of b and the true variance of b either. However, we can increase the number of observations (by putting more labor hours for digging out the data).

Now, when we have both N and T dimensions, we can rewrite the pooled estimate of b as

$$\hat{b}_{\text{panel}} = \frac{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}x_{it}y_{it}}{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}x_{it}^2} = b + \frac{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}x_{it}u_{it}}{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}x_{it}^2},$$

and similarly

$$\hat{b}_{\text{panel}} - b = \left(\frac{1}{\sqrt{NT}}\right)\frac{A_{NT}}{B_{NT}}, \text{ let say}$$

and

$$A_{NT} \Longrightarrow^{d} N\left(0, \Omega_A^2\right), \quad B_{NT} \longrightarrow^{p} Q_B \text{ as } N, T \to \infty \text{ jointly.} \tag{35}$$

Then we have

$$\sqrt{NT}\left(\hat{b}_{\text{panel}} - b\right) \Longrightarrow^{d} N\left(0, Q_B^{-1}\Omega_A^2 Q_B^{-1}\right) \tag{36}$$

Now consider the above three criteria for the power of the test. Does panel data enable us to know either true value of variance of b? The answer is no. Then what about the last one? Does panel data enable us to use more observations? The answer is not straightforward. In practice, one often face the situation like this. When one use one dimensional data (for example time series), one may choose or select the longest time series data for y_t and x_t. Denote the size of the sample as T_s. Now if s/he has to use a panel data, usually s/he scarifies the lengthy time series in order to increase the cross section units. Denote the time series s/he will use for the panel data as T. From the direct calculation, we have the condition for the panel gain given by

$$N > T_s/T$$

That is, if you have 300 of T_s for one series but have to use only 30 of T in order to use the panel data, then the minimum number of the cross sections – you have to obtain – should be larger than 10.

However, we may need much larger cross sections if y_s and x_s are $I(1)$ (or in other words, nonstationary). In this case, the limiting distribution for $\hat{b}$ is different from a normal distribution (actually it becomes $\left(\int B_x du\right)\left(\int B_x^2 dr\right)^{-1}$) and also the convergence rate becomes T rather than $\sqrt{T}$. Hence the minimum condition for the panel gain changes as

$$\sqrt{N} > T_s/T.$$

In the above example, you need at least $\sqrt{N} > 10$ or $N > 100$.

Unfortunately, the most of macro data are nonstationary. So the important question becomes that how many observations should be scarified to use the panel data. Let k be the fraction of the sample you have to sacrifice to use additional N cross sections. Then we have

$$\sqrt{N} > \frac{T_s}{T}, \text{ or } \sqrt{N} > \frac{T_s}{(1-k)\,T_s} = \frac{1}{1-k},$$

so that

$$N > \left(\frac{1}{1-k}\right)^2.$$

To decode this formula, let say you have 120 monthly time series observations initially. In order to use the panel data, if you have to use 10 annual observations, then $T_s/T = 120/10 = 12$, so that the minimum N becomes 144. Remember that the power of a test with $N = 144$ and $T = 10$ will be exactly same as the power of the test with $T = 120$ and $N = 1$. However, if you can still use monthly observations but loose 2 years observations, then $T_s/T = 120/96 = 1.25$, so that the minimum N becomes 1.56 which is less than 2. Hence the power of a test with $N = 2$ and $T = 96$ will be larger than that with $N = 1$ and $T = 120$.

So the conclusion follows:

Recommendation (How to Construct a Panel Data)

1. When you are interested in the correlation among level variables, you should use the panel data set which contains more T, or the largest of $N \times T^2$ rather than $N \times T$.

2. When you are interested in the correlation among (quasi) difference variables (such as growth rates), you should use the panel data which total number of observations $(= N \times T)$ is largest.

13.1.2 How to Calculate the Covariance Matrix

Here we are asking how to estimate Ω_A^2 and Q_B in (35) and (36). First consider Ω_A^2 which can be defined as

$$\begin{aligned} \Omega_A^2 &= \frac{1}{NT} E \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} u_{it} \right)^2 \\ &= \frac{1}{NT} E \left(x_{11} u_{11} + ... + x_{1T} u_{1T} + x_{21} u_{21} + ... + x_{NT} u_{NT} \right)^2 \\ &= \frac{1}{NT} E \left(x_{11}^2 u_{11}^2 + ... + x_{1T}^2 u_{1T}^2 + x_{21}^2 u_{21}^2 + ... + x_{NT}^2 u_{NT}^2 \right) \\ & \quad + E \, (\text{cross products}) \end{aligned} \tag{37}$$

If $E(u_{i1} u_{i2}) \neq 0$ due to serial correlation, then in general the expected values of the cross product terms are not equal to zero.

White (1980) suggests the use of the so called 'heteroskedasticity consistent estimator' which is given by

$$\hat{\Omega}_A^2 = \frac{1}{N} \sum_{i=1}^{N} \tilde{X}_i' \hat{u}_i \hat{u}_i' \tilde{X}_i$$

where $\hat{u}_i = (\hat{u}_{i1}, ..., \hat{u}_{iT})'$, $\tilde{X}_i = (\tilde{x}_{i1}, ..., \tilde{x}_{iT})'$. So that the sample covariance matrix becomes

$$V\left(\hat{b}\right) = \left(\sum_{i=1}^{N} \tilde{X}_i' \tilde{X} \right)^{-1} \left(\sum_{i=1}^{N} \tilde{X}_i' \hat{u}_i \hat{u}_i' \tilde{X}_i \right) \left(\sum_{i=1}^{N} \tilde{X}_i' \tilde{X} \right)^{-1} \tag{38}$$

and its associated t-statistic becomes

$$t_{\hat{b}} = \frac{\hat{b}}{\sqrt{\left(\sum_{i=1}^{N} \tilde{X}_i' \tilde{X} \right)^{-1} \left(\sum_{i=1}^{N} \tilde{X}_i' \hat{u}_i \hat{u}_i' \tilde{X}_i \right) \left(\sum_{i=1}^{N} \tilde{X}_i' \tilde{X} \right)^{-1}}} \tag{39}$$

Note that if x_{it} and u_{it} are *iid*, then the above formula can be simplified as

$$V\left(\hat{b}\right) = \hat{\sigma}_u^2 \left(\sum_{i=1}^{N} \tilde{X}_i' \tilde{X} \right)^{-1} \tag{40}$$

which is the sample variance reported in canned statistical packages.

Here are a couple of very important facts:

Recommendation

1. Usually the sample variance in (2-5) is larger than that in (40). This implies that when there is either heteroskedasticity or autocorrelation, the standard t-ratio is much larger than its true value.

2. When T is fixed but N is large, the $t_{\hat{b}}$ in (39) is distributed as a normal. So the standard critical value can be used here. However when T is large but N is small, the t ratio asymptotically follows a $t-$distribution with $N-1$ degrees of freedom under homoskedasticity. (Hansen, 2007 Journal of Econometrics, 'Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when T is large')

13.2 Testing

13.2.1 Some Basic Facts on Statistical Testing

Size and Power The size of a test stands for the rejection rate of the null when the null hypothesis is true, meanwhile the power of a test implies the rejection rate of the null when the alternative is true. Usually we set the size of a test at the significance level. For example, the critical value for the 5% significance level for a normal distribution (for two sides test) is 1.95. In other words, we permit ourselves that we would make a wrong decision at the 5% level. (5 out of 100 times). Setting a smaller size means that you want to be more conservative or don't want to make any mistake, but at the same time it also implies that the power of the test will be reduced.

Size Distortion You set the size at the 5% significance level. However (especially in the finite sample), a test does not produce exactly the 5% of the size. If a test over-rejects the null (when the null is true), then we say that the test suffers from oversize distortion. The opposite case is undersize distortion. Usually the undersized test is acceptable since it simply implies that you will make less mistake. The oversize problem becomes serious. The oversized test usually rejects the null very often even when the null is true.

Size Problem in Panel Data In univariate case, usually a well designed statistic does not suffer from the size distortion as n (the number of observations) goes to infinity. For example, the standard t-test for the univariate AR(1) regression produces somewhat serious size distortion with small T, but as $T \to \infty$, the size distortion goes away.

$$y_t = a + \rho y_{t-1} + u_t, \quad t_{\hat{\rho}} = \frac{\hat{\rho}}{\sqrt{V(\hat{\rho})}} \text{ for } \rho < 1 \tag{41}$$

It is because the asymptotic variance of $\hat{\rho}$ is designed in this way. However, in the panel data, the t-ratio produces more size distortion as $N \to \infty$ for fixed T.

$$y_{it} = a_i + \rho y_{it-1} + u_{it}, \quad t_{\hat{\rho}} = \frac{\hat{\rho}_{\text{lsdv}}}{\sqrt{V(\hat{\rho}_{\text{lsdv}})}} \text{ for } \rho < 1 \tag{42}$$

The underlying reason is simple. When T is small, the test statistic in (41) produces a small size distortion. In the panel data, the size distortion becomes cumulated as N increases. Similarly, as $T \to \infty$ for a fixed N, the usual panel statistic in (39) produces more size distortion if there is heteroskedasticity in the error terms.

13.2.2 Fixed versus Random Effects.

LSDV estimator is 'robust' and consistent whether or not the fixed effects a_i in (34) are correlated with x_{it}. Meanwhile the GLS (or random effects estimator) is 'efficient' and consistent only when a_i is not correlated with regressors. When the number of observations are small (such as moderately small N and T), the GLS becomes an attractive estimator if a_i is not correlated with regressors. Naturally econometricians have developed various test statistics to investigate if this condition holds or not.

There are broadly two ways to test the orthogonality between a_i and x_{it}. The first method is based on the pooled OLS regression residuals, and the second method is based on the difference between LSDV and GLS. We discuss the first method, first.

Breusch & Pagan (1980)'s LM Test BP tests if

$$\begin{aligned} H_0 &: a_i = a, \text{ for all } i, \\ H_A &: a_i \neq a \text{ for any } i \end{aligned} \tag{43}$$

When u_{it} in (34) is not serially correlated, these hypotheses can be rewritten as

$$\begin{aligned} H_0 &: E\left(\hat{e}_{it}\hat{e}_{is}\right) = 0 \text{ for all } i, \\ H_A &: E\left(\hat{e}_{it}\hat{e}_{is}\right) \neq 0 \text{ for any } i \end{aligned}$$

where $\hat{e}_{it}$ is the pooled OLS regression residuals. That is,

$$\hat{e}_{it} = y_{it} - \hat{a} - \hat{b}_{\text{pols}} x_{it}.$$

The test statistic is given by

$$LM = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^{N} \left(\sum_{t=1}^{T} \hat{e}_{it} \right)^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{e}_{it}^2} - 1 \right]^2 \Longrightarrow^d \chi_1^2$$

Note that

$$\mathrm{E}\left(\sum_{t=1}^{T} \hat{e}_{it}\right)^2 = \mathrm{E}\left(\sum_{t=1}^{T} \hat{e}_{it}^2 + \sum_{t=1}^{T}\sum_{s\neq t}^{T} e_{it}e_{is}\right)$$

and under H_0, we have

$$\mathrm{E}\left(\sum_{t=1}^{T} \hat{e}_{it}\right)^2 = \mathrm{E}\left(\sum_{t=1}^{T} \hat{e}_{it}^2 + \sum_{t=1}^{T}\sum_{s\neq t}^{T} e_{it}e_{is}\right) = \mathrm{E}\left(\sum_{t=1}^{T} \hat{e}_{it}^2\right)$$

since the expectation of the cross product terms become zero. For large T and N, also note that under the alternative and no serial correlation among u_{it}, we have

$$\mathrm{E}\left(\sum_{t=1}^{T} \hat{e}_{it}\right)^2 \geq \mathrm{E}\left(\sum_{t=1}^{T} \hat{e}_{it}^2\right)$$

since

$$\mathrm{E}\left(e_{it}e_{is}\right) = \mathrm{E}\left(\mu_i^2 + u_{it}u_{is}\right) = \sigma_\mu^2 > 0.$$

It is important to note that if u_{it} is serially correlated, then BP's LM test fails.

Hausman's Specification Test Hausman test is fairly a general test for misspecification, and can be applied to test the null hypothesis in (43). Under the null hypothesis

$$\text{plim}_{N,T\to\infty} \hat{b}_{\text{LSDV}} = \text{plim}_{N,T\to\infty} \hat{b}_{\text{GLS}}$$

since two estimators are both consistent. However, under the alternative, we have

$$\text{plim}_{N,T\to\infty}\hat{b}_{\text{LSDV}} = b \text{ but } \text{plim}_{N,T\to\infty}\hat{b}_{\text{GLS}} \neq b$$

so that

$$\text{plim}_{N,T\to\infty}\left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right) \neq 0$$

Hence we can test H_0 by examining if the distance between $\hat{b}_{\text{GLS}}$ and $\hat{b}_{\text{LSDV}}$ is equal to zero or not. A typical test statistic in this case is given by

$$\left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right)' \left[Var\left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right)\right]^{-1} \left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right)' \Longrightarrow^d \chi_k^2$$

when the dimension of $\hat{b}$ is k. For a single regressor case, we have simply

$$\frac{\left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right)^2}{Var\left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right)} \Longrightarrow^d \chi_1^2$$

Note that under H_0,

$$Var\left(\hat{b}_{\text{GLS}} - \hat{b}_{\text{LSDV}}\right) = \hat{\sigma}_u^2 \left[\sum_{i=1}^N \sum_{t=1}^T \tilde{x}_{it}^2\right]^{-1} - \left[\sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \hat{\omega}_{ij}\tilde{x}_{it}\tilde{x}_{jt}\right]^{-1}$$

where $\hat{\omega}_{ij}$ is the ith and jth element of $\hat{\Omega}^{-1}$ and $\hat{\Omega}$ is defined at (22).

14 Dynamic Panel Regression I (Issues and Problems)

More than the quarter of theoretical studies on the panel data is focused on the dynamic panel regression. Modelling the 'dynamics' in the panel data is critically important. First we address where 'dynamic adjustment form' comes from.

14.1 Source of Serial Correlation

14.1.1 Univariate Series

Many economic variables such as income, consumption, wage, etc have the following transitional path.

$$y_{it} = y_i^* + (y_{i0} - y_i^*) e^{-\beta t}$$

where y_i^* is the steady state outcome. Note that all variables are in logarithm. Rewrite this model as

$$y_{it} = \left[y_i^* + (y_{i0} - y_i^*) e^{-\beta(t-1)}\right] e^{-\beta} + y_i^* \left(1 - e^{-\beta}\right) = y_i^* \left(1 - e^{-\beta}\right) + e^{-\beta} y_{it-1}$$

By letting $\rho = e^{-\beta}$, $a_i = y_i^*$ and adding a random error which can be exogeneous i.i.d. measurement errors, transitory shocks, etc, then we have

$$y_{it} = a_i (1 - \rho) + \rho y_{it-1} + u_{it}, \text{ for } t = 1, ..., T.$$

This simple growth model generates the time dependence between y_{it} and y_{it-1}.

In other words, all variables (growing variables such as wage, income, height etc) are serially correlated during transition periods.

14.1.2 General Regressions

In general regression models, the serial correlation occurs whenever the regression is not balanced. To understand the balancing concept, consider a simple regression model given by

$$y_{it} = \alpha_i + bx_{it} + u_{it} \tag{44}$$

Suppose that

1. y_{it} has a linear trend. $b \neq 0$. Regardless x_{it} has a linear trend or not, u_{it} contains a linear trend. So you have to include a trend in the regression. Why?

   (a) Let $y_{it} = a_{iy} + c_i t + y_{it}^o$, and $x_{it} = a_{ix} + d_i t + x_{it}^o$. You don't want to assume that the deterministic trend terms have a common relationshop since you can't write it as

   $$c_i t = a + b\,(d_i t) + e_{it}. \tag{45}$$

   Simply because the dependent variable is purely nonstochastic. Even when the dependent variable has a stochastic component (such as $\zeta_{it} = c_i t + \epsilon_{it}$, and $\zeta_{it} = a + b\,(d_i t) + e_{it}$, as long as $c_i \neq bd_i$ for any i, the error term includes a linear trend component.

   (b) The interest relation must be between y_{it}^o and x_{it}^o. In this case, you have to eliminate the trend term in the first place by including a linear trend component in the regression

   (c) If you are interested in analyzing growth rates in y_{it} and x_{it}, then you have to take the first difference to approximate the stochastic growth components. That is,

   $$\Delta y_{it} = \alpha_i + b\Delta x_{it} + \text{error}_{it} \tag{46}$$

2. y_{it} is serially correlated but x_{it} is not. Then u_{it} is serially correlated. (the opposite is not true) In this case, you may want to run the dynamic panel regression

   $$y_{it} = a_i + \rho y_{it-1} + bx_{it} + \gamma x_{it-1} + \varepsilon_{it} \tag{47}$$

   (a) From (44), you have

   $$u_{it} = \rho u_{it-1} + \varepsilon_{it}. \tag{48}$$

   Here I assume that the error term follows AR(1) structure for simplicity.

   (b) Then you have

   $$\rho y_{it-1} = \alpha_i \rho + b\rho x_{it-1} + \rho u_{it-1}. \tag{49}$$

   Subtracting (49) from (44) yields (47).

3. y_{it} is not serially correlated but x_{it} is very persistent (ρ is near unity). And more importantly $b \neq 0$. Then the regression in (44) is not well specified. Simply it becomes unbalanced regression. In this case, to balance out the serial correlation, u_{it} should be negatively correlated with x_{it}.

   (a) Example: Stock return predictability & UIP:

   $$y_{it} = a_i + bx_{it} + u_{it}$$

   where y_{it} is either stock return or depreciation rates, which are almost white noisy. x_{it} is either interest rate differential (for UIP), or dividend ratio (stock return). Both interest rate differential or dividend ratio is highly serially correlated. If $b \neq 0$, then x_{it} should be negatively correlated with u_{it}.

   (b) Hence u_{it} is serially correlated in this case also.

14.2 Modeling Dynamic Panel Regression

There are several types of dynamic panel regressions. Depending on the regression types, the properties of LSDV estimators are quite different. Hence modeling dynamic panel regression becomes very important.

$$\text{M1:} \qquad y_{it} = a_i + \beta x_{it} + u_{it}, \quad u_{it} = \rho u_{it-1} + \varepsilon_{it} \tag{50}$$

$$\text{M2:} \qquad y_{it} = a_i + \rho y_{it-1} + \beta x_{it} + \varepsilon_{it} \tag{51}$$

where I didn't include common time effects and linear trend components either. Note that M1 and M2 can be restated as

$$\text{M1:} \qquad z_{it} = \alpha_i + u_{it}, \quad z_{it} = y_{it} - \beta x_{it}, \quad u_{it} = \rho u_{it-1} + \varepsilon_{it} \tag{52}$$

$$\text{M2:} \qquad y_{it} = \alpha_i + u_{it}, \quad u_{it} = \rho u_{it-1} + e_{it}, \quad e_{it} = \beta x_{it} + \varepsilon_{it} \tag{53}$$

Note that in M1, x_{it} is correlated with y_{it} in level. Meanwhile in M2, x_{it} is correlated with the quasi-differenced y_{it}. Alternatively we can rewrite M1 as

$$\text{M1: } y_{it} = a_i + \rho y_{it-1} + \beta x_{it} + \gamma x_{it-1} + \varepsilon_{it}. \tag{54}$$

Hence if (51) is true, then (54) is not misspecified. Simply γ becomes zero if (51) is true. However, if (54) or M1 is true, then (51) becomes misspecified, which results in inconsistent estimator for β as well as ρ in (51). In this sense, (54) nests (51).

The economic interpretations are different across models. M1 states that the quasi-difference $(y_{it} - \rho y_{it-1})$ is explained by x_{it}. Meanwhile M2 implies that the level of y_{it} is explained by x_{it}. Hence usually x_{it} in (51) is assumed to follow a white noisy process (no serial correlation). Meanwhile x_{it} in (54) does not have such restriction.

14.3 Inconsistency of LSDV estimator

Here we analyze why the LSDV estimator under fixed effects becomes inconsistent as $N \to \infty$ but fixed T. The model we study is given by

$$y_{it} = a_i + \rho y_{it-1} + u_{it}, \quad u_{it} \sim iid\left(0, \sigma_u^2\right)$$

Nickell Bias (1981, Econometrica) Nickell extends the so-called 'Kendall' (1954, Biometrika) bias to the panel data setting.

1. To understand Kendall bias, we consider an univariate simple AR(1) model with constant

$$y_t = a + \rho y_{t-1} + u_t. \tag{55}$$

   The OLS estimator is given by

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \tilde{y}_{t-1}\tilde{y}_t}{\sum_{t=2}^{T} \tilde{y}_{t-1}^2},$$

   and its expectation gives

$$E\hat{\rho} = E\left[\frac{\sum_{t=2}^{T} \tilde{y}_{t-1}\tilde{u}_t}{\sum_{t=2}^{T} \tilde{y}_{t-1}^2}\right] := E\frac{A_T}{B_T}$$

   From Marriott and Pope (1954, Biometrika), we have

$$E\frac{A_T}{B_T} = \frac{EA_T}{EB_T}\left[1 - E\left(C_T\right)\right]$$

$$E\left(C_T\right)=\frac{Cov\left(A_T B_T\right)}{E\left(A_T\right)E\left(B_T\right)}+\frac{Var\left(B_T\right)}{\left[E\left(B_T\right)\right]^2}$$

Note that $E\left(C_T\right)\neq 0$ usually due to asymetric distribution of $\hat{\rho}$. In the finite sample, the empirical distribution of $\hat{\rho}$ is not a normal but skewed left a little bit. This asymetric distribution yields the small sample bias but usually it goes away quickly as T increases

2. The major bias arises from the first term EA_T/EB_T. To see this

$$\frac{EA_T}{EB_T}=\rho+\frac{E\sum_{t=2}^{T}\tilde{y}_{t-1}\tilde{u}_t}{E\sum_{t=2}^{T}\tilde{y}_{t-1}^2}$$

Note that

$$\begin{aligned}
E\sum_{t=2}^{T}\tilde{y}_{t-1}\tilde{u}_t &= E\sum_{t=2}^{T}\left(y_{t-1}-\bar{y}\right)\left(u_t-\bar{u}\right)=E\sum_{t=2}^{T}y_{t-1}u_t-\frac{1}{T}E\left(\sum_{t=2}^{T}y_{t-1}\right)\left(\sum_{t=2}^{T}u_t\right)\\
&= 0-\frac{1}{T}E\left(\sum_{t=2}^{T}y_{t-1}\right)\left(\sum_{t=2}^{T}u_t\right)
\end{aligned}$$

Since

$$y_t=a+\rho y_{t-1}+u_t=\frac{a}{1-\rho}+\sum_{j=0}^{\infty}\rho^j u_{t-j}$$

so that $Ey_tu_s=0$ for all $t<s$. However

$$E\left(\sum_{t=2}^{T}y_{t-1}\right)\left(\sum_{t=2}^{T}u_t\right)=E\left(y_1+...+y_{T-1}\right)\left(u_2+...+u_T\right)$$

and note that

$$Ey_1u_1=\sigma_u^2,\quad Ey_2u_1=E\left(\rho y_1+u_2\right)u_1=\rho\sigma_u^2,...$$

hence this term is not eqaul to zero.

3. Finally we have

$$E\hat{\rho}=E\frac{A_T}{B_T}=\rho-b_1\left(T\right)-b_2\left(T\right)$$

where

$$\begin{aligned}
b_1\left(T\right) &= \frac{E\sum_{t=2}^{T}\tilde{y}_{t-1}\tilde{u}_t}{E\sum_{t=2}^{T}\tilde{y}_{t-1}^2}=-\frac{1+\rho}{T}+O\left(T^{-2}\right)\\
b_2\left(T\right) &= -\frac{2\rho}{T}+O\left(T^{-2}\right)
\end{aligned}$$

It is important to know that the first bias, $b_1(T)$, comes from the correlation between $\tilde{y}_{t-1}$ and $\tilde{u}_t$ (which are the regressor and the regression error after demeaning transformation), and the second bias, $b_2(T)$, comes from the asymmetric distribution of $\hat{\rho}$.

4. In panel regressions, this first part of the small time series bias remains permanently when $N \to \infty$. However the second part of the small bias goes away. The underlying reason is straightforward. As $N \to \infty$, the distribution of $\hat{\rho}_{\text{LSDV}}$ becomes symmetric. Hence the bias arised from asymmetric distribution goes away simply. However the first bias $b_1(T)$ does not go away since this bias is arised because of the time series correlation between the regressor, $\tilde{y}_{t-1}$, and the regression error, $\tilde{u}_t$. More formally, we states

$$\begin{aligned}
\operatorname{plim}_{N\to\infty}(\hat{\rho}_{\text{LSDV}} - \rho) &= \operatorname{plim}_{N\to\infty} \frac{\frac{1}{N}\sum_{i=1}^N \sum_{t=2}^T \tilde{y}_{it-1}\tilde{u}_{it}}{\frac{1}{N}\sum_{i=1}^N \sum_{t=2}^T \tilde{y}_{it-1}^2} \\
&= \frac{\operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \sum_{t=2}^T \tilde{y}_{it-1}\tilde{u}_{it}}{\operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \sum_{t=2}^T \tilde{y}_{it-1}^2} \\
&= \frac{E\frac{1}{N}\sum_{i=1}^N \sum_{t=2}^T \tilde{y}_{it-1}\tilde{u}_{it}}{E\frac{1}{N}\sum_{i=1}^N \sum_{t=2}^T \tilde{y}_{it-1}^2} \\
&= -\frac{1+\rho}{T} + O\left(T^{-2}\right)
\end{aligned}$$

Asymptotic Bias when $\rho = 1$ Nickell (1981) shows the asymptotic bias (or inconsistency of $\hat{\rho}_{\text{LSDV}}$) when $\rho < 1$. Here we study how the expression of the bias formula badly fails when $\rho = 1$.

1. Consider the following latent model

$$y_t = \alpha + y_t^o, \quad y_t^o = \rho y_{t-1}^o + u_t$$

then we have

$$y_t = \alpha(1-\rho) + \rho y_{t-1} + u_t$$

so that, if $\rho = 1$, then

$$y_t = y_{t-1} + u_t = \sum_{s=1}^t u_s = u_1 + \ldots + u_t$$

2. In the panel data, we have

$$Ey_{it}^2 = E\left(u_{i1} + ... + u_{it}\right)^2 = t\sigma_u^2 \text{ for } Eu_{it}^2 = \sigma_u^2 \text{ for all } i.$$

$$E\frac{1}{T-1}\sum_{t=2}^{T} y_{it-1}^2 = \frac{1}{T-1}\sum_{t=2}^{T} E\left(u_{i1} + ... + u_{it}\right)^2 = \sigma_u^2 \frac{1}{T-1}\sum_{t=1}^{T-1} t = \sigma_u^2 \frac{T}{2}.$$

3. Prove that

$$E\left(\hat{\rho}_{\text{LSDV}} - 1\right) = -\frac{3}{T} + O\left(T^{-2}\right) < -\frac{2}{T} + O\left(T^{-2}\right)$$

14.4 Inconsistency of the Pooled OLS Estimator

Derive the inconsistency of the pooled OLS estimator

$$E\left(\hat{\rho}_{\text{POLS}}\right) = ?$$

1. We are running

$$y_{it} = a + \rho y_{it-1} + e_{it}, \quad e_{it} = a_i - a + u_{it}$$

2. The POLS estimator is given by

$$\hat{\rho}_{\text{POLS}} = \rho + \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it-1}\right)\left(e_{it-1} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} e_{it-1}\right)}{\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it-1}\right)^2}$$

Note that

$$E\left(\left[\alpha_i - \alpha\right] + y_{it-1}^o\right)\left(\left[\alpha_i - \alpha\right](1-\rho) + u_{it}\right) = \sigma_\alpha^2$$

and

$$E\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(y_{it-1} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it-1}\right)^2 = \sigma_\alpha^2 + \sigma_y^2$$

3. Let $\eta = \sigma_\alpha^2/\sigma_u^2$. And express the inconsistency in terms of η.

14.5 Asymptotic Distribution of LSDV estimator

$$\begin{aligned}
\hat{\rho}_{\text{LSDV}} - \rho &= \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}\tilde{u}_{it}}{\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^{2}} \\
&= -\frac{\sum_{i=1}^{N}\left(\sum_{t=2}^{T}y_{it-1}\right)\left(\sum_{t=2}^{T}u_{it}\right)}{\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^{2}} + \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-1}u_{it}}{\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^{2}}
\end{aligned}$$

$$\frac{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-1}u_{it}}{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^{2}} \Longrightarrow^{d} N\left(0, 1-\rho^{2}\right)$$

since

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-1}u_{it} \Longrightarrow^{d} N\left(0, \frac{\sigma_u^4}{1-\rho^2}\right)$$

Now we have

$$\begin{aligned}
\sqrt{NT}\left(\hat{\rho}_{\text{LSDV}} - \rho\right) &= -\frac{1+\rho}{T}\sqrt{NT} + \frac{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-1}u_{it}}{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^{2}} \\
&= -\left(1+\rho\right)\sqrt{\frac{N}{T}} + \frac{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}y_{it-1}u_{it}}{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\tilde{y}_{it-1}^{2}}
\end{aligned}$$

If $\frac{N}{T} \to c$ as $N, T \to \infty$,

$$\sqrt{NT}\left(\hat{\rho}_{\text{LSDV}} - \rho\right) \Longrightarrow^{d} -\left(1+\rho\right)c + N\left(0, 1-\rho^{2}\right)$$

If $\frac{N}{T} \to \infty$ as $N, T \to \infty$,

$$\sqrt{NT}\left(\hat{\rho}_{\text{LSDV}} - \rho\right) \to^{p} \infty$$

If $\frac{N}{T} \to 0$ as $N, T \to \infty$, then

$$\sqrt{NT}\left(\hat{\rho}_{\text{LSDV}} - \rho\right) \Longrightarrow^{d} N\left(0, 1-\rho^{2}\right)$$

Empirical Example Nominal wage $= y_{it}$, $S_i =$ treatment variable or dummy

$$y_{it} = \alpha + \beta S_i + u_{it}, \quad u_{it} = \rho u_{it-1} + \varepsilon_{it}, \quad \varepsilon_{it} \sim iidN\left(0, \sigma^2\right)$$

where S_i is a dummy variable. Suppose that u_{it} is serially correlated ($\rho \neq 0$).

Q1: Find the limiting distribution of $\hat{\beta}$ and $\hat{\alpha}$.

First transform the regression as

$$\begin{aligned} y_{it} - \frac{1}{N}\sum_{i=1}^{N} y_{it} &= \beta\left(S_i - \frac{1}{N}\sum_{i=1}^{N} S_i\right) + \left(u_{it} - \frac{1}{N}\sum_{i=1}^{N} u_{it}\right) \\ \tilde{y}_{it} &= \beta \tilde{S}_i + \tilde{u}_{it}, \text{ let say} \end{aligned}$$

Then

$$\hat{\beta} = \frac{\sum_{i=1}^{N} \tilde{S}_i \left(\sum_{t=1}^{T} \tilde{y}_{it}\right)}{\sum_{i=1}^{N} \tilde{S}_i^2} = \beta + \frac{\sum_{i=1}^{N} \tilde{S}_i \left(\sum_{t=1}^{T} \tilde{u}_{it}\right)}{\sum_{i=1}^{N} \tilde{S}_i^2}$$

Let

$$\hat{\beta} - \beta = \frac{\frac{1}{N}\sum_{i=1}^{N} \tilde{S}_i \left(\sum_{t=1}^{T} \tilde{u}_{it}\right)}{\frac{1}{N}\sum_{i=1}^{N} \tilde{S}_i^2}.$$

Assume that

$$S_i = \begin{cases} 0 \text{ if } i \in G_1 \text{ or } i = 1, ..., \frac{N}{2} \\ 1 \text{ if } i \notin G_1, \text{ or } i = \frac{N}{2}+1, ..., N \end{cases}.$$

$$E\left[\sum_{i=1}^{N} \tilde{S}_i \left(\sum_{t=1}^{T} \tilde{u}_{it}\right)\right]^2 = E\left[\sum_{i=1}^{N} \tilde{S}_i T \bar{u}_i\right]^2$$

where

$$\bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} \tilde{u}_{it}.$$

Observe this

$$\begin{aligned} & E\left[\sum_{i=1}^{N} \tilde{S}_i \left(\sum_{t=1}^{T} \tilde{u}_{it}\right)\right]^2 \\ = \; & E\left[-\frac{1}{2}\sum_{i=1}^{N/2}\sum_{t=1}^{T} \tilde{u}_{it} + \frac{1}{2}\sum_{i=N/2+1}^{N}\sum_{t=1}^{T} \tilde{u}_{it}\right]^2 \\ = \; & E\left[\frac{1}{4}\left(\sum_{i=1}^{N/2}\sum_{t=1}^{T} \tilde{u}_{it}\right)^2 + \frac{1}{4}\left(\sum_{i=N/2+1}^{N}\sum_{t=1}^{T} \tilde{u}_{it}\right)^2 - \frac{1}{2}\left(\sum_{i=1}^{N/2}\sum_{t=1}^{T} \tilde{u}_{it}\right)\left(\sum_{i=N/2+1}^{N}\sum_{t=1}^{T} \tilde{u}_{it}\right)\right] \end{aligned}$$

Note that if there is no cross section dependence, then the last third term becomes zero. Hence we have

$$
\begin{aligned}
E\left[\sum_{i=1}^{N}\tilde{S}_i\left(\sum_{t=1}^{T}\tilde{u}_{it}\right)\right]^2 &= \frac{1}{4}E\left(\sum_{i=1}^{N/2}\sum_{t=1}^{T}\tilde{u}_{it}\right)^2+\frac{1}{4}E\left(\sum_{i=N/2+1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\right)^2 \\
&= \frac{1}{4}\frac{N}{2}E\left(\frac{2}{N}\sum_{i=1}^{N/2}\sum_{t=1}^{T}\tilde{u}_{it}\right)^2+\frac{1}{4}\frac{N}{2}E\left(\frac{2}{N}\sum_{i=N/2+1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\right)^2 \\
&= \frac{NT}{8}\frac{\sigma^2}{(1-\rho)^2}+\frac{NT}{8}\frac{\sigma^2}{(1-\rho)^2}=\frac{NT}{4}\frac{\sigma^2}{(1-\rho)^2}
\end{aligned}
$$

where we use the fact

$$
E\left(\sum_{t=1}^{T}\tilde{u}_{it}\right)^2=\frac{\sigma^2}{(1-\rho)^2}. \quad \text{(To students: Prove this)}
$$

Note that

$$
E\left(\sum_{t=1}^{T}\tilde{u}_{it}\right)^2 > E\sum_{t=1}^{T}\tilde{u}_{it}^2=\frac{\sigma^2}{1-\rho^2}
$$

Solution: Use panel robust HAC estimator. Prove this.

Next, Consider the convergence rate: => must be $\sqrt{NT}$. Why?

Limiting Distribution: Major (nice) term and nuisance term For LSDV.

Nice term:

$$
G_{NT}=\frac{\sum^{NT}y_{it-1}u_{it}}{\sum^{NT}\tilde{y}_{it-1}^2}
$$

Nuisance term:

$$
N_{NT}=-\frac{1}{T}\frac{\sum^{N}\left(\sum^{T}y_{it-1}\right)\left(\sum^{T}u_{it}\right)}{\sum^{NT}\tilde{y}_{it-1}^2}
$$

Note that

$$
\hat{\rho}_{LSDV}-\rho=\frac{\sum^{NT}\tilde{y}_{it-1}\tilde{u}_{it}}{\sum^{NT}\tilde{y}_{it-1}^2}=G_{NT}+N_{NT}=G_{NT}+O_p\left(\frac{1}{?}\right),
$$

and G_{NT} is $O_p\left(\frac{1}{\sqrt{NT}}\right)$.

$$
\sqrt{NT}G_{NT}=\frac{\frac{1}{\sqrt{NT}}\sum^{NT}y_{it-1}u_{it}}{\frac{1}{NT}\sum^{NT}\tilde{y}_{it-1}^2}\to^d N\left(0,V^2\right)
$$

but

$$
\begin{aligned}
N_{NT} &= -\frac{1}{T}\frac{\frac{1}{NT}\sum^{N}\left(\sum^{T} y_{it-1}\right)\left(\sum^{T} u_{it}\right)}{\frac{1}{NT}\sum^{NT}\tilde{y}_{it-1}^{2}} \\
&= -\frac{1}{T}\frac{\frac{1}{N}\sum^{N}\left(\frac{1}{\sqrt{T}}\sum^{T} y_{it-1}\right)\left(\frac{1}{\sqrt{T}}\sum^{T} u_{it}\right)}{\frac{1}{NT}\sum^{NT}\tilde{y}_{it-1}^{2}} = -\frac{1}{T}\frac{\frac{1}{N}\sum^{N} O_p(1)\, O_p(1)}{O_p(1)} \\
&= -\frac{1}{\sqrt{N}}\frac{1}{T}\frac{\frac{1}{\sqrt{N}}\sum^{N} O_p(1)\, O_p(1)}{O_p(1)} = \frac{O_p(1)}{\sqrt{N}T} = O_p\left(\frac{1}{\sqrt{N}T}\right)
\end{aligned}
$$

Hence

$$
\sqrt{NT}\left(\hat{\rho}_{LSDV} - \rho\right) = \sqrt{NT}G_{NT} + \sqrt{NT}N_{NT} = O_p(1) + O_p\left(\frac{1}{\sqrt{T}}\right)
$$

so that as $T \to \infty$, we can ignore the second term.

Sample Final Exam:

Part I: Definition and Explanation

Q1: Cointegration

Q2: Unit Root Test

Q3: Weakly Stationarity

Q4: Newey and West Estimator

Q5: Panel Robust Covariance Estimator

Q6: White Heteroskedasticity Consistent Estimator

Q7: Nickell Bias

Q8: Relationship among between, within and pooled estimators

Q9: First Difference GMM/IV estimator in Dynamic Panel Regression

Q10: Hausman Test for Fixed Effects

Q11: Granger Causality Test

Q12: Error Correction Model

Part II: Proof and Derivation

Consider the following DGP

$$y_{it} = a_i + y_{it}^o, \quad y_{it}^o = \rho y_{it-1}^o + u_{it}, \quad u_{it} \sim iid\,(0,1)\,, \quad y_{i0}^o = u_{i0}.$$

Q1: Assume $\rho = 1$. You run the following regression

$$y_{it} = \alpha y_{it-1} + e_{it} \tag{56}$$

(a) Show that the pooled OLS estimator in (56) becomes consistent for fixed T and large N. That is,

$$\text{plim}_{N\to\infty}\hat{\alpha}_{\text{pols}} = 1$$

(b) Derive the limiting distribution of $\hat{\alpha}_{\text{pols}}$ when $N, T \to \infty$ jointly.

Now you add fixed effects.

$$y_{it} = \beta_i + \alpha y_{it-1} + \varepsilon_{it} \tag{57}$$

(c) Show that the within group estimator in (57) becomes inconsistent. (for fixed T large N).

(d) Suppose that $N/T \to 0$ as $N, T \to \infty$. Derive the limiting distribution of $\hat{\alpha}_{\mathrm{FE}}$.

Q2: Assume $|\rho| < 1$. You run (56).

(a) Find the moment conditions that the pooled OLS becomes consistent.

(b) Under the conditon of (a), derive the limiting distribution of $\hat{\alpha}_{\mathrm{pols}}$.